

ADVANCING GENOMICS FOR RARE DISEASE DIAGNOSIS AND THERAPY DEVELOPMENT

EDITED BY: Zhichao Liu, Weida Tong, Tielu Shi, Mike Mikailov and Ruth Roberts
PUBLISHED IN: Frontiers in Pharmacology, Frontiers in Genetics and
Frontiers in Pediatrics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88966-162-6

DOI 10.3389/978-2-88966-162-6

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ADVANCING GENOMICS FOR RARE DISEASE DIAGNOSIS AND THERAPY DEVELOPMENT

Topic Editors:

Zhichao Liu, National Center for Toxicological Research (FDA), United States

Weida Tong, National Center for Toxicological Research (FDA), United States

Tielu Shi, East China Normal University, China

Mike Mikailov, United States Food and Drug Administration, United States

Ruth Roberts, Apconix, United Kingdom

Citation: Liu, Z., Tong, W., Shi, T., Mikailov, M., Roberts, R., eds. (2020). Advancing Genomics for Rare Disease Diagnosis and Therapy Development. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-162-6

Table of Contents

- 06 Editorial: Advancing Genomics for Rare Disease Diagnosis and Therapy Development**
Zhichao Liu, Ruth Roberts, Tielu Shi, Mike Mikailov and Weida Tong
- 09 Phenotype and Molecular Characterizations of 30 Children From China With NR5A1 Mutations**
Yanning Song, Lijun Fan and Chunxiu Gong
- 18 Phenotype-Driven Virtual Panel is an Effective Method to Analyze WES Data of Neurological Disease**
Xu Wang, Xiang Shen, Fang Fang, Chang-Hong Ding, Hao Zhang, Zhen-Hua Cao and Dong-Yan An
- 29 Growth Concerns in Coffin–Lowry Syndrome: A Case Report and Literature Review**
Ying Lv, Liuyan Zhu, Jing Zheng, Dingwen Wu and Jie Shao
- 34 Two Novel AGXT Mutations Cause the Infantile Form of Primary Hyperoxaluria Type I in a Chinese Family: Research on Missed Mutation**
Xiulan Lu, Weijian Chen, Liping Li, Xinyuan Zhu, Caizhi Huang, Saijun Liu, Yongjia Yang and Yaowang Zhao
- 43 Excess of Rare Missense Variants in Hearing Loss Genes in Sporadic Meniere Disease**
Alvaro Gallego-Martinez, Teresa Requena, Pablo Roman-Naranjo and Jose A. Lopez-Escamez on behalf of the Meniere Disease Consortium (MeDiC)
- 55 Next Generation Sequencing and Animal Models Reveal SLC9A3R1 as a New Gene Involved in Human Age-Related Hearing Loss**
Giorgia Giroto, Anna Morgan, Navaneethakrishnan Krishnamoorthy, Massimiliano Cocca, Marco Brumat, Sissy Bassani, Martina La Bianca, Mariateresa Di Stazio and Paolo Gasparini
- 70 Biochemical, Molecular, and Clinical Characterization of Patients With Primary Carnitine Deficiency via Large-Scale Newborn Screening in Xuzhou Area**
Wei Zhou, Huizhong Li, Ting Huang, Yan Zhang, Chuanxia Wang and Maosheng Gu
- 79 A Novel Homozygous Frameshift Variant in XYLT2 Causes Spondyloocular Syndrome in a Consanguineous Pakistani Family**
Mehran Kausar, Elaine Guo Yan Chew, Hazrat Ullah, Mariam Anees, Chiea Chuen Khor, Jia Nee Foo, Outi Makitie and Saima Siddiqi
- 86 Compound Heterozygous CHAT Gene Mutations of a Large Deletion and a Missense Variant in a Chinese Patient With Severe Congenital Myasthenic Syndrome With Episodic Apnea**
Zhimei Liu, Li Zhang, Danmin Shen, Changhong Ding, Xinying Yang, Weihua Zhang, Jiuwei Li, Jie Deng, Shuai Gong, Jun Liu, Suyun Qian and Fang Fang

- 93 ***Growth Pattern in Chinese Children With 5 α -Reductase Type 2 Deficiency: A Retrospective Multicenter Study***
Xiu Zhao, Yanning Song, Shaoke Chen, Xiumin Wang, Feihong Luo, Yu Yang, Linqi Chen, Ruimin Chen, Hui Chen, Zhe Su, Di Wu and Chunxiu Gong
- 106 ***Immunohistochemical Evaluation of Histological Change in a Chinese Milroy Disease Family With Venous and Skin Abnormalities***
Sijia Zhang, Xihui Chen, Lijuan Yuan, Shuyan Wang, Dangzhi Moli, Shujuan Liu and Yuanming Wu
- 113 ***Comparative Analysis for the Performance of Variant Calling Pipelines on Detecting the de novo Mutations in Humans***
Yu Liang, Li He, Yiru Zhao, Yinyi Hao, Yifan Zhou, Menglong Li, Chuan Li, Xuemei Pu and Zhining Wen
- 122 ***A Novel Nonsense Mutation in FERMT3 Causes LAD-III in a Pakistani Family***
Saba Shahid, Samreen Zaidi, Shariq Ahmed, Saima Siddiqui, Aiysha Abid, Shabbir Malik and Tahir Shamsi
- 127 ***Next-Generation Sequencing Analysis Reveals Novel Pathogenic Variants in Four Chinese Siblings With Late-Infantile Neuronal Ceroid Lipofuscinosis***
Xiao-Tun Ren, Xiao-Hui Wang, Chang-Hong Ding, Xiang Shen, Hao Zhang, Wei-Hua Zhang, Jiu-Wei Li, Chang-Hong Ren and Fang Fang
- 137 ***The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics***
Ruili Huang, Ivan Grishagin, Yuhong Wang, Tongan Zhao, Jon Greene, John C. Obenauer, Deborah Ngan, Dac-Trung Nguyen, Rajarshi Guha, Ajit Jadhav, Noel Southall, Anton Simeonov and Christopher P. Austin
- 150 ***A New SLC10A7 Homozygous Missense Mutation Responsible for a Milder Phenotype of Skeletal Dysplasia With Amelogenesis Imperfecta***
Virginie Laugel-Haushalter, Séverine Bär, Elise Schaefer, Corinne Stoetzel, Véronique Geoffroy, Yves Alembik, Naji Kharouf, Mathilde Huckert, Pauline Hamm, Joseph Hemmerlé, Marie-Cécile Manière, Sylvie Friant, Hélène Dollfus and Agnès Bloch-Zupan
- 158 ***Molecular Genetics Analysis of 70 Chinese Families With Muscular Dystrophy Using Multiplex Ligation-Dependent Probe Amplification and Next-Generation Sequencing***
Dong Wang, Min Gao, Kaihui Zhang, Ruifeng Jin, Yuqiang Lv, Yong Liu, Jian Ma, Ya Wan, Zhongtao Gai and Yi Liu
- 170 ***COL1A1/2 Pathogenic Variants and Phenotype Characteristics in Ukrainian Osteogenesis Imperfecta Patients***
Lidiia Zhytnik, Katre Maasalu, Andrey Pashenko, Sergey Khmyzov, Ene Reimann, Ele Prans, Sulev Kõks and Aare Märtson
- 184 ***Novel Neonatal Variants of the Carbamoyl Phosphate Synthetase 1 Deficiency: Two Case Reports and Review of Literature***
Beibei Yan, Chao Wang, Kaihui Zhang, Haiyan Zhang, Min Gao, Yuqiang Lv, Xiaoying Li, Yi Liu and Zhongtao Gai
- 195 ***Association of SERPINC1 Gene Polymorphism (rs2227589) With Pulmonary Embolism Risk in a Chinese Population***
Yongjian Yue, Qing Sun, Lu Xiao, Shengguo Liu, Qijun Huang, Minlian Wang, Mei Huo, Mo Yang and Yingyun Fu

- 205** *Systematically Analyzing the Pathogenic Variations for Acute Intermittent Porphyrria*
Yibao Fu, Jinmeng Jia, Lishu Yue, Ruiying Yang, Yongli Guo, Xin Ni and Tieliu Shi
- 216** *Frequent Mutations of VHL Gene and the Clinical Phenotypes in the Largest Chinese Cohort With Von Hippel–Lindau Disease*
Baoan Hong, Kaifang Ma, Jingcheng Zhou, Jiufeng Zhang, Jiangyi Wang, Shengjie Liu, Zhongyuan Zhang, Lin Cai, Ning Zhang and Kan Gong
- 225** *Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders*
Alireza Shahryari, Marie Saghaeian Jazi, Saeed Mohammadi, Hadi Razavi Nikoo, Zahra Nazari, Elaheh Sadat Hosseini, Ingo Bartscher, Seyed Javad Mowla and Heiko Lickert
- 250** *Genotype–Phenotype Association Analysis Reveals New Pathogenic Factors for Osteogenesis Imperfecta Disease*
Jingru Shi, Meng Ren, Jinmeng Jia, Muxue Tang, Yongli Guo, Xin Ni and Tieliu Shi
- 262** *Transcriptome Analysis Identifies Piwi-Interacting RNAs as Prognostic Markers for Recurrence of Prostate Cancer*
Yuanli Zuo, Yu Liang, Jiting Zhang, Yingyi Hao, Menglong Li, Zhining Wen and Yun Zhao
- 270** *Sodium Taurocholate Cotransporting Polypeptide (NTCP) Deficiency Hidden Behind Citrin Deficiency in Early Infancy: A Report of Three Cases*
Hui Lin, Jian-Wu Qiu, Yaqub-Muhammad Rauf, Gui-Zhi Lin, Rui Liu, Li-Jing Deng, Mei Deng and Yuan-Zong Song
- 278** *Clinical Assessments and EEG Analyses of Encephalopathies Associated With Dynamin-1 Mutation*
Hua Li, Fang Fang, Manting Xu, Zhimei Liu, Ji Zhou, Xiaohui Wang, Xiaofei Wang and Tongli Han
- 288** *Neurologic Manifestations as Initial Clinical Presentation of Familial Hemophagocytic Lymphohistiocytosis Type2 Due to PRF1 Mutation in Chinese Pediatric Patients*
Wei-xing Feng, Xin-ying Yang, Jiu-wei Li, Shuai Gong, Yun Wu, Wei-hua Zhang, Tong-li Han, Xiu-wei Zhuo, Chang-hong Ding and Fang Fang



Editorial: Advancing Genomics for Rare Disease Diagnosis and Therapy Development

Zhichao Liu^{1*}, Ruth Roberts^{1,2,3}, Tielu Shi^{1†}, Mike Mikailov⁴ and Weida Tong^{1*}

¹ Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, US Food and Drug Administration, Jefferson, AR, United States, ² Department of Drug Safety, Apconix, Alderley Edge, United Kingdom, ³ Department of Biosciences, University of Birmingham, Birmingham, United Kingdom, ⁴ Office of Science and Engineering Labs, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, United States

Keywords: rare diseases, NGS, genomics, diagnosis, drug discovery

OPEN ACCESS

Edited and reviewed by:

Alastair George Stewart,
The University of Melbourne, Australia

*Correspondence:

Zhichao Liu
zhichao.liu@fda.hhs.gov
Weida Tong
Weida.tong@fda.hhs.gov

†Present address:

Tielu Shi,
The Center for Bioinformatics and
Computational Biology, The Institute of
Biomedical Sciences and School of
Life Sciences, School of Statistics, Key
Laboratory of Advanced Theory and
Application in Statistics and Data
Science-Ministry of Education,
East China Normal University,
Shanghai, China

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 25 August 2020

Accepted: 07 September 2020

Published: 25 September 2020

Citation:

Liu Z, Roberts R, Shi T, Mikailov M and
Tong W (2020) Editorial: Advancing
Genomics for Rare Disease Diagnosis
and Therapy Development.
Front. Pharmacol. 11:598889.
doi: 10.3389/fphar.2020.598889

Editorial on the Research Topic

Advancing Genomics for Rare Disease Diagnosis and Therapy Development

Rare diseases affect only a small percentage of the population and are often chronic and potentially life-threatening. There are more than 7,000 known rare diseases, and yet fewer than 700 approved treatment options are available. Progress made with the use of emerging technologies such as next-generation sequencing (NGS) and bioengineering holds great promise in advancing rare disease diagnosis and therapy development (Liu et al., 2019). This Research Topic brings together a multi-faceted approach to improve rare disease diagnosis and treatment development using genomic technologies, including (1) causal genetic variant identification (2) experimental designs for genetic testing studies; (3) population-specific genetic testing; (4) genotype-phenotype association; (5) enhanced diagnostic power; (6) retrospective and prospective studies; (7) reproducibility of NGS-based genetic testing; and (8) data management and bioinformatics pipeline development.

Identification of disease-causing variants is essential for improving our understanding of disease history and accelerating diagnostic biomarker discovery for rare diseases. NGS technology provides unprecedented speed and resolution to reveal individual genetic makeup, underpinning the identification of rare disease causal variants. Girotto et al. employed targeted sequencing to screen a large cohort of 464 Italian patients to prioritize genetic variants related to age-related hearing loss (ARHL). They found unreported genetic variants in SLC9A3R1 and further confirmed the finding using an in-vivo zebrafish knock-in model. Gallego-Martinez et al. conducted a targeted sequencing analysis of 890 Sporadic Ménière's disease (MD) patients to identify the rare missense variants associated with hearing loss-related genes. These studies helped design a "fit-for-purpose" gene panel to diagnose MD. Liu et al. used targeted sequencing to find a compound heterozygous inherited deletion in the CHAT gene in a Chinese patient with severe congenital myasthenic syndrome with episodic apnea (CMS-EA). Besides targeted sequencing, there are increasing applications of whole genome sequencing/whole exome sequencing (WGS/WES) to detect complex genetic variants and provide complete genetic information in support of rare disease diagnosis. Yan et al. identified novel neonatal variants associated with Carbamoyl phosphate synthetase I deficiency (CPS1D) using WES, and further verified their findings based on a comprehensive literature survey; this expanded our knowledge of the genetic variants of the CPS1 gene and associated phenotypes.

NGS that deploys trio-based and proband sampling design allows for more sensitive detection of *de novo* mutations that are present only in children, providing useful information on variants in recessive or imprinted disorders by inheritance. Kausar et al. carried out WES analysis on three spondyloocular syndrome (SOS) patients in a consanguineous Pakistani family and found a novel homozygous frameshift variant in XYLT2. Sanger sequencing was further used to confirm the findings and to reveal an autosomal recessive inheritance pattern. Shahid et al. reported a novel, homozygous FERMT3 nonsense mutation (c.286C > T, p.Q96₋) in the proband of an infant in a Pakistani family by using targeted sequencing. Novel causal variants may also facilitate the prenatal diagnosis of Leukocyte adhesion deficiency-III (LAD3). Zhang et al. associated the genetic variants (i.e., FLT4: c.3075G>A) identified from a Chinese trio Milroy disease (MD) family to different histological changes, providing more insight into the pathogenic impact of FLT4 mutations. Laugel-Haushalter et al. employed WES to detect SLC10A7 mutations in Amelogenesis imperfecta (AI) in an affected daughter from a consanguineous family, indicating a diversity of phenotypes associated with the mutations. Ren et al. identified three novel mutations (c.1551+1insTGAT in TPP1, c.244G>T in CLN6, c.554-5A>G in MFSD8) in genes associated with Late-Infantile Neuronal Ceroid Lipofuscinosis (LINCLs). The authors conducted third generation sequencing (TGS) with downstream pathological assessment and functional analysis on four late-infantile NCL siblings with similar phenotypic symptoms and found mutations located in different genes. Lu et al. used Sanger sequencing and WES to investigate a Chinese family in which two siblings were affected by the infantile form of primary hyperoxaluria type 1 (PH1). Two novel missense mutations were identified for the infantile form of PH1 by WES. Furthermore, the authors found that the same AGXT genotype caused the same infantile form of PH1 within the family.

A better understanding of allelic frequency across different populations is key to implementing “fit-for-purpose” diagnostic tools in rare diseases. Zhytnik et al. presented a Sanger sequencing analysis on 94 Ukrainian Osteogenesis imperfecta (OI) families. They identified 27 novel COL1A1/2 pathogenic variants, indicating that the spectrum of OI genotypes may differ between populations. Yue et al. conducted a case-control population study to screen the allele frequency of SERPINC1 variant rs2227589 in the Chinese population using the Sequenom assay. The study suggested that variant rs2227589 was associated with an increased risk of antithrombin deficiency in pulmonary embolism (PTE). Fu et al. carried out a meta-analysis by collecting 117 HMBS gene mutations from acute intermittent porphyria (AIP) patients and evaluated mutational impact on corresponding protein structures and functions. The authors found population disparities within 23 genetic variants across eight different ethnic groups, providing important information on population-based diagnostic biomarker development for AIP.

Although NGS can provide more information on the detailed genetic makeup of rare disease patients, it remains challenging to

precisely identify pathogenic variants with potential clinical applications. Several studies reviewed in this Research Topic explored different phenotypic anchoring strategies to establish the genotype-phenotype association in rare diseases. Hong et al. screened 540 patients from 187 unrelated Chinese Von Hippel-Lindau (VHL) families for 19 frequent VHL mutations, looking for associations between allelic frequency and clinical phenotypes. Furthermore, a Kaplan–Meier survival analysis was carried out to link allelic frequency with onset age. Lin et al. conducted a phenotypic association study, and found that Sodium Taurocholate Cotransporting Polypeptide (NTCP) deficiency could be covered up by citrin deficiency during early infancy for neonatal patients with the pathogenic variant c.800C > T (p.Ser267Phe) in gene SLC10A1. Feng et al. investigated clinical manifestations and genetic abnormalities in children with Familial hemophagocytic lymphohistiocytosis Type 2 (FHL2) using WES and enhanced magnetic resonance imaging (MRI). Lv et al. reported a case study of a mutant c.2185C > T in the RPS6KA3 gene associated with Coffin-Lowry syndrome (CLS), identified by targeted sequencing with magnetic resonance imaging (MRI). Further, they discussed the efficacy and safety of the application of growth hormone analogs in patients with CLS. Song et al. investigated the phenotype of patients with NR5A1 gene mutations in 30 Chinese patients (11 boys and 19 girls). No gender difference was identified regarding associated phenotypes due to genetic variants of NR5A1, but NR5A1 genetic variants were linked to different clinical manifestations. Zhao et al. conducted a retrospective multicenter study of 141 patients with 5 α -reductase type 2 deficiency (5 α RD). They investigated growth patterns and their association with clinical parameters such as luteinizing hormone (LH).

A combination of NGS and other advanced analytical technologies will enhance diagnostic power and the identification of disease-causing variants. Zhou et al. integrated NGS and tandem mass spectrometry (MS/MS) to screen 236,368 newborns for Primary carnitine deficiency (PCD) to precisely diagnose PCD. Wang D. et al. proposed a hybrid approach by combining multiplex ligation-dependent probe amplification (MLPA) and NGS to improve the precise identification of genetic variants in 70 Chinese families with suspected Muscular dystrophy (MD) probands.

Retrospective and prospective study using meta-analysis of reported genetic variants, especially for rare diseases, is an effective way to generate hypotheses and enhance understanding of underlying molecular mechanisms. Shi et al. implemented a meta-analysis to categorize the clinical phenotypes and genotypes based on variant types for 155 OI patients collected from the literature. The authors reported that three phenotypes (bone deformity, DI, walking with assistance) were enriched in two variation types (the Gly-substitution missense; and groups of frameshift, nonsense, and splicing variations). Li et al. conducted a prospective study to characterize the phenotypic, genetic, and electroencephalographic features of children with DNMI mutation-related epileptic encephalopathy. The authors suggested that patients carrying pathogenic variants in the

GTPase or middle domains present different epileptic encephalopathy and neurodevelopmental symptoms. Shahryari et al. reviewed 20 approved human gene and cell-based gene therapy products with great promise in treating devastating rare diseases and cancers. The pros and cons of these 20 gene therapy products were compared, and potential solutions for further improvement were provided.

Reproducible NGS-based genetic testing is vital for a successful clinical diagnosis of rare diseases. Liang et al. carried out a comparative analysis of a *de novo* variant calling with the state-of-the-art bioinformatics pipeline. The authors found a suboptimal concordance among the different calling algorithms and proposed a filtering strategy to improve the reproducibility of *de novo* variant identification.

Knowledge management and bioinformatics software development is essential for filtering and prioritizing clinically relevant genetic variants in rare diseases. Wang X. et al. introduced “Mingjian” software to improve genetic variant annotations. “Mingjian” software is a self-updating genetic disease computer-supported diagnostic system that integrates variant annotation databases, including HPO, OMIM, HGMD, with sophisticated statistical measures. Clinical utility of the software was verified with WES by using genetic variants detected from 30 neurological patients. Zuo et al. explored novel genetic elements such as piwi-

interacting RNAs (piRNAs) for their potential as prognostic biomarkers in the recurrence of prostate cancer using network analysis. Huang et al. created NCATS BioPlanet, a comprehensive, integrated pathway resource consisting of a universe of 1,658 human pathways sourced from publicly available, manually curated sources. It is a promising tool to facilitate research in systems biology, toxicology, and chemical genomics.

The power and promise of genomics for the diagnosis of rare diseases are exemplified in this special issue. However, many essential aspects of genomics in rare diseases were not included in this review of a small collection of articles. For example, drug repositioning for rare disease treatment development, standardization of rare disease terminology for clinical application, and novel NGS technologies are also of significant importance in widening the horizon of genomics application. We hope this Research Topic will trigger a broader interest and heightened community discussion toward the advancement of rare disease diagnosis and treatment development.

AUTHOR CONTRIBUTIONS

ZL wrote the first draft of the editorial. WT, RR, MM, and TS revised the editorial.

REFERENCE

Liu, Z., Zhu, L., Roberts, R., and Tong, W. (2019). Toward Clinical Implementation of Next-Generation Sequencing-Based Genetic Testing in Rare Diseases: Where Are We? *Trends Genet.* 35 (11), 852–867. doi: 10.1016/j.tig.2019.08.006

Disclaimer: The views presented in this article do not necessarily reflect current or future opinions or policies of the US Food and Drug Administration. Any mention of commercial products is for clarification and not intended as an endorsement.

Conflict of Interest: RR is co-founder and co-director of Apconix, an integrated toxicology and ion channel company that provides expert advice on non-clinical

aspects of drug discovery and drug development to academia, industry, and not-for-profit organizations.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Roberts, Shi, Mikailov and Tong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Phenotype and Molecular Characterizations of 30 Children From China With *NR5A1* Mutations

Yanning Song¹, Lijun Fan¹ and Chunxiu Gong^{1,2*}

¹ Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China, ² Beijing Key Laboratory for Genetics of Birth Defects, Beijing Children's Hospital, Capital Medical University, Beijing, China

Background: Patients harboring *NR5A1* mutations have a wide spectrum of phenotypes.

Objective: To investigate the phenotype of patients with *NR5A1* gene mutations from a 30 Chinese patient cohort.

Methods: We reported the clinical features of children with *NR5A1* gene mutations and compared them between two groups of patients with social genders of male (boys group) and female (girls group).

Results: Thirty patients with *NR5A1* mutations ranging from 2 months to 17 years of age were studied. There were 11 boys and 19 girls who were identified when they visited the hospital. The patients were verified as having testes without a uterus and ovaries by B-mode ultrasound. There was no difference between boys and girls in terms of the Prader stage ($p = 0.086$), but the position of the testes was higher in girls than in boys ($p = 0.013$). The patients' average height is -0.43 SDS according to the normal boys' height with SDS (while their average target height was 0.07 SDS). However, there was no such difference between boys and girls ($p > 0.05$). Although the basal LH and post-hCG testosterone (T) levels were not different ($p > 0.05$), but the basal FSH level, LH/FSH ratio, and INHB level were decreased in girls ($p = 0.002$; $p = 0.001$; $p = 0.006$). All of the mothers of the patients reported to have normal pregnancies. We found 24 patients (80%) with *de novo* mutations in the *NR5A1* gene; 5 patients had inherited mutations from their mothers, and one inherited from the father. Only the mothers of patients 16 and 18 showed premature ovarian failure at the time of reporting. Among 26 disease associated mutations, 14 novel mutations that have been reported the first time and p.R87C is the most common Among the other 12 had had been reported, the p.R313C is the most common.

Conclusion: Patients with 46, XY *NR5A1* mutations presented a wide spectrum of external genitalia characteristics and severe Sertoli cell impairment. The p.R87C and p.R313C mutations appeared to be common (10%) in this group, and 14 new mutations were identified, improving our understanding the genotype phenotype correlations.

Keywords: *NR5A1* mutation, novel mutation, phenotype and genotype, prader grade, steroli cell

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Michaël R. Laurent,
University Hospitals Leuven, Belgium
Bohu Pan,
National Center for Toxicological
Research (FDA), United States

*Correspondence:

Chunxiu Gong
chunxiugong@sina.com

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 09 July 2018

Accepted: 08 October 2018

Published: 30 October 2018

Citation:

Song Y, Fan L and Gong C (2018)
Phenotype and Molecular
Characterizations of 30 Children From
China With *NR5A1* Mutations.
Front. Pharmacol. 9:1224.
doi: 10.3389/fphar.2018.01224

Abbreviations: ACTH, adrenocorticotrophic hormone; AMH, anti-Müllerian hormone; Cor, cortisol; DBD, DNA-binding domain; DSD, disorder of sex development; FSH, follicle-stimulating hormone; GnRHa, gonadotropin-releasing hormone agonist; hCG, human chorionic gonadotropin; INHB, inhibin B; LBD, ligand-binding domain; LH, luteinizing hormone; POI, premature ovarian insufficiency; SDS, standard deviation score; T, testosterone.

INTRODUCTION

Disorders of sex development (DSDs) are defined as congenital conditions in which the development of chromosomal, gonadal, or phenotypic sex is abnormal (Ono and Harley, 2013). DSDs have been divided into three groups: Sex chromosome DSDs, 46, XX DSDs and 46, XY DSDs. The incidence of complete gonadal dysgenesis is estimated to be 1:80000 in newborns (Michala et al., 2008). One gene, *NR5A1* gene located on chr 9q33.3, has emerged play a major role as a common genetic cause in 10–20% of 46, XY DSD cases in the last few years (Suntharalingham et al., 2015; Takasawa et al., 2017), it encodes steroidogenic factor-1 (SF-1). SF-1 is a key regulator of steroidogenesis and reproductive development that controls several steps of adrenal and gonadal development. It stimulates the expression of several genes required for the development and maintenance of the male differentiation cascade. This gene also regulates the expression of *LHCGR* and the steroidogenic enzymes *STAR*, *CYP11A1*, and *CYP17A1* in Leydig cells, which are required for testosterone biosynthesis. *NR5A1* also increases the expression of insulin-like polypeptide 3 (*INSL3*) (Schimmer and White, 2010), which regulates testicular descent and is a survival factor for male germ cells in adults. The phenotypical spectrum encompasses from hypospadias (Tantawy et al., 2014), ambiguous genitalia, such as a hypoplastic phallus (Werner et al., 2015), to a complete external female appearance (Allali et al., 2011). The *NR5A1* gene has one nontranslated exon (exon 1) and six other coding exons (exon 2–7). SF1 has two zinc finger DNA-binding domains (DBDs), a ligand-binding domain (LBD), two functional activation domains (AF-1 and AF-2), an accessory region and a hinge region. The DBD contains a core with two Cys4 zinc finger motifs and a highly conserved Ftz-F1 box motif that is potentially involved in interactions with DNA (Parker and Schimmer, 1997).

Achermann et al. (1999, 2002) initially reported in 1999 and again in 2002 that heterozygous mutations in the *NR5A1* gene could cause adrenal insufficiency for patient with 46, XY severe testicular dysplasia. However, subsequent reports found only gonad dysfunction without adrenal insufficiency (Mallet et al., 2004; Philibert et al., 2007; Reuter et al., 2007; Kohler et al., 2008, 2009; Tajima et al., 2009). To date, only 4 given heterozygous mutations (p.G35D p.G35E, p.R92Q, and p.R255L) can cause adrenal insufficiency combined with gonadal dysfunction have been reported (Achermann et al., 1999, 2002; Orekhova et al., 2017).

Patients with 46, XY DSD present with normal AMH levels without a uterus or fallopian tubes (Coutant et al., 2007; Brandt et al., 2013), or have low concentrations of AMH and detectable Müllerian structures on B-mode ultrasound (Brandt et al., 2013). Some patients with low levels of AMH at birth but without apparent Müllerian structures have also been reported (Coutant et al., 2007; Kohler et al., 2009). In most cases, testosterone level is low during the neonatal period. Therefore, the phenotype has been known ranging from ambiguous genitalia to female external genitalia at birth (Tantawy et al., 2014; Yagi et al., 2015). However, there are also reports stating that patients with normal testosterone concentrations at birth or even at

puberty may show spontaneous pubertal progression or obvious virilization. This phenomenon suggests that the function of Leydig cells is sufficient for some patients. Patients showing persistently elevated FSH concentrations and low INHB and AMH concentrations have progressive gonadal failure that appears to occur with age, especially affecting the Sertoli cells; this finding is in accordance with the results of another study in which they followed up several patients younger than 30 years of age who presented with progressive gonadal dysgenesis after adolescence (Fabbri et al., 2014; Yagi et al., 2015; Werner et al., 2017).

Recently, in addition to causing 46, XY DSDs and adrenal dysfunction, missense mutations in *NR5A1* were identified as a cause of 46, XX testicular/ovotesticular disorders of sexual development (Baetens et al., 2017; Igarashi et al., 2017; Takasawa et al., 2017). While a genotype–phenotype relationship has not been established to date, approximately 120 *NR5A1* mutations have been documented in the Human Gene Mutation Database¹. Nonsense mutations are the most common. In this study, we evaluated the clinical features and genotypes of 30 Chinese Han patients with 46, XY DSDs.

MATERIALS AND METHODS

Patients

Patients ranging in age from 2 months to 17 years with various degrees of ambiguous external genitalia and *NR5A1* gene mutations were recruited from 30 unrelated families. All of the patients had been confirmed to have 46, XY karyotype. The clinical diagnoses in the patients with *NR5A1* mutations were based on incomplete virilization features, such as hypospadias, microphallus, cryptorchidism, clitoromegaly and complete female external genitalia. All of the patients also underwent *AR* and *SRD5A2* genetics analysis to exclude androgen insensitivity syndrome and 5 α -reductase type 2 deficiency, respectively. The informed consent for participation in the study were documented. The research protocol was approved by the Ethics Committee of Beijing Children's Hospital, Capital Medical University.

Clinical Information

The same pediatric endocrinologist performed these physical examinations and assessments. These information included age, social gender, chief complaint, family history, height, weight, facial features, clitoris/penis length, testicular position, urethral and vaginal meatus, electrolyte levels, and liver and kidney function, etc. Pituitary hormones and T concentrations were measured, and B-mode ultrasound or MRI was used to examine the patients' kidneys, adrenal glands, pelvic gonads and ducts. These patients were grouped based on their social gender of either male or female. The phenotype, hormones and gene mutations were compared between the two groups. The outcomes of the results led the individuals to undergo selective surgery or medical treatment for boys or girls for adaption to the reared gender.

¹<http://www.hgmd.cf.ac.uk/ac/gene.php?gene=NR5A1>

Clinical Features: Meatus and Testes Position and Classification

(i) Prader classification included the following values (Moshiri et al., 2012): Prader stage 0–1 = 1, Prader stage 2 = 2, Prader stage 3 = 3, Prader stage 4 = 4, and Prader stage 5 = 5. A Prader stage ≤ 3 was considered a severe phenotype. (ii) The testis position was classified with the following scores: abdominal = 1, inguinal = 2, labia = 3, and scrotum = 4. The lower the score was, the more serious the position was considered to be.

Hormone measurement was determined as follows. First, we examined the basal T concentrations. If T was at a prepubertal concentration, an hCG stimulation test was performed; namely, an injection of 15.00 IU of hCG per day was administered for 4 consecutive days. Peripheral blood samples were obtained after 12 h. T, FSH, LH, ACTH, and Cor were evaluated by radioimmunoassay techniques. MAGLUMI®2000. AMH was used for electrochemiluminescence, and INHB was evaluated by an ELISA assay.

Molecular Analysis

The *NR5A1*, *AR*, and *SRD5A2* genes were detected by the Beijing Key Laboratory for Genetics of Birth Defects and then confirmed by Sanger sequencing from January 2010 to July 2017. The sequencing data were analyzed by the authors.

Evaluation of Variant Pathogenicity

The genomics data are based on comparison with the NCBI reference sequence NM_004959. PolyPhen-2², SIFT³ and the ACMG guidelines were used to predict the impact of the identified mutations on protein function.

Statistical Analysis

We compared the clinical Prader stages, testes position values, basal T levels, T levels after hCG stimulation and LH/FSH ratios. In this study, the normally distributed values are described as the mean \pm standard deviation and were compared by *t*-tests. The nonnormally distributed data are described by the median and were compared with the Mann-Whitney U test. A 2-tailed *p*-value < 0.05 was considered statistically significant for all analyses. All statistical analyses were conducted using SPSS Statistics version 17.0.

RESULTS

Clinical Features

There were 30 patients with 46, XY DSDs who had *NR5A1* mutations. The patients' gonadal tissues were all testes without residual Müllerian or ovotesticular structures. Considering gender identity at birth, there were 19 girls and 11 boys. Thirteen of the girls presented with an inguinal mass, and the remaining 6 had obvious virilization at the first visit (older than 10 years with an obvious "Adam's apple," hoarse voice, and clitoris virilization).

The Prader stage of the external genitalia ranged from 0 to 4. One of the 11 boys had a micropenis, and the others had hypospadias (Prader stage of 2–5). There was no difference between genders when comparing Prader stages ($p = 0.086$), but the position of the testes was higher in girls than in boys ($p = 0.013$). The average height of all subjects was -0.43 SDS according to the normal boys' height standard (their average target height was 0.07 SDS), but there was no difference between boys and girls ($p > 0.05$). The bone ages were not clearly different. All mothers of the patients reported a normal pregnancy. The mothers of patients 16 and 18 carried the same mutations as the probands and showed delayed puberty and POI (menarche at 14 and 18 years and menopause at 40 and 36 years). Two sisters of patient 18 carried the same mutation and had delayed puberty, and two nephews (sons of the elder sister's) showed 46, XY hypospadias. The other patients' family histories were negative. The clinical characteristics of the patients are shown in **Tables 1, 2**.

Hormone Measurements

The average ACTH level was 21.2 ± 7.1 pg/ml (normal, 0–46), and the average Cor level was 9.6 ± 3.1 μ g/dl (normal, 5–25). The average basal LH/FSH ratio of the patients was 0.23 ± 0.29 , and the basal LH and post-hCG T levels showed no differences between boys and girls ($p > 0.05$); however, the basal FSH level, LH/FSH ratio, and INHB level were lower in girls than in boys ($p = 0.002$; $p = 0.001$; $p = 0.006$) (**Table 3**). There were 21 patients in prepuberty, 1 patient in mini-puberty (his T level was 81.5 ng/dl), and 8 undergoing spontaneous puberty, with basal T concentrations ranging from 99.1 to 289 ng/dl. The T concentrations were > 100 ng/dl after hCG stimulation, except in 2 patients who had T concentrations lower than 20 ng/dl.

Molecular Analysis of the *NR5A1* Gene

There were 24/30 patients with *de novo* mutations, accounting for 80% of the affected patients. Five mutations were from the mothers, including the mothers of patients 16 and 18 who showed premature ovarian failure (menopause at 40 and 36 years). The ages of the other 3 mothers were 42, 28, and 38 years old, and these mothers experienced normal puberty and normal menses. Patient 22 inherited the mutation from a 23 years old asymptomatic father.

NR5A1 genes were studied and showed 26 mutations. Twenty alleles were affected including 2 cuttings, 3 deletions and 2 insertions. Twelve are reported here: p.M1I, p.R84C, p.R92W, p.Q107*, p.P206Tfs*20, p.P216Afs*10, p.D293N, p.R313H, p.R313C, p.A351E, p.C412*, and c.1138 + 1G > A. Fourteen mutations have never been reported: p.S21E, p.G26R, p.T29M, p.C33*, p.V83M, p.R87C, p.104-105del, p.Y201*, p.T252=, p.G328R, p.Q417Rfs*13, p.E425Rfs*5, p.S430I, and c.245-2A > T. The mutations p.R87C and p.R313C accounted for 10% (3/30) of all mutations, and the rest of the mutations occurred only once each. We also found that exon 4 was the most commonly affected in 40% of patients. Exon 4 was affected in 72.7% of boys and 21% of girls, and exon 5 was the next most commonly affected (6/30). No mutations were found in exon 3. There were no differences in the clinical features between the

²<http://genetics.bwh.harvard.edu/pph2/>

³<http://sift.jcvi.org/>

TABLE 1 | Clinical data of 30 children with *NR5A1* mutations ID.

	Height (SDS)	PG	TP	Basal						P-hCG	AMH	INHB	Family History
				LH (mIU/ml)	FSH (mIU/ml)	T (ng/dl)	LH/FSH	ACTH (pg/ml)	Cor (ug/dl)				
1	−0.33	3	2	0.1	2.62	<20	0.04	27	9.1	183.7	23	83	N
2	0.19	2	2	0.07	1.94	10.08	0.04	12	7.2	300			N
3	−1.12	2	2	6.51	31.9	99.1	0.2	14	8.5	526			N
4	0.67	3	3	2.33	9.88	<20	0.24	15	7	120			N
5	−1.43	3	2	0.42	1.04	<20	0.4	24	9.8	231	1.43	32.5	N
6	−2.48	2	2	0.21	3.17	<20	0.07	42	11.8	168	23	99	N
7	−1.9	4	2	19.2	67	289	0.29	27	8.5	N/A			N
8	−0.22	3	2	0.22	4	<20	0.06	23	10.2	101			N
9	1.24	2	1	14.2	66	130	0.22	22	7.3	147			N
10	0.37	0	2	1.43	9.08	<20.0	0.16	24	11	219			N
11	0.31	4	2	4.25	22.2	121	0.19	23	9.6	632			N
12	−1.49	1	2	2.76	36.1	71.8	0.08	15	14.2	469	2.19	20.18	N
13	−0.33	2	2	0.1	2.13	<20	0.05	25	8.3	<20	7.89	33.6	N
14	0.11	0	3	3.67	30.1	86	0.12	18	8	579	23	59	N
15	−0.33	2	2	0.07	1.94	10.08	0.04	20	10.1	197	23	57.2	N
16	0.73	3	1	14.5	49.2	171	0.29	21	12.7	N/A	1.73	2.46	46, XX POI
17	0.31	2	2	3.95	28.4	209	0.14	18	16.5	N/A			N
18	−0.6	4	4	22.24	46.43	123.5	0.48	29	9.1	N/A	0.19	7.56	46, XX POI + 46, XX DSD
19	−0.93	2	2	4	23.4	100	0.17	34	6.9	N/A			N
20	0.75	3	4	0.46	2.83	61	0.16	11	5	595	23	88	N
21	−1.5	5	2	0.1	1.3	20	0.08	18	13	164	10.96	89.2	N
22	−1.21	4	2	3.54	17.2	81.5	0.21	19	8	N/A			N
23	−2.27	3	4	N/A	N/A	N/A	–	13	11.2	N/A			N
24	0.37	3	4	2.87	17.6	141	0.16	20	5.2	312			N
25	−0.68	3	3	0.1	3.2	<20	0.31	14	6.5	<20	23	39.3	N
26	−0.87	3	2	1.23	2.38	116	0.52	33	15.8	274	23	21.5	N
27	0.95	3	4	0.83	1.7	20	0.49	19	14.8	157	6.09	52.65	N
28	0.03	4	4	1.24	1.46	20	0.85	17	6.9	321	23	188.5	N
29	−0.1	3	4	0.16	9.27	20	0.02	14	5.6	420	23	260	N
30	−1.12	4	4	3.81	5.92	253	0.64	24	9.7	N/A	3.31	231.9	N

DBD and LBD ($p = 0.506$). The detailed gene mutants are shown in **Figures 1, 2**.

boys had repaired hypospadias, and others had treatment for a small penis. All of them are now enjoying their gender.

Follow-Up

The duration of follow-up was 2.96 ± 2.3 y. Sixteen of the 19 girls changed their genders to male, including five patients (patients 3, 7, 8, 11, and 17) who had spontaneous puberty. They had hypospadias repaired, including some after treatment with T undecanoate. Patients 16 and 19 used a GnRHa to inhibit gonadal development and waited until psychological maturation to make their own decisions regarding surgery at an appropriate time. Nine children were treated with T undecanoate to improve the appearance of a small penis. None of the patients underwent gonadectomy. Three girls’ parents tended to rear them as girls. One had no T response to hCG administration, another had a good T response, and another 11 years old patient had a basal T level > 120 ng/dl. For these three children, we advised the parents against surgery until the patients could make their own decision and encouraged GnRHa treatment when necessary. Five of 11

DISCUSSION

DSD patients with *NR5A1* mutations have demonstrated phenotypic variability without a clear genotype–phenotype relationship since the first case was reported. The external genitalia can present as normal, ambiguous, severe hypospadias or a female type (Camats et al., 2012; Domenice et al., 2016). Additionally, several reports have shown patients with 46, XX DSDs have ovarian malformations or POI. Until now, no relationship has been established (Lourenco et al., 2009). This study also showed a variety of clinical phenotypes in 30 Chinese patients. Previous study for heterozygous *NR5A1*+/– mice only revealed adrenal insufficiency during stress conditions and showed significant adrenal hyperplasia, demonstrating that normal gene dosage of SF-1 is required for mounting an adequate stress response (Bland et al., 2000). Though none of our 30

TABLE 2 | Genetic analysis results of 30 children with *NR5A1* mutations.

ID	Nucleotide Mutation	Amino acid Mutation	Exon	Mutants From	ACMG classification	Domain
1	c.63C > T	p.S21F	Exon 2	<i>De novo</i>	Uncertain significance	DBD
2	c.86C > T	p.T29M	Exon 2	<i>De novo</i>	Pathogenic	DBD
3	c.274C > T	p.R92W	Exon 4	<i>De novo</i>	Pathogenic	DBD
4	c.319C > T	p.Q107*	Exon 4	<i>De novo</i>	Pathogenic	DBD
5	c.305-310del	p.104-105del	Exon 4	<i>De novo</i>	Pathogenic	DBD
6	c.645_c.646insG	p.P216Afs*10	Exon 4	Mother	Pathogenic	Hinge Region
7	c.937C > T	p.R313C	Exon 5	<i>De novo</i>	Pathogenic	LBD
8	c.937C > T	p.R313C	Exon 5	<i>De novo</i>	Pathogenic	LBD
9	c.1052C > A	p.A351E	Exon 6	Mother	Pathogenic	LBD
10	c.1273delG	p.E425Rfs*5	Exon 7	<i>De novo</i>	Pathogenic	LBD
11	c.1289G > T	p.S430I	Exon 7	<i>De novo</i>	Likely pathogenic	LBD
12	c.1236C > A	p.Cys412*	Exon 7	Mother	Pathogenic	LBD
13	c.1138 + 1G > A	–	–	<i>De novo</i>	Pathogenic	–
14	c.245-2A > T	–	–	<i>De novo</i>	Pathogenic	–
15	c.982G > A	p.G328R	Exon 5	<i>De novo</i>	Pathogenic	LBD
16	3G > A	P.M1I	Exon 2	Mother	Pathogenic	DBD
17	c.1250delA	p.Q417Rfs*13	Exon 7	<i>De novo</i>	Pathogenic	LBD
18	c.99C > A	p.C33X	Exon 2	Mother	Pathogenic	DBD
19	c.938G > A	p.R313H	Exon 5	<i>De novo</i>	Pathogenic	LBD
20	c.76G > A	p.G26R	Exon 2	<i>De novo</i>	Uncertain significance	DBD
21	247G > a	p.V83M	Exon 4	<i>De novo</i>	Uncertain significance	DBD
22	c.250C > T	p.R84C	Exon 4	Father	Likely pathogenic	DBD
23	c.259C > T	p.R87C	Exon 4	<i>De novo</i>	Likely pathogenic	DBD
24	c.259C > T	p.R87C	Exon 4	<i>De novo</i>	Likely pathogenic	DBD
25	c.259C > T	p.R87C	Exon 4	<i>De novo</i>	Likely pathogenic	DBD
26	c.614dupC	p.P206Tfs*20	Exon 4	<i>De novo</i>	Pathogenic	Hinge Region
27	c.603T > A	p.Y201X	Exon 4	<i>De novo</i>	Pathogenic	Hinge Region
28	c.937C > T	p.R313C	Exon 5	<i>De novo</i>	Pathogenic	LBD
29	c.877G > A	p.D293N	Exon 5	<i>De novo</i>	Uncertain significance	LBD
30	c.756C > A	p.T252 =	Exon 4	<i>De novo</i>	Uncertain significance	LBD

The bold values mean the most frequency mutations.

TABLE 3 | Hormonal differences between boys and girls with *NR5A1* mutations.

Social gender	PG	Hight	T	LH	FSH	LH/FSH	INHB
Male	3 IQR (2–3)	–0.33 SDS (median)	282.87 ± 176.34	1.67 ± 1.56	7.84 ± 7.92	0.33 ± 0.26	121.38 ± 92.21
Female	2 IQR (3–4)	–0.68 SDS (median)	278.05 ± 193.67	5.34 ± 7.12	22.95 ± 22.61	0.17 ± 0.12	38.89 ± 31.67
<i>p</i>	0.086	0.384	<i>p</i> > 0.05	<i>p</i> > 0.05	0.002	0.006	0.001

patients had adrenal insufficiency, they may underlie subtle forms of subclinical adrenal insufficiency, which may become life threatening during traumatic stress. So it's necessary to assess adrenal function regularly in order to make adequate preparations under stress.

The gonads were all testes without any residual Müllerian structures. Although the Prader stage at the first visit showed no difference between boys and girls, that may be mainly due to the fact that the decision about social gender by the parents is not always the best one and as well as these patients undergoing masculine development gradually after birth. This phenomenon

also suggests that the Leydig cells may have considerable to be functional and can lead to virilization. This finding is consistent with previous research (Tantawy et al., 2012; Fabbri et al., 2014).

T concentrations were within the normal range in most patients, with mini-puberty or spontaneous puberty being displayed, and prepubertal patients had a good T response after hCG stimulation, except for two patients who likely had no remnant Leydig cells. We also observed remarkably increased FSH concentrations and a low LH/FSH ratio. The ratio was much lower in girls than in boys, which suggested that the female phenotype was a more severe type and that *NR5A1*

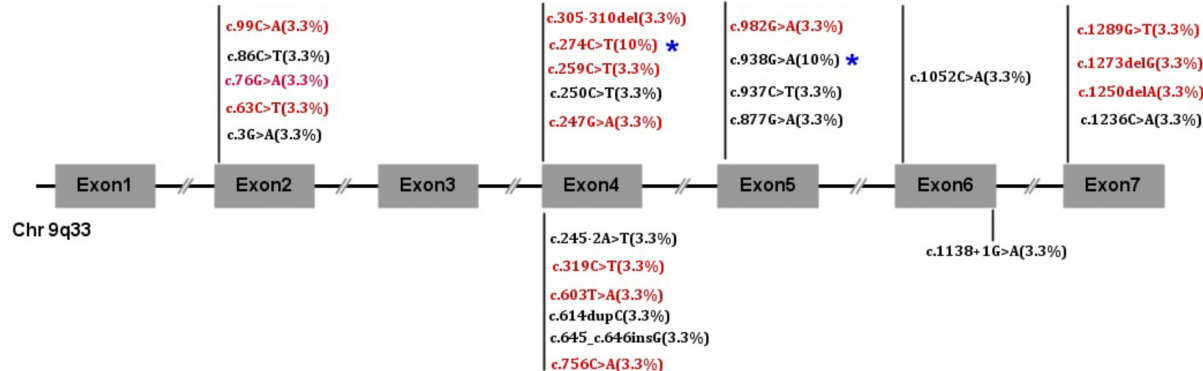
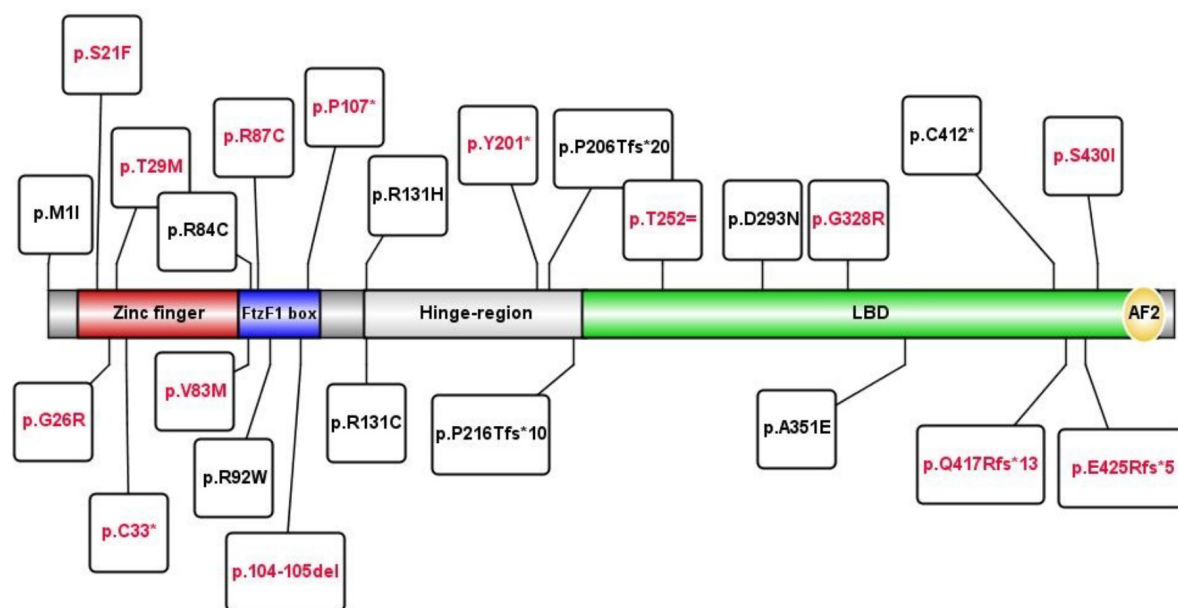


FIGURE 1 | *NR5A1* mutation frequency in 30 children. Red font indicates novel mutations, and black font indicates reported mutations.



Two other splice mutants: c.1138+1G>A and c.245-2A>T

FIGURE 2 | Relationship between mutations and the domain of the *NR5A1* gene. Red font indicates novel mutations, and black font indicates reported mutations.

mutations may more severely impair Sertoli cells than Leydig cells. S. Tantawy (Tantawy et al., 2012) observed similar results when examining gonadal function; Sertoli cell function gradually decreased with age, resulting in oligozoospermia or azoospermia. Decreased AMH can be associated with incomplete regression of Müllerian structure remnants with *NR5A1* mutations (Allali et al., 2011). In this group, patients had high FSH concentrations without Müllerian structures, demonstrating that there was enough AMH in the fetus to inhibit the growth of Müllerian structures. Whether or not patients develop Müllerian structures depends on the speed of attenuation of Sertoli cells. Therefore, as Tantawy suggested, Sertoli cell function decreases gradually with age (Allali et al., 2011; Tantawy et al., 2012). Two patients (patients 16 and 18) had a positive family history. The family

members carried mutations associated with 46, XY DSDs or 46, XX premature ovarian failure. On the other hand, in some families, the carriers had no clinical manifestations, such as the father and mother of some patients in this study, which may be related to incomplete penetrance or young age (Philibert et al., 2007; Lourenco et al., 2009). We report there was a "temporary" family history that was negative or an unreliable negative family history because we noticed that in all the literature we cited above, the researchers only recorded the menstrual cycle of the mother but not the age of menarche or menopause. Therefore, it was difficult to evaluate whether there was true incomplete penetrance or ambiguous family history. The same is the case when the mutant is inherited from the father.

After checking the available databases, we did not find on relation of height and *NR5A1* mutations carrier. Peycelon et al. (2017) reported that *NR5A1* mutations can cause growth retardation before 1 year of age, and this gap in height increases with age. However, another study showed that the heights of patients with *NR5A1* mutations were similar to those of their peers, especially among adolescent children, which may be associated with the secretion of T (Tantawy et al., 2012). In this study, the average height of the 30 children was lower than their target heights. The bone age was similar to that of normal children, indicating that T may play a role in growth and development, which is identical to the findings of Tantawy's study (Tantawy et al., 2012).

Gender rearing for DSD patients has been a hot topic of debate. Most 46, XY girls with *NR5A1* mutations will undergo progressive masculinity if the gonads are not removed. As shown above, boys with *NR5A1* mutations can undergo spontaneous puberty (Tantawy et al., 2012; Fabbri et al., 2014). Additionally, patients with *NR5A1* mutations and preserved fertility have also been reported, suggesting that a 46, XY individual with an *NR5A1* mutation reared as a boy has certain advantages (Bashamboo et al., 2010; Philibert et al., 2011; Yagi et al., 2015); this finding was confirmed by the case of the father of one of our patients who had the same mutation as the patient but had preserved fertility. In this group, five mothers had the same mutation, and 2 of them had premature ovarian failure, indicating that 46, XX individuals with *NR5A1* mutations can have a natural pregnancy. The 3 other mothers, with an average age of 36 years, and one father were normal, but this finding must be followed up to determine if these individuals had premature failure. These facts suggest that an early sex change may be hasty. Caution should be taken when early gonadectomy is considered to lead to irreparable changes. In this study, with the exception of three patients whose parents insisted on rearing them as girls, all others underwent repair of hypospadias, were reared as boys and continued to do well until now. We therefore suggest that GnRHa protects gonadal function and inhibits the masculinization process until the child reaches psychological maturity.

NR5A1 mutations are associated with a wide spectrum of gonadal development disorders, ranging from DSDs to oligo/azoospermia in 46, XY individuals (Camats et al., 2012; Domenice et al., 2016). In this study, 26 different mutations were found including 14 novel mutations. The mutations were mainly located in exon 4 and not in exon 3. This finding was different from those of previous studies in that the mutation sites were dispersed (Baetens et al., 2017; Rocca et al., 2018). The phenotypes of the reported mutations were similar to those previously published (Reuter et al., 2007; Allali et al., 2011; Camats et al., 2012; Yagi et al., 2015; Domenice et al., 2016; Fabbri et al., 2016; Baetens et al., 2017; Igarashi et al., 2017). We also identified 14 novel mutations that contributed to the improving our understanding on this subject.

The DBD region is crucial for transcription factors to bind to the DNA promoter and induce clinical expression. Studies

have shown that mutations in the DBD may be more severe than mutations in other regions (Venselaar et al., 2010; Fabbri et al., 2016). However, in this group, we did not find that mutations in the DBD were more severe than those affecting the LBD. Changes in p.G26E, causing severe phenotypes, and p.T29R, leading to moderate phenotypes, have been reported (Domenice et al., 2016). In this study, patient 20 who had a p.G26R mutation and patient 2 who had a p.T29M mutation presented as Prader stage 2–3. This finding was different from those of previous research and may be related to the protein conformation. The variant p.C33* directly interacts with the first zinc finger protein. It is known that cysteine affects the 3-D structure of the whole protein; therefore, cysteine changes lead to an inability of DNA to interact with the zinc finger, causing loss of the whole protein function. The patient only showed obvious masculinity at puberty with good testicular function. The Ftz-F1 box is considered a stabilizing region (DBD) for SF-1 binding to DNA and is highly conserved in different species (Little et al., 2006). The patients with p.V83M, p.R87C and p.104-105del mutations in the Ftz-F1 box motif presented as Prader stages 3-5 in this study. Thus, the clinical phenotype of the for patients with mutation in DBD domain remained to be investigated. The hinge region is an important domain for the interaction of the LBD with other proteins. Patient 27 had a moderate phenotype with the mutant p.Y201*, leading to SF-1 losing the LBD region, which was consistent with the findings of previous studies (Baetens et al., 2017). LBD mutations may have varying effects depending on their location and alterations in ligand specificity/recognition (Domenice et al., 2016). The p.E425Rfs*5 and p.Q417Rfs*13 mutations result in a truncated protein that loses the AF-2 region, leading to a loss of protein function. The two patients who presented as Prader stage 0 and Prader stage 2 had relatively severe phenotypes. The p.S430I mutation located in the LBD C-terminus causes the loss of hydrogen bonds in the protein core and disrupts correct protein folding (Venselaar et al., 2010); one patient with this mutation presented as Prader stage 4. Therefore, the severity of the clinical phenotype caused by the *NR5A1* mutation site is not necessarily dependent on the affected domain. The c.245-2A > T mutation was not included in the ExAC and was not an SNP, indicating that it was a rare mutation. The one patient who presented with this mutation presented as Prader stage 0, indicating that this mutation may likely promote a prematurely truncated protein and lead to a severe phenotype. There were no clear relationships among the novel mutations. Functional studies are necessary to verify the pathogenicity of these novel mutations and should be carried out in the future.

Limitations

First, the sample size may have been too small to establish a relationship between genotypes and phenotypes, and our future work will continue to increase the sample size to further focus on this topic. Second, the follow-up time can be extended to be longer period of time. Third, functional experiments will be needed to characterize the novel mutants, and we will perform these experiments in the following steps. Furthermore, compared

to mass spectrometry, the use of a radioimmunoassay may be insufficient for detecting T levels. Lastly, only *AR* and *SRD5A2* were examined in some patients, and other genes related to 46, XY DSDs were excluded. Therefore, we will use a gene panel or next-generation sequencing (NGS)-based approach to make the diagnosis more comprehensive.

CONCLUSION

Patients with 46, XY *NR5A1* mutations can clinically present with a wide spectrum of external genitalia and more severe Sertoli cell impairment than Leydig cell impairment. Mutations occurred often in exon 4. The novel p.R87C and reported p.R313C mutations appeared to be common (10%) in this group. The 14 new mutations enriched the mutation database and illuminated the particularity of Chinese people. The specific genotype–phenotype relationship remained to be established during future work.

REFERENCES

- Achermann, J. C., Ito, M., Ito, M., Hindmarsh, P. C., and Jameson, J. L. (1999). A mutation in the gene encoding steroidogenic factor-1 causes XY sex reversal and adrenal failure in humans. *Nat. Genet.* 22, 125–126. doi: 10.1038/9629
- Achermann, J. C., Ozisik, G., Ito, M., Orun, U. A., Harmanci, K., Gurakan, B., et al. (2002). Gonadal determination and adrenal development are regulated by the orphan nuclear receptor steroidogenic factor-1, in a dose-dependent manner. *J. Clin. Endocrinol. Metab.* 87, 1829–1833. doi: 10.1210/jcem.87.4.8376
- Allali, S., Muller, J. B., Brauner, R., Lourenco, D., Boudjenah, R., Karageorgou, V., et al. (2011). Mutation analysis of *NR5A1* encoding steroidogenic factor 1 in 77 patients with 46, XY disorders of sex development (DSD) including hypospadias. *PLoS One* 6:e24117. doi: 10.1371/journal.pone.0024117
- Baetens, D., Stoop, H., Peelman, F., Todeschini, A. L., Rosseel, T., Coppeters, F., et al. (2017). *NR5A1* is a novel disease gene for 46,XX testicular and ovotesticular disorders of sex development. *Genet. Med.* 19, 367–376. doi: 10.1038/gim.2016.118
- Bashamboo, A., Ferraz-de-Souza, B., Lourenco, D., Lin, L., Sebire, N. J., Montjean, D., et al. (2010). Human male infertility associated with mutations in *NR5A1* encoding steroidogenic factor 1. *Am. J. Hum. Genet.* 87, 505–512. doi: 10.1016/j.ajhg.2010.09.009
- Bland, M. L., Jamieson, C. A., Akana, S. F., Bornstein, S. R., Eisenhofer, G., Dallman, M. F., et al. (2000). Haploinsufficiency of steroidogenic factor-1 in mice disrupts adrenal development leading to an impaired stress response[J]. *Proc. Natl. Acad. Sci. U.S.A.* 97, 14488–14493. doi: 10.1073/pnas.97.26.14488
- Brandt, T., Blanchard, L., Desai, K., Nimkarn, S., Cohen, N., Edelmann, L., et al. (2013). 46,XY disorder of sex development and developmental delay associated with a novel 9q33.3 microdeletion encompassing *NR5A1*. *Eur. J. Med. Genet.* 56, 619–623. doi: 10.1016/j.ejmg.2013.09.006
- Camats, N., Pandey, A. V., Fernandez-Cancio, M., Andaluz, P., Janner, M., Toran, N., et al. (2012). Ten novel mutations in the *NR5A1* gene cause disordered sex development in 46,XY and ovarian insufficiency in 46,XX individuals. *J. Clin. Endocrinol. Metab.* 97, E1294–E1306. doi: 10.1210/jc.2011-3169
- Coutant, R., Mallet, D., Lahlou, N., Bouhours-Nouet, N., Guichet, A., Coupris, L., et al. (2007). Heterozygous mutation of steroidogenic factor-1 in 46,XY subjects may mimic partial androgen insensitivity syndrome. *J. Clin. Endocrinol. Metab.* 92, 2868–2873. doi: 10.1210/jc.2007-0024
- Domenice, S., Machado, A. Z., Ferreira, F. M., Ferraz-de-Souza, B., Lerario, A. M., Lin, L., et al. (2016). Wide spectrum of *NR5A1*-related phenotypes in 46,XY and 46,XX individuals. *Birth Defects Res. C Embryo Today* 108, 309–320. doi: 10.1002/bdrc.21145

AUTHOR CONTRIBUTIONS

CG conceived and designed the study, provided critical comments and edited the manuscripts. YS analyzed the data and wrote the first draft. YS and LF collected the data. All authors read and approved the final manuscript.

FUNDING

This work was funded by the Public Health Project for Residents in Beijing (Z151100003915103) and the National Key Research and Development Program of China (2016YFC0901505).

ACKNOWLEDGMENTS

We thank Yang Wei and all the other researchers at the Beijing Key Laboratory for Genetics of Birth Defects for the gene pathogenicity analysis.

- Fabbri, H. C., de Andrade, J. G., Soardi, F. C., de Calais, F. L., Petroli, R. J., Maciel-Guerra, A. T., et al. (2014). The novel p.Cys65Tyr mutation in *NR5A1* gene in three 46,XY siblings with normal testosterone levels and their mother with primary ovarian insufficiency. *BMC Med. Genet.* 15:7. doi: 10.1186/1471-2350-15-7
- Fabbri, H. C., Ribeiro, D. A. J., Maciel-Guerra, A. T., Guerra-Junior, G., and de Mello, M. P. (2016). *NR5A1* loss-of-function mutations lead to 46,XY partial gonadal dysgenesis phenotype: report of three novel mutations. *Sex. Dev.* 10, 191–199. doi: 10.1159/000448013
- Igarashi, M., Takasawa, K., Hakoda, A., Kanno, J., Takada, S., Miyado, M., et al. (2017). Identical *NR5A1* missense mutations in two unrelated 46,XX individuals with testicular tissues. *Hum. Mutat.* 38, 39–42. doi: 10.1002/humu.23116
- Kohler, B., Lin, L., Ferraz-de-Souza, B., Wieacker, P., Heidemann, P., Schroder, V., et al. (2008). Five novel mutations in steroidogenic factor 1 (SF1, *NR5A1*) in 46,XY patients with severe underandrogenization but without adrenal insufficiency. *Hum. Mutat.* 29, 59–64. doi: 10.1002/humu.20588
- Kohler, B., Lin, L., Mazen, I., Cetindag, C., Biebermann, H., Akkurt, I., et al. (2009). The spectrum of phenotypes associated with mutations in steroidogenic factor 1 (SF-1, *NR5A1*, Ad4BP) includes severe penoscrotal hypospadias in 46,XY males without adrenal insufficiency. *Eur. J. Endocrinol.* 161, 237–242. doi: 10.1530/EJE-09-0067
- Little, T. H., Zhang, Y., Matulis, C. K., Weck, J., Zhang, Z., Ramachandran, A., et al. (2006). Sequence-specific deoxyribonucleic acid (DNA) recognition by steroidogenic factor 1: a helix at the carboxy terminus of the DNA binding domain is necessary for complex stability. *Mol. Endocrinol.* 20, 831–843. doi: 10.1210/me.2005-0384
- Lourenco, D., Brauner, R., Lin, L., De Perdigo, A., Weryha, G., Muresan, M., et al. (2009). Mutations in *NR5A1* associated with ovarian insufficiency. *N. Engl. J. Med.* 360, 1200–1210. doi: 10.1056/NEJMoa0806228
- Mallet, D., Bretones, P., Michel-Calemard, L., Dijoud, F., David, M., and Morel, Y. (2004). Gonadal dysgenesis without adrenal insufficiency in a 46, XY patient heterozygous for the nonsense C16X mutation: a case of SF1 haploinsufficiency. *J. Clin. Endocrinol. Metab.* 89, 4829–4832. doi: 10.1210/jc.2004-0670
- Michala, L., Goswami, D., Creighton, S. M., and Conway, G. S. (2008). Swyer syndrome: presentation and outcomes. *BJOG* 115, 737–741. doi: 10.1111/j.1471-0528.2008.01703.x
- Moshiri, M., Chapman, T., Fechner, P. Y., Dubinsky, T. J., Shnorhavorian, M., Osman, S., et al. (2012). Evaluation and management of disorders of sex development: multidisciplinary approach to a complex diagnosis. *Radiographics* 32, 1599–1618. doi: 10.1148/rg.326125507

- Ono, M., and Harley, V. R. (2013). Disorders of sex development: new genes, new concepts. *Nat. Rev. Endocrinol.* 9, 79–91. doi: 10.1038/nrendo.2012.235
- Orekhova, A. S., Kalinchenko, N., Morozov, I. A., Vasilyev, E. V., Rubtsov, P. M., Dedov, I. I., et al. (2017). A Novel Mutation in the critical p-box residue of steroidogenic factor-1 presenting with XY sex reversal and transient adrenal failure. *Horm. Res. Paediatr.* 89, 450–454. doi: 10.1159/000481776
- Parker, K. L., and Schimmer, B. P. (1997). Steroidogenic factor 1: a key determinant of endocrine development and function. *Endocr. Rev.* 18, 361–377. doi: 10.1210/edrv.18.3.0301
- Peycelon, M., Mansour-Hendili, L., Hyon, C., Collot, N., Houang, M., Legendre, M., et al. (2017). Recurrent intragenic duplication within the NR5A1 gene and severe proximal hypospadias. *Sex. Dev.* 11, 293–297. doi: 10.1159/000485909
- Philibert, P., Polak, M., Colmenares, A., Lortat-Jacob, S., Audran, F., Poulat, F., et al. (2011). Predominant Sertoli cell deficiency in a 46,XY disorders of sex development patient with a new NR5A1/SF-1 mutation transmitted by his unaffected father. *Fertil. Steril.* 95, 1785–1788. doi: 10.1016/j.fertnstert.2010.11.035
- Philibert, P., Zenaty, D., Lin, L., Soskin, S., Audran, F., Leger, J., et al. (2007). Mutational analysis of steroidogenic factor 1 (NR5a1) in 24 boys with bilateral anorchia: a French collaborative study. *Hum. Reprod.* 22, 3255–3261. doi: 10.1093/humrep/dem278
- Reuter, A. L., Goji, K., Bingham, N. C., Matsuo, M., and Parker, K. L. (2007). A novel mutation in the accessory DNA-binding domain of human steroidogenic factor 1 causes XY gonadal dysgenesis without adrenal insufficiency. *Eur. J. Endocrinol.* 157, 233–238. doi: 10.1530/EJE-07-0113
- Rocca, M. S., Ortolano, R., Menabo, S., Baronio, F., Cassio, A., Russo, G., et al. (2018). Mutational and functional studies on NR5A1 gene in 46,XY disorders of sex development: identification of six novel loss of function mutations. *Fertil. Steril.* 109, 1105–1113. doi: 10.1016/j.fertnstert.2018.02.123
- Schimmer, B. P., and White, P. C. (2010). Minireview: steroidogenic factor 1: its roles in differentiation, development, and disease. *Mol. Endocrinol.* 24, 1322–1337. doi: 10.1210/me.2009-0519
- Suntharalingham, J. P., Buonocore, F., Duncan, A. J., and Achermann, J. C. (2015). DAX-1 (NR0B1) and steroidogenic factor-1 (SF-1, NR5A1) in human disease. *Best Pract. Res. Clin. Endocrinol. Metab.* 29, 607–619. doi: 10.1016/j.beem.2015.07.004
- Tajima, T., Fujiwara, F., and Fujieda, K. (2009). A novel heterozygous mutation of steroidogenic factor-1 (SF-1/Ad4BP) gene (NR5A1) in a 46, XY disorders of sex development (DSD) patient without adrenal failure. *Endocr. J.* 56, 619–624. doi: 10.1507/endocrj.K08E-380
- Takasawa, K., Igarashi, M., Ono, M., Takemoto, A., Takada, S., Yamataka, A., et al. (2017). Phenotypic variation in 46,XX disorders of sex development due to the NR5A1 p.R92W variant: a sibling case report and literature review. *Sex. Dev.* 11, 284–288. doi: 10.1159/000485868
- Tantawy, S., Lin, L., Akkurt, I., Borck, G., Klingmuller, D., Hauffa, B. P., et al. (2012). Testosterone production during puberty in two 46,XY patients with disorders of sex development and novel NR5A1 (SF-1) mutations. *Eur. J. Endocrinol.* 167, 125–130. doi: 10.1530/EJE-11-0944
- Tantawy, S., Mazen, I., Soliman, H., Anwar, G., Atef, A., El-Gammal, M., et al. (2014). Analysis of the gene coding for steroidogenic factor 1 (SF1, NR5A1) in a cohort of 50 Egyptian patients with 46,XY disorders of sex development. *Eur. J. Endocrinol.* 170, 759–767. doi: 10.1530/EJE-13-0965
- Venselaar, H., Te, B. T., Kuipers, R. K., Hekkelman, M. L., and Vriend, G. (2010). Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 11:548. doi: 10.1186/1471-2105-11-548
- Werner, R., Monig, I., August, J., Freiberg, C., Lunstedt, R., Reiz, B., et al. (2015). Novel insights into 46,XY disorders of sex development due to NR5A1 gene mutation. *Sex. Dev.* 9, 260–268. doi: 10.1159/000442309
- Werner, R., Monig, I., Lunstedt, R., Wunsch, L., Thorns, C., Reiz, B., et al. (2017). New NR5A1 mutations and phenotypic variations of gonadal dysgenesis. *PLoS One* 12:e176720. doi: 10.1371/journal.pone.0176720
- Yagi, H., Takagi, M., Kon, M., Igarashi, M., Fukami, M., and Hasegawa, Y. (2015). Fertility preservation in a family with a novel NR5A1 mutation. *Endocr. J.* 62, 289–295. doi: 10.1507/endocrj.EJ14-0340

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer BP and handling Editor declared their shared affiliation.

Copyright © 2018 Song, Fan and Gong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Phenotype-Driven Virtual Panel Is an Effective Method to Analyze WES Data of Neurological Disease

Xu Wang^{1*†}, Xiang Shen^{2†}, Fang Fang¹, Chang-Hong Ding¹, Hao Zhang², Zhen-Hua Cao² and Dong-Yan An²

¹ Department of Neurology, Beijing Children's Hospital, National Centre for Children's Health, Capital Medical University, Beijing, China, ² Running Gene Inc., Beijing, China

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University, China

Reviewed by:

Olimpia Musumeci,
Università degli Studi di Messina, Italy
Zhiping Liu,
Augusta University, United States

*Correspondence:

Xu Wang
zfwx05@126.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 15 July 2018

Accepted: 13 December 2018

Published: 09 January 2019

Citation:

Wang X, Shen X, Fang F, Ding C-H,
Zhang H, Cao Z-H and An D-Y
(2019) Phenotype-Driven Virtual Panel
Is an Effective Method to Analyze
WES Data of Neurological Disease.
Front. Pharmacol. 9:1529.
doi: 10.3389/fphar.2018.01529

Objective: Whole Exome Sequencing (WES) is an effective diagnostic method for complicated and multi-system involved rare diseases. However, annotation and analysis of the WES result, especially for single case analysis still remain a challenge. Here, we introduce a method called phenotype-driven designing “virtual panel” to simplify the procedure and assess the diagnostic rate of this method.

Methods: WES was performed in samples of 30 patients, core phenotypes of probands were then extracted and inputted into an in-house software, “Mingjian” to calculate and generate associated gene list of a virtual panel. Mingjian is a self-updating genetic disease computer supportive diagnostic system that based on the databases of HPO, OMIM, HGMD. The virtual panel that generated by Mingjian system was then used to filter and annotate candidate mutations. Sanger sequencing and co-segregation analysis among the family were then used to confirm the filtered mutants.

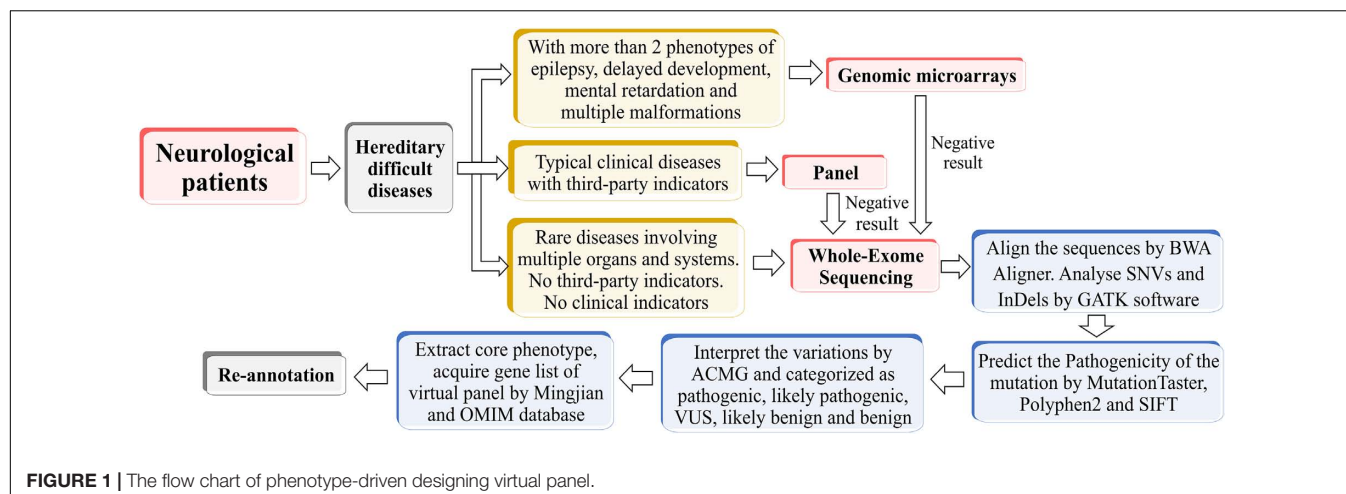
Result: We first used phenotype-driven designing “virtual panel” to analyze the WES data of a patient whose core phenotypes are ataxia, seizures, esotropia, puberty and gonadal disorders, and global developmental delay. Two mutations, c.430T > C and c.640G > C in PMM2 were identified by this method. This result was also confirmed by Sanger sequencing among the family. The same analysing method was then used in the annotation of WES data of other 29 neurological rare disease patients. The diagnostic rate was 65.52%, which is significantly higher than the diagnostic rate before.

Conclusion: Phenotype-driven designing virtual panel could achieve low-cost individualized analysis. This method may decrease the time-cost of annotation, increase the diagnostic efficiency and the diagnostic rate.

Keywords: WES, phenotype-driven, virtual panel, rare disease, annotation

INTRODUCTION

Rare Disease is defined as disease affected less than one in 2000 citizens in Europe, or less than one in 1250 in the United States (Schieppati et al., 2008). Rare diseases often start in childhood and accompanied by multisystem disorders which affect life quality of patients (Dodge et al., 2011; Elliott and Zuryski, 2015; Wright et al., 2018). Moreover, 33% of rare disease children die before 5 years old (Wright et al., 2018). There are now approximately 10,000 rare diseases



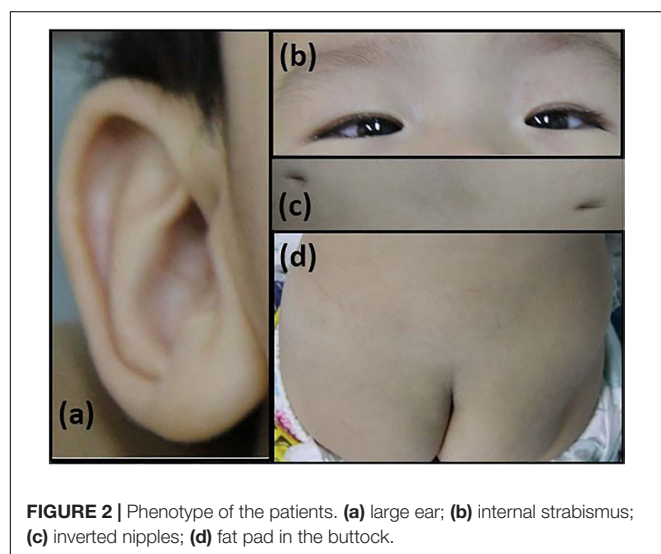
(Elliott and Zuryski, 2015), about 4 of 5 rare disease patients are thought to have a genetic base (Plaia et al., 2010; Dodge et al., 2011) especially monogenic disorder (Stolk et al., 2006). For some rare disorders such as tuberous sclerosis complex, phenotypes may vary among individuals due to heterogeneous manifestations. Merely diagnosis based on clinical presentations could be a great challenge (Bai et al., 2017). Hence, gene sequencing for the pathogenic genes is vital for understanding the cause of diseases.

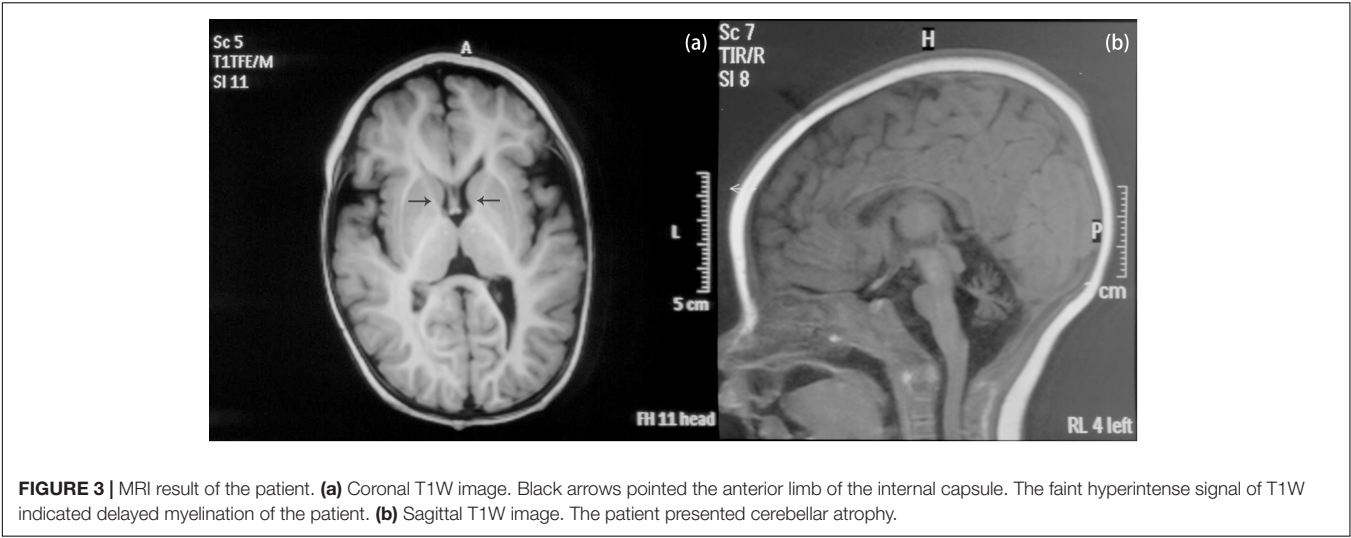
The mainstream of gene sequencing includes genomic microarrays, Sanger sequencing and Next-Generation sequencing (NGS). Genomic microarrays are low-resolution method for detection of 50~100 kb copy number variation (Speicher and Carter, 2005). For small insertion or deletion less than 50 kb, Sanger sequencing and NGS could fulfill the task. Sanger sequencing, due to limited throughput, is only used when a specific gene is selected. Different diseases could have similar clinical presentations such as ataxia and mental retardation. At the same time, a disease may be caused by

various genes. It is difficult to determine the pathogenic gene in every patient to perform Sanger sequencing. NGS offers much higher throughput that can facilitate sequencing up to 1000s of gene once. In addition, since sheared DNA is sequenced parallelly multiple times, therefore lower error rate is achieved compared to Sanger sequencing. Moreover, the recent study showed that NGS could also be used to detect Copy Number variation that larger than 100 kb (de Ligt et al., 2013; Feng et al., 2017). Therefore, it has been increasingly used in rare disease diagnosis.

For NGS, the range of detection object could vary from multiple disease-associated genes (gene panel), whole exome (Whole-Exome Sequencing) to whole genome (Whole-Genome Sequencing). For gene panel, various genes affected several similar diseases or diseases in the same system could be detected at the same time. Since it focuses on the specific genes, the data size is generally smaller than Whole-Exome Sequencing (WES) and Whole-Genome Sequencing (WGS), the result is easy to analyze and interpret. Although convenient, the gene list of a particular panel is constant; meanwhile, the discovery of disease-associated gene is developing. The newly discovered gene on one hand cannot be added to the already made panel, and further analysis cannot be performed. On the other hand, updating gene list every day is, however, impractical, costly and with less sense. Gene panels are at present insufficient for detection and is not recommended by most of the genetics and clinicians (Biesecker and Green, 2014; Wenger et al., 2017; Ewans et al., 2018; Jin et al., 2018).

WGS, mostly based on Illumina technology, is the sequencing method covers most part of the human genome. Although easy to perform, it is costly and time consuming to analyze and interpret data. On average, 3–4 million mutations could be discovered in each individual (Ashley et al., 2010; Lupski et al., 2010; Roach et al., 2010; Sobreira et al., 2010; Bainbridge et al., 2011). In the meantime, the mutations in the intronic region except for the ones near splicing sites are hard to predict the relative risk of phenotype, since the function of the intronic gene is still mostly undiscovered, and the mutation frequency in the intron is considerably high (Tabor et al., 2002; Abecasis et al., 2010). It





is hard to estimate which mutation is deleterious. Research also presented that WGS has limited significance at the present stage (Alfares et al., 2018). By contrast, the exome represents 1–2% protein-coding gene of the whole genome thus more exomes could be sequenced per run (Gilissen et al., 2012). The result of WES is more accessible to interpret since non-synonymous mutations in the coding region could directly lead to amino acid change then affect the protein structure and function. This method could also help identify not only the unknown pathological mutations but also the undiscovered mutations (Liu et al., 2012). Re-analysis of WES data was also proved to significantly increase the diagnostic rate (Alfares et al., 2018). The cost of WES is also much lower than WGS (Gilissen et al., 2012) at present. Although the number of variants is cut down to the range between 20,000 and 50,000 (Ashley et al., 2010; Lupski et al., 2010; Roach et al., 2010; Sobreira et al., 2010; Bainbridge et al., 2011; Gilissen et al., 2012), it is still difficult to analyze and identify the pathogenicity of every variant, especially for detection of single

case because of lower efficiency and time consuming. Meanwhile, due to the analysis strategy with less-efficacy, the diagnostic rate of WES with unspecific analysis was relatively low, approximately 25–30% (Yang et al., 2013; Lee et al., 2014; Shashi et al., 2016).

After carrying out, investigating and studying WES in clinic for many years, the combination of clinical information and gene sequencing is increasingly suggested in disease diagnosis (Jin et al., 2018). Here, we developed a method called “Phenotype-driven designing virtual panel,” a method that concentrates in analysing the genes of diseases with related phenotypes. The gene lists of phenotype-associated diseases were generated by a system called “Mingjian.” After inputting all phenotypes of the patient, the system will automatically list the associated genes and rank the gene by the corresponding number of phenotypes. This method is proved to improve the diagnostic rate significantly in our further test.

METHODS

Whole-Exome Sequencing

Proband DNA was sequenced to discover the causal gene. DNA was isolated from peripheral blood using a DNA Isolation

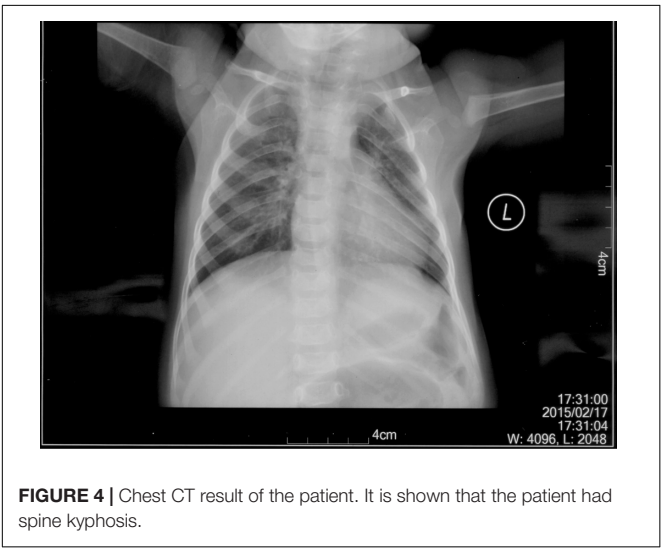


TABLE 1 | Mass spectrum result of patient’s blood.

	Result of patient		Reference for children (6 months to 1 year old)	Ratio
His	99.916	↑	0.00–79.30	1.260
Tyr	19.083	↓	19.40–79.40	0.240
Thr	87.512	↑	22.00–64.20	1.363
Phe/Tyr	1.980	↑	0.23–1.20	1.650
C5DC	0.099	↑	0.00–0.08	1.232
C0/C2	0.568	↓	0.82–2.40	0.237

Enhancement of His, Thr, Phe/Tyr, C5DC and regression of Tyr and C0/C2 indicated liver dysfunction of the patient.

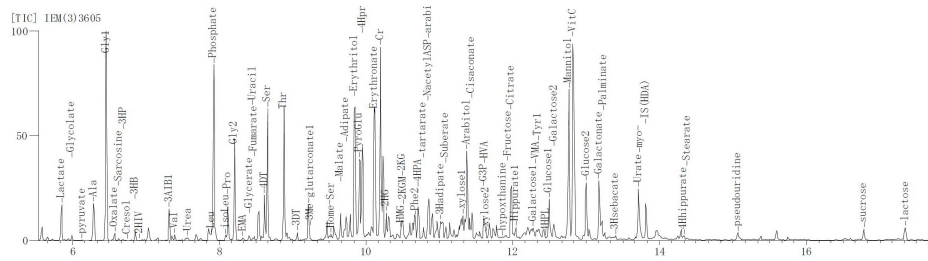


TABLE 2 | Blood test result of the patient's serum.

Kit (Bioteke, AU1802). 1 µg genomic DNA was fragmented into 200–300 bp length by Covaris Acoustic System. The DNA fragments were then processed by end-repairing, A-tailing and adaptor ligation, a 4-cycle pre-capture PCR amplification, targeted sequences capture. Captured DNA fragments were eluted and amplified by 15 cycle post-capture PCR. The final products were sequenced with 150 bp paired-end reads on Illumina HiSeq X platform according to the standard manual.

Sanger Sequencing

¹<http://bio-bwa.sourceforge.net/>

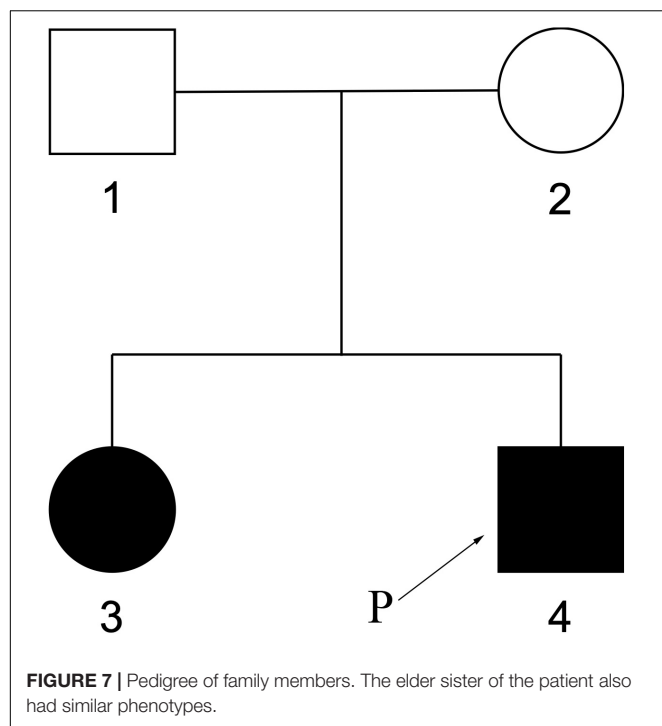
²www.broadinstitute.org/gatk

³annovar.openbioinformatics.org/en/latest/

⁴<http://omim.org/>

No	项目	结果	生物参考区间	单位	No	项目	结果	生物参考区间	单位
1	谷丙转氨酶(ALT*)	108	↑ 9-50	IU/L	21	白球比值(A/G)	2.41	↑ 1.2-2.4	
2	谷草转氨酶(AST*)	107	↑ 15-40	IU/L					
3	总蛋白(TP)	56.9	↓ 65-85	g/L					
4	白蛋白(ALB)	40.2	↓ 40-55	g/L					
5	碱性磷酸酶(ALP*)	180	50-240	IU/L					
6	谷氨酰转氨酶(GGT*)	13	10-60	IU/L					
7	总胆红素(TBIL)	7.8	1.7-20	μmol/L					
8	直接胆红素(DBIL)	0.10	0-6	μmol/L					
9	胆碱酯酶(CHE)	3226	↓ 4300-13200	IU/L					
10	前白蛋白(PA)	186.3	↓ 470-420	mg/L					
11	血肌酐(CREA*)	16.00	↓ 44-133	μmol/L					
12	尿素(UREA*)	5.68	1.8-7.1	mmol/L					
13	葡萄糖(GLU*)	4.34	3.61-6.11	mmol/L					
14	钾(K*)	5.05	3.5-5.3	mmol/L					
15	钠(NA*)	134.00	↓ 137-147	mmol/L					
16	氯(CL*)	108.8	99-110	mmol/L					
17	碳酸氢根(CO2)	18.50	↓ 22-36	mmol/L					
18	乳酸脱氢酶(LDH*)	432	↓ 100-240	IU/L					
19	阴离子间隙(AG)	13.75		mmol/L					
20	总胆汁酸(TBA)	6.29	0-10	μmol/L					

FIGURE 6 | The blood test result of the patient's serum. Alleviation of alanine aminotransferase (ALT) and Aspartate aminotransferase (AST) and deduction of plasma cholinesterase indicated liver dysfunction of the patient.



Chromas Lite v2.01 (Technelysium Pty Ltd., Tewantin, QLD, Australia).

A CASE OF A DIAGNOSTIC ODYSSEY

The patient is an 8 months old boy who was born to a normal non-consanguineous Han family by normal vaginal delivery at full-term. He had tonic seizure epilepsy with sustaining state when he first came to our hospital. His symptoms get alleviated obviously after taking levetiracetam 40 mg/kg per day. The milestone development and comprehensive development of the patient was also delayed. Physical examination: the head circumference of the patient was 41 cm, anterior fontanel was 1*1 cm. He had internal strabismus but could chase light, he also presented large ear, low nose, inverted nipples, low muscle tension with muscle strength-4, weak tendon reflex, poor head control, round back, fat pad in buttock, bilateral cryptorchidism and short penis. His body always leaned forward when sitting (**Figure 2**). He could not open his mouth or speak actively. He could neither grab things initiatively. Laboratory result: MRI result presented cerebellar atrophy and delayed myelination (**Figure 3**); chest CT showed spine kyphosis (**Figure 4**); EMG result showed neurogenic damage; the LC-MS/MS result of blood (**Table 1**), GC-MS result of urine (**Figure 5**) and blood test of patient's serum (**Table 2** and **Figure 6**) indicated abnormal liver function.

The elder sister of the patient, 8 years old, also shows somehow similar phenotypes. At 2 years of age, she started to have tonic epilepsy and ataxia, mental retardation, so far

can only speak 2–3 words phrase. The pedigree was shown in **Figure 7**.

The clinical presentation involved multiple systems and thus, even he has got treated at many hospitals and screened by existing detection methods, the disease was still unclear.

RESULTS

The Gene List of Phenotype-Driven Virtual Panel

Extracting and inputting the core phenotypes: Ataxia, Seizures, Esotropia, Puberty and Gonadal disorders, Global

TABLE 3 | Gene list exported by Software Mingjian according to the inputting core phenotypes.

Number of consilient phenotypes	Gene lists
6	PMM2 CEP290
5	GBA, POLG, GP1BB, HSD17B4, PEX1, PEX6, ERCC2 BCS1L, DOCK8, PEX10, TCF4, PEX12, ERCC6, RRM2B, PEX26, PEX2, ERCC4, PEX16, GRIN2B, PEX5, ERCC1, WDR73, PEX3, K1F1A, PEX14, PEX19, PEX11B, ADGRG1, C100RF2
4	ABCD1, SCN1A, ABCC8, PTS, SURF1, BTBD, NPC1, GCH1, ASL, CDKL5, ASS1, ATM, PRF1, GAMT, PDHA1, CPS1, OFD1, PLA2G6, SOX10, ETHE1, GJA1, ADSL, PROKR2, FGFR1, PPT1, FKRP, OTX2, POMGNT1, NPC2, SCO2, SIL1, BBS2, UNC13D, POMT1, TBX1, BBS1, STXBP1, BBS10, NDUFS4, ALMS1, GJC2, STXBP2, NPHP1, BRAF, HESX1, NDUFV1, ECHS1, MKKS, ERCC8, GMPPB, BBS12, NDUFS8, TUBB2B, POLR1C, COQ2, MKS1, SUCLG1, FMR1, BBS4, POLR3B, SPR, RAB3GAP1, ADLH5A1, RAF1, NDUFAF2, SDHA, EDNRB, CC2D2A, RARS2, ARL6, TSEN54, SUOX, SLC17A5, MBD5, POMT2, SCN2A, MMADHC, SCN9A, MFSD8, NDUFS2, SLC25A1, BBS7, POLR3A, PCNT, NDUFS6, EDN3, PDHX, PNKP, BBS9, WWOX, PSAP, DPM1, DYRK1A, NDUFA1, PET100, TTC8, ALG6, FKTN, DLD, NDUFS1, TMEM216, BBS5, SDCCAG8, SLC19A3, SYNGAP1, HIBCH, NDUFS7, COX6B1, NDUFAF1, MTFMT, SLC6A19, ALG1, LARGE, ERCC3, NOTCH1, CTC1, KCNJ10, GLI2, IFT172, TRIM32, NDUFS3, LIPT1, DOCK6, DYNC1H1, NDUFAF3, SCO1, NDUFB9, SLC46A1, NDUFA2, TMEM138, TMEM138, NDUFB3, DLL4, NDUFAF5, TTC19, GABRA1, COA3, FOXRED1, STX11, COX10, SLC25A4, DEAF1, ACO2, NDUFV2, B3GALNT2, GRIN1, APOPT1, NUBPL, TSFM, CDH15, NDUFA12, CYC1, WPCP, RAB3GAP2, RFT1, TACO1, COX14, TMEM231, TMEM237, NDUFA11, GRM1, NDUFAF6, ZNF423, RPIA, KIRREL3, ATP5A1, NDUFA4, IFT27, COMT, PDSS2, NDUFAF4, UQCRC2, LZTFL1, EOGT, UQCRCQ, NDUFA9, COX15, NDUFA10, UQCRC2, UQCRC3, DHFR, BBIP1, PDP1, CACNG2, PLXND1, COX20, ARHGAP31, RBPJ, EPB41L1, NIN, CTDP1, MYO5A, UQCRCB, NAT8L, LYRM7, FASTKD2, ZNF592, C5ORF42, ND3, ND2, TRNV, ND5, ND4, ND1, ATP6, CYTB, ND6, TRNL1, COX2, COX3, TRNK, RNU4ATAC, COX1, TRNW
3	OTC, DMD, PROC, SDHB... altogether 441 genes
2	GLA, PAH, GCK, GALT... altogether 543 genes
1	HBB, LDLR, MLH1... altogether 1427 genes

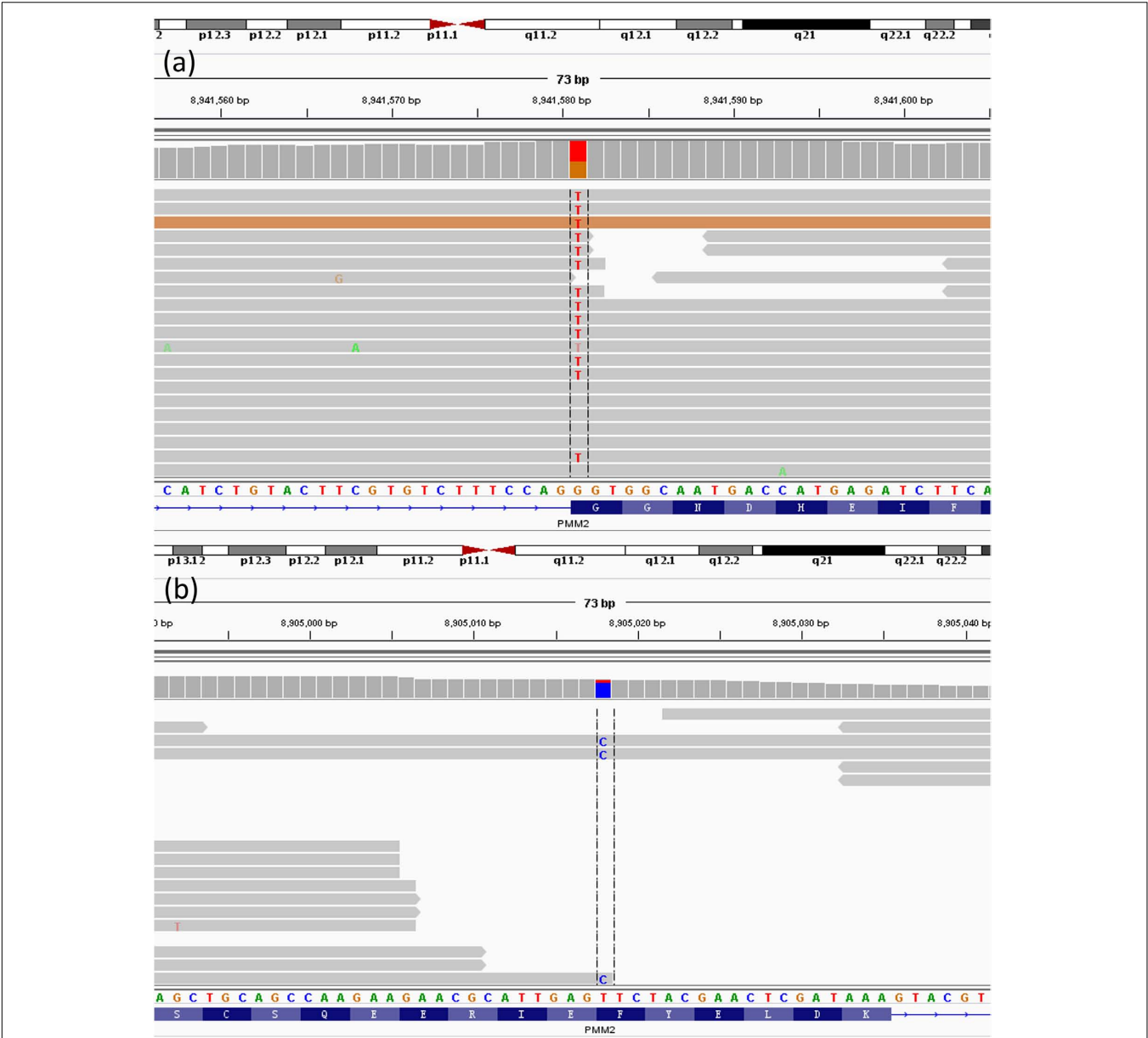


FIGURE 8 | Next-Generation result of patient. The result shows the two heterozygous mutations in the PMM2 gene. **(a)** c.640G > C; **(b)** c.430T > G.

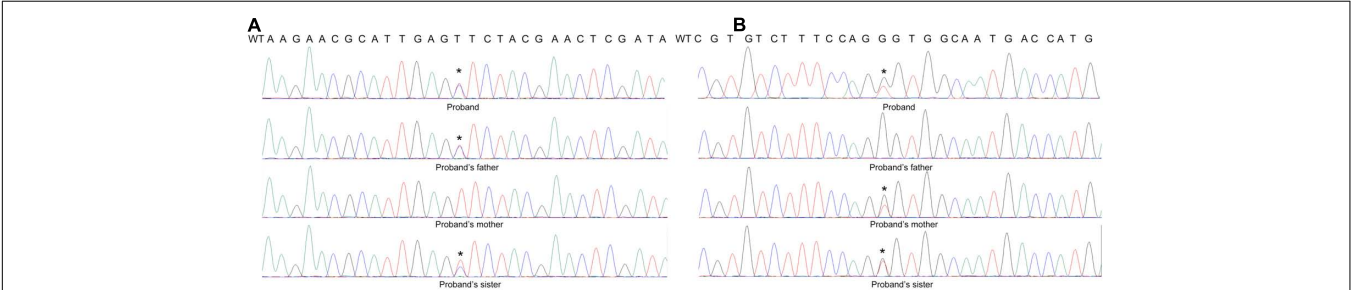


FIGURE 9 | Sanger Sequence result of the patient's family. The result shows that **(A)** the proband's father was the heterozygous carrier of the c.430T > C mutation, while **(B)** the proband's mother carried the c.640G > C mutation. The proband's sister is also the carrier of the compound heterozygous mutations.

splice sites	alteration within used splice site, likely to disturb normal splicing effect	gDNA position score	detection sequence	exon-intron border
Acceptor lost	58901	sequence motif lost	–	wt: ccag GGTG mu: ccag.CGTG
Acc marginally increased	58902	wt: 0.8914 / mu: 0.9172 (marginal change - not scored)	wt: TGTACTTCGTGCTTTCCAGGGTGGCAATGACCATGAGATC mu: TGTACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATC	cagg GTGG
Acc marginally increased	58896	wt: 0.3636 / mu: 0.4068 (marginal change - not scored)	wt: TTCATCTGACTTCGTGCTTTCCAGGGTGGCAATGACCATGAGATC mu: TTCATCTGACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATC	tctt TCCA
Acc marginally increased	58900	wt: 0.9541 / mu: 0.9606 (marginal change - not scored)	wt: TCTGTACTTCGTGCTTTCCAGGGTGGCAATGACCATGAGATC mu: TCTGTACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATC	tcca GGGT
Acc marginally increased	58893	wt: 0.8257 / mu: 0.8796 (marginal change - not scored)	wt: GCCTTCATCTGACTTCGTGCTTTCCAGGGTGGCAATGACCATGAGATC mu: GCCTTCATCTGACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATC	gtgt CTTT
Acc marginally increased	58899	wt: 0.9131 / mu: 0.9347 (marginal change - not scored)	wt: ATCTGTACTTCGTGCTTTCCAGGGTGGCAATGACCATGAGATC mu: ATCTGTACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATC	ttcc AGGG
Acc marginally decreased	58901	wt: 0.9826 / mu: 0.8961 (marginal change - not scored)	wt: CTGTACTTCGTGCTTTCCAGGGTGGCAATGACCATGAGATC mu: CTGTACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATC	ccag GGTG
Donor increased	58905	wt: 0.57 / mu: 0.79	wt: AGGCTGGCAATGACC mu: AGGCTGGCAATGACC	GGTG gcaa
Donor increased	58900	wt: 0.61 / mu: 0.77	wt: TTTCAGGGTGGCAA mu: TTTCAGCGTGGCAA	TCCA gggt
Donor increased	58904	wt: 0.38 / mu: 0.89	wt: CAGGGTGGCAATGAC mu: CAGCGTGGCAATGAC	GGGT ggca
Acc gained	58904	0.86	mu: TACTTCGTGCTTTCCAGCGTGGCAATGACCATGAGATCTT	gcgt GSCA

FIGURE 10 | Indicated splicing change by MutationTaster (Schwarz et al., 2014). This mutation might disturb the exon-intron border.

developmental delay, Autosomal recessive (inheritance pattern). The gene list exported by Mingjian is listed in **Table 3**.

Result of Whole-Exome Sequencing

Analysing the gene from gene list generated by Mingjian according to the core phenotypes, two heterozygous mutations in PMM2 gene had been found, c.430T > C in exon 5 (chr16:8905018 T > C) and c.640G > C in exon 8 (chr16:8941581G > C). These nucleotide substitutions would result in alterations in amino acid, F144L and G214R, respectively (**Figure 8**).

Further Sanger Sequencing result showed the proband's father is the heterozygous carrier of the c.430T > C mutation, while the proband's mother carries the c.640G > C mutation. The proband's sister with the same clinical presentation also carries all these two mutations. Thus, the proband is the compound heterozygous for the PMM2 p.F144L/p.G214R mutations (**Figure 9**).

Mutation p.F144L is a pathologic mutation that has been reported before. This mutation could create a new site for restriction enzyme SacI causing extra splicing (Kondo et al., 1999). Another mutation p.G214R has not been reported before, however, there is another reported disease-causing mutation at the same position (c. 640G > A, G214S) (Schollen et al., 2002; Vicario et al., 2017). Since this mutation is absent from controls (PM2), detected in trans with a pathogenic variant (PM3), located at the same position with a reported pathogenic missense change (PM5), this variant was classified as “likely pathogenic” according to ACMG guidelines (Richards et al., 2015). Prediction of this mutation by MutationTaster, Provan and SIFT also turned out to be disease causing (probability > 0.99), deleterious (score = −7.66) and damaging (score = 0), respectively. The result of MutationTaster (Schwarz et al., 2014) also indicated splice site change caused by the mutation (**Figure 10**), however mRNA experiment was not successfully performed to prove it.

Result of Other Patients

To assess the diagnostic rate of this method, “phenotype-driven virtual panel,” we decided to use the same method to analyze more neurological patients.

Clinical Information of the Patients

The clinical phenotypes of 29 patients were listed in **Table 4**.

Patients were collected from the neurology department of Beijing Children's Hospital. Of the 29 patients, 19 patients (65%) are male, 10 patients (35%) are female. The ages range from 4 months to 17 years 6 months. Most patients have an intellectual disability. More precise clinical information, phenotypes and gene sequencing result were available in **Supplementary Material**.

Sequencing Results of Patients

The gene sequencing results of these 29 patients was listed in **Table 5**.

DISCUSSION

Rare diseases, especially the ones involving multisystem are challenges for clinical diagnosis. For example, the PMM2 case described here involves not only the nervous system but also muscle, gonad, liver, spine, etc. It is hard to distinguish the fundamental factors of the pathogenesis by only examine clinical symptoms. Judging merely based on the clinical information, misdiagnosis was definitely not a rare event, especially in the generation without gene detection. A patient in our hospital who was previously diagnosed as Crouzon syndrome was finally proved to be Cytochrome P450 oxidoreductase deficiency by NGS (Hao et al., 2018). Misdiagnosis can result in a completely different treatment and might have possibility in leading deterioration. The efficacy of treatment might also be affected when the optimal treatment time is missed. Thus, gene sequencing is essential in the diagnosis of rare diseases.

TABLE 4 | Phenotype of 29 patients with neurological diseases.

Patient No.	Age range	Phenotypes
1	2–4	Gait instability, worse in the dark; dystonia
2	0–2	Intellectual disability; hearing abnormality; congenital cataract; talipes equinovarus; brain atrophy
3	>4	Intellectual disability; dark skin; abnormal facial shape
4	2–4	Global development delay; autism; optic atrophy; gait disturbance
5	0–2	Intellectual disability; abnormal facial shape; dystonia; muscle weakness
6	>4	Seizures; epileptic encephalopathy; intellectual disability; microcephaly
7	2–4	Intellectual disability; abnormal facial shape; autism
8	>4	Intellectual disability; short stature
9	0–2	Seizures; Intellectual disability; microcephaly; abnormal ear morphology
10	2–4	Seizures; Intellectual disability; abnormality of metabolism/homeostasis
11	0–2	Intellectual disability; seizures
12	0–2	Intellectual disability; hypertonia; esotropia; abnormality of metabolism/homeostasis
13	>4	Delayed gross motor development; Intellectual disability; agenesis of corpus callosum
14	>4	Intellectual disability; cryptorchidism; Short stature
15	2–4	Intellectual disability; autism
16	2–4	Intellectual disability; autism
17	0–2	Intellectual disability; elevated urine guanidinoacetic acid
18	2–4	Delayed gross motor development; Intellectual disability, ulnar claw
19	0–2	Seizures; global development delay; high palate
20	2–4	Seizures
21	>4	Intellectual disability; autism
22	>4	Seizures; glutaric aciduria
23	0–2	Seizures (VB6 improvement); intellectual disability; dyspepsia
24	0–2	Intellectual disability; microcephaly; abnormality of metabolism/homeostasis
25	0–2	Seizures; intellectual disability; vitamin B6 deficiency
26	2–4	Intellectual disability; cerebellar atrophy
27	2–4	Seizure; arachnoid cyst
28	0–2	Intellectual disability; autism
29	0–2	Multiple-malformation; dystonia

Core phenotypes of patients with the neurological inherited disease are similar, i.e., ataxia, seizures, esotropia, global developmental delay, puberty and gonadal disorders in this case. It is almost impossible to only rely on clinicians' experience to diagnose and determine candidate genes. Evaluating pathogenicity of the candidate mutations, confirming the gene function, excluding not associated mutations, choosing the clinically meaningful variants for Sanger Sequencing according to the similarity of

clinical presentation is the traditional way to annotate (Jin et al., 2018). However, it is unavoidable that the function and related diseases of the redundant phenotype-unrelated mutants will be analyzed. Here, the phenotype-driven designing “virtual panel” method could automatically filter the genes that is unrelated to the patient's symptoms, so that the analyser could only focus on the mutations in phenotype-related genes. This method can decrease the genes that should be analyzed, shorten the analysing time and make a more efficient annotation.

Moreover, designing traditional gene panel is a manual work, there might be bias occurring when selecting the gene list in the panel. Also, gene list in produced panel is constant, updating panel aligning with new discoveries is expensive and time-consuming. The virtual panel we run is designed by computer software “Mingjian,” which could avoid the bias due to personal cognition and judgement. In addition, “Mingjian” is according to the database of HPO, OMIM, and HGMD which includes all the known possible genes related to the phenotypes. Since it is actually “virtual,” updating the gene list is not an obstacle. Thus, it could contain all the present discovered, phenotype-related genes. Besides, all the undiagnosed cases can be re-analyzed when more disease-causing mutations are discovered and more linkages between disease and variations are established. Also, every patient has distinct phenotypes, a designed panel may not be applicable for every patient. Phenotype-driven “virtual panel” is based on the phenotypes of the patients, it may simply achieve low-cost individualized analysis when typical and standardized core phenotypes are extracted.

Consequently, we carried out this method in the diagnosis of more patients with neurological diseases to access the diagnostic rate. In 29 cases of patients, 21 of 29 patients were found carrying mutations in related genes. However, according to the inheritance pattern of genes, 2 heterozygous mutations of autosomal recessive genes were excluded. Other 19 of 29 patients were all confirmed with corresponding mutations by Sanger Sequencing.

For the rest of 10 patients who didn't confirm with the relevant mutations, it may fit one of the following conditions. First, the disease-causing mutations may locate in the undefined genes or genes that have not been experimentally proved to be associated with such neurological diseases. For example, we have found that NCAM1 polymorphisms is associated with autism in a previously undiagnosed case in year 2014 (Zhang et al., 2014). This kind of cases may be solved in the future due to development of research. Secondly, some mitochondrial gene mutations may also be involved but are outside the detection range of Whole Exome Sequencing. The symptoms of most mitochondrial diseases include seizures, mental retardation, developmental delay, metabolic disorders, muscle problems and visual disorders as well (Fang et al., 2017). Both mitochondrial DNA and nuclear DNA mutations may

TABLE 5 | Gene sequencing result of 29 patients with neurological diseases.

Patient No.	Gene	Position	Nucleotide variant	Protein variant	Inheritance pattern
1	COMP	chr19:18896846	c.1418A > G	p.D473G	AD
2	DYNC1H1	chr14:102499762	c.10354G > A	p.A3452T	AD
	CRYBB2	chr22:25627591	c.470C > G	p.P157R	
3	ADNP	chr20:49520469	c.64dupA	p.I22fs	AD
4	SYNE2	chr14:64675492	c.18218T > C	p.I6073T	AD
5	COL6A3	chr2:238245098	c.8645C > T	p. P2882L	AD/AR
		chr2:238275918	c.4912G > A	p.A1638T	
6	CHD2	chr15:93498742	c.1809G > T	p.K603N	AD
		chr15:93498743	c.1809+1G > T	Splicing	
7	HCFC1	chrX:153220254	c.3596G > C	p.R1199P	XR
8	SUCLG1	chr2:84652709	c.884G > A	p.V282I	AR
		chr2:84660557	c.592A > G	p.I198V	
9	SMARCAL1	chr2:217285085	c.926G > A	p.S309N	AR
		chr2:217332750	c.2225C > T	p.T742M	
10	MTHFR	chr1:11856378	c.665C > T	p.A222V	AR
		chr1:11863038	c.136C > T	p.R46W	
11	CDKL5	chrX:18582616	c.119C > T	p.A40V	XD
12	PDGFRB	chr5:149512504	c.936G > C	p.E312D	AD
13	TUBA1A	chr12:49579133	c.1016G > T	p.R339L	AD
14	SMC1A	chrX:53441721	c.331T > G	p.F111V	XD
15	HUWE1	chrX:53578276	c.9047A > C	p.E3016A	Unknown
	MAPT	chr17:44060834	c.664C > G	p.R222G	AD
16	MECP2	chrX:153296153	c.1162C > T	p.P388S	XD/XR
	KCNC1	chr11:17793707	c.1066G > A	p.V356M	AD
17	DYNC1H1 XDH	chr14:102463472	c.3665A > G	p.N1222S	AD AR
		chr2:31596756	c.1669G > A	p.D557N	
		chr2:31598377	c.1471G > A	p.A491T	
18	IGHMBP2	chr11:68702842	c.1708C > T	p.R570X	AR
		chr11:68704545	c.2598_2599del	p.K866Sfs	
19	CHD2	chr15:93563380	c.5045A > G	p.D1682G	AD
	CSF1R	chr5:149433641	c.2909_2910insATCA	p.Q970fs	AD
	EZH2	chr7:148544336	c.55G > A	p.V19I	AD
20	ND	ND	ND	ND	ND
21	ND	ND	ND	ND	ND
22	ND	ND	ND	ND	ND
23	ALDH7A1	chr5:125919644	c.454A > G	p.I152V	AR
24	ND	ND	ND	ND	ND
25	ND	ND	ND	ND	ND
26	SLC22A5	chr5:131728257	c.1400C > G	p.S467C	AR
27	ND	ND	ND	ND	ND
28	ND	ND	ND	ND	ND
29	ND	ND	ND	ND	ND

Twenty one of Twenty nine patients have been sequenced with a suspected gene, however, 2 of them have not corresponded with the inheritance pattern, i.e., autosome recessive gene with only one mutation. AD, autosomal dominant; AR, autosomal recessive; XD, X-linked dominant; XR, X-linked recessive; ND, not detected with related mutations.

contribute to dysfunction in mitochondria (Liu et al., 2014, 2015; Fang et al., 2017). Therefore, the disease-causing variants in these undiagnosed cases may be located in mitochondrial DNA. Moreover, insertion or deletion which is larger than 50 kb or chromosomal inversion may also cause disease. However, these mutations could not be identified by NGS due to technical limitations. This may not be a rare event since we previously diagnosed a novel DDC gene deletion in the

patients who was suspected to carry mutations in DDC gene but only diagnosed with single missense variant (Dai et al., 2018).

Overall, the diagnostic rate in this study was $19/29 = 65.52\%$, which far exceeds the known diagnostic rate of Whole-Exome Sequencing (25–30%). Therefore, the phenotype-driven virtual panel is an effective method to analyze WES data of neurological disease.

DATA AVAILABILITY STATEMENT

All the clinical and genetic data of the cases reported in this study have been submitted to the rare disease database, eRAM, at <http://www.unimd.org/eram/>.

ETHICS STATEMENT

This study was carried out is approved by Capital Medical University Beijing Children's Hospital Ethics Committee (Ethics Number: 2018-k-63). The protocol was approved by the Capital Medical University Beijing Children's Hospital Ethics Committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

CONSENT FOR PUBLICATION

The patient's parents gave written informed consent to studies and publication of clinical information, images and sequencing data.

REFERENCES

- Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., Gibbs, R. A., et al. (2010). A map of human genome variation from population-scale sequencing. *Nature* 467, 1061–1073. doi: 10.1038/nature09534
- Alfares, A., Aloraini, T., Subaie, L. A., Alissa, A., Qudsi, A. A., Alahmad, A., et al. (2018). Whole-genome sequencing offers additional but limited clinical utility compared with reanalysis of whole-exome sequencing. *Genet. Med.* 20, 1328–1333. doi: 10.1038/gim.2018.41
- Ashley, E. A., Butte, A. J., Wheeler, M. T., Chen, R., Klein, T. E., Dewey, F. E., et al. (2010). Clinical assessment incorporating a personal genome. *Lancet* 375, 1525–1535. doi: 10.1016/S0140-6736(10)60452-7
- Bai, D., Zhao, J., Li, L., Gao, J., and Wang, X. (2017). Analysis of genotypes and phenotypes in Chinese children with tuberous sclerosis complex. *Sci. China Life Sci.* 60, 763–771. doi: 10.1007/s11427-017-9091-x
- Bainbridge, M. N., Wisniewski, W., Murdock, D. R., Friedman, J., Gonzaga-Jauregui, C., Newsham, I., et al. (2011). Whole-genome sequencing for optimized patient management. *Sci. Transl. Med.* 3:87re3. doi: 10.1126/scitranslmed.3002243
- Biesecker, L. G., and Green, R. C. (2014). Diagnostic clinical genome and exome sequencing. *N. Engl. J. Med.* 370, 2418–2425. doi: 10.1056/NEJMra1312543
- Dai, L., Ding, C., and Fang, F. (2018). A novel Ddc gene deletion mutation in two Chinese mainland siblings with aromatic L-amino acid decarboxylase deficiency. *Brain Dev.* doi: 10.1016/j.braindev.2018.08.003 [Epub ahead of print].
- de Ligt, J., Boone, P. M., Pfundt, R., Vissers, L. E., Richmond, T., Geoghegan, J., et al. (2013). Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum. Mutat.* 34, 1439–1448. doi: 10.1002/humu.22387
- Dodge, J. A., Chigladze, T., Donadieu, J., Grossman, Z., Ramos, F., Serlicorni, A., et al. (2011). The importance of rare diseases: from the gene to society. *Arch. Dis. Child.* 96, 791–792. doi: 10.1136/adc.2010.193664
- Elliott, E., and Zuryski, Y. (2015). Rare diseases are a 'common' problem for clinicians. *Aust. Fam. Physician* 44, 630–633.
- Ewans, L. J., Schofield, D., Shrestha, R., Zhu, Y., Gayevskiy, V., Ying, K., et al. (2018). Whole-exome sequencing reanalysis at 12 months boosts diagnosis and is cost-effective when applied early in Mendelian disorders. *Genet. Med.* doi: 10.1038/gim.2018.39 [Epub ahead of print].
- Fang, F., Liu, Z., Fang, H., Wu, J., Shen, D., Sun, S., et al. (2017). The clinical and genetic characteristics in children with mitochondrial disease in China. *Sci. China Life Sci.* 60, 746–757. doi: 10.1007/s11427-017-9080-y

AUTHOR CONTRIBUTIONS

XW and FF designed the study. XW, FF, and C-HD collected the clinical data. XS, HZ, and Z-HC performed the WES. XS and D-YA analyzed the genetic data. XW, XS, and HZ wrote the manuscript. All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We are grateful to all of the family members for their participation in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.01529/full#supplementary-material>

- Feng, Y., Chen, D., and Wong, L.-J. C. (2017). "Detection of copy number variations (Cnvs) based on the coverage depth from the next generation sequencing data," in *Next Generation Sequencing Based Clinical Molecular Diagnosis of Human Genetic Disorders*, ed. L. J. Wong (Cham: Springer), 13–22.
- Gilissen, C., Hoischen, A., Brunner, H. G., and Veltman, J. A. (2012). Disease gene identification strategies for exome sequencing. *Eur. J. Hum. Genet.* 20, 490–497. doi: 10.1038/ejhg.2011.258
- Hao, C., Guo, J., Guo, R., Qi, Z., Li, W., and Ni, X. (2018). Compound heterozygous variants in Por gene identified by whole-exome sequencing in a Chinese pedigree with cytochrome P450 oxidoreductase deficiency. *Pediatr. Invest.* 2, 90–95. doi: 10.1002/ped4.12035
- Jin, Y., Zhang, L., Ning, B., Hong, H., Xiao, W., Tong, W., et al. (2018). Application of genome analysis strategies in the clinical testing for pediatric diseases. *Pediatr. Investig.* 2, 72–81. doi: 10.1002/ped4.12044
- Kondo, I., Mizugishi, K., Yoneda, Y., Hashimoto, T., Kuwajima, K., Yuasa, I., et al. (1999). Missense mutations in phosphomannomutase 2 gene in two Japanese families with carbohydrate-deficient glycoprotein syndrome type 1. *Clin. Genet.* 55, 50–54. doi: 10.1034/j.1399-0004.1999.550109.x
- Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., et al. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 312, 1880–1887. doi: 10.1001/jama.2014.14604
- Liu, P., Morrison, C., Wang, L., Xiong, D., Vedell, P., Cui, P., et al. (2012). Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* 33, 1270–1276. doi: 10.1093/carcin/bgs148
- Liu, Z., Fang, F., Ding, C., Wu, H., Lyu, J., and Wu, Y. (2014). [Sucla2-related encephalomyopathic mitochondrial Dna depletion syndrome: a case report and review of literature]. *Zhonghua Er Ke Za Zhi* 52, 817–821.
- Liu, Z., Fang, F., Ding, C., Zhang, W., Li, J., Yang, X., et al. (2015). [Diagnosis of mitochondrial disorders in children with next generation sequencing]. *Zhonghua Er Ke Za Zhi* 53, 747–753.
- Lupski, J. R., Reid, J. G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D. C., Nazareth, L., et al. (2010). Whole-genome sequencing in a patient with charcot-marie-tooth neuropathy. *N. Engl. J. Med.* 362, 1181–1191. doi: 10.1056/NEJMoa0908094
- Plaiasu, V., Nanu, M., and Matei, D. (2010). Rare Disease Day – at a glance. *Maedica (Buchar)* 5, 65–66.

- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30
- Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., Shannon, P. T., et al. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328, 636–639. doi: 10.1126/science.1186802
- Schieppati, A., Henter, J. I., Daina, E., and Aperia, A. (2008). Why rare diseases are an important medical and social issue. *Lancet* 371, 2039–2041. doi: 10.1016/S0140-6736(08)60872-7
- Schollen, E., Martens, K., Geuzens, E., and Matthijs, G. (2002). Dhplc analysis as a platform for molecular diagnosis of congenital disorders of glycosylation (CdG). *Eur. J. Hum. Genet.* 10, 643–648. doi: 10.1038/sj.ejhg.5200858
- Schwarz, J. M., Cooper, D. N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nat. Methods* 11, 361–362. doi: 10.1038/nmeth.2890
- Shashi, V., Mcconkie-Rosell, A., Schoch, K., Kasturi, V., Rehder, C., Jiang, Y. H., et al. (2016). Practical considerations in the clinical application of whole-exome sequencing. *Clin. Genet.* 89, 173–181. doi: 10.1111/cge.12569
- Sobreira, N. L., Cirulli, E. T., Avramopoulos, D., Wohler, E., Oswald, G. L., Stevens, E. L., et al. (2010). Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. *PLoS Genet.* 6:e1000991. doi: 10.1371/journal.pgen.1000991
- Speicher, M. R., and Carter, N. P. (2005). The new cytogenetics: blurring the boundaries with molecular biology. *Nat. Rev. Genet.* 6, 782–792. doi: 10.1038/nrg1692
- Stolk, P., Willemsen, M. J., and Leufkens, H. G. (2006). Rare essentials: drugs for rare diseases as essential medicines. *Bull. World Health Organ.* 84, 745–751. doi: 10.2471/BLT.06.031518
- Tabor, H. K., Risch, N. J., and Myers, R. M. (2002). Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat. Rev. Genet.* 3, 391–397. doi: 10.1038/nrg796
- Vicario, M., Cali, T., Cieri, D., Vallesse, F., Bortolotto, R., Lopreiato, R., et al. (2017). A novel Pmca3 mutation in an ataxic patient with hypomorphic phosphomannomutase 2 (Pmm2) heterozygote mutations: biochemical characterization of the pump defect. *Biochim. Biophys. Acta* 1863, 3303–3312. doi: 10.1016/j.bbdis.2017.08.006
- Wenger, A. M., Guturu, H., Bernstein, J. A., and Bejerano, G. (2017). Systematic reanalysis of clinical exome data yields additional diagnoses: implications for providers. *Genet. Med.* 19, 209–214. doi: 10.1038/gim.2016.88
- Wright, C. F., Fitzpatrick, D. R., and Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nat. Rev. Genet.* 19, 253–268. doi: 10.1038/nrg.2017.116
- Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* 369, 1502–1511. doi: 10.1056/NEJMoa1306555
- Zhang, J., Wang, A., Li, Y., Lu, X., Wang, F., and Fang, F. (2014). Association of Ncam1 polymorphisms with autism and parental age at conception in a Chinese Han population. *Genet. Test. Mol. Biomarkers* 18, 690–694. doi: 10.1089/gtmb.2014.0055

Conflict of Interest Statement: XS, HZ, Z-HC, and D-YA were employed by company Running Gene Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Shen, Fang, Ding, Zhang, Cao and An. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Growth Concerns in Coffin–Lowry Syndrome: A Case Report and Literature Review

Ying Lv¹, Liuyan Zhu¹, Jing Zheng², Dingwen Wu² and Jie Shao^{1*}

¹ Department of Pediatric Health Care, The Children's Hospital, Zhejiang University School of Medicine, Hangzhou, China,

² Department of Gene Screening Laboratory, The Children's Hospital, Zhejiang University School of Medicine, Hangzhou, China

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Bohu Pan,
National Center for Toxicological
Research (FDA), United States
Zhuopei Hu,
University of Arkansas for Medical
Sciences, United States

*Correspondence:

Jie Shao
shaojie@zju.edu.cn

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Pediatrics

Received: 09 September 2018

Accepted: 24 December 2018

Published: 25 January 2019

Citation:

Lv Y, Zhu L, Zheng J, Wu D and
Shao J (2019) Growth Concerns in
Coffin–Lowry Syndrome: A Case
Report and Literature Review.
Front. Pediatr. 6:430.
doi: 10.3389/fped.2018.00430

Mutation of RPS6KA3 can induce Coffin–Lowry syndrome, an X-linked syndrome. The case here reported manifests its signature characteristic of short stature, facial dysmorphism, development retardation, hearing defect. The mutation of RPS6KA3 we detected by NGS analysis is c.2185 C > T. The short stature is a noteworthy problem we discuss here to improve the patient's growth and development. The efficacy and safety of application of growth hormone analogs on patients with CLS are not confirmed and need to be carefully considered.

Keywords: RSK2, Coffin–Lowry, growth retardation, pervasive development disorder, growth hormone

BACKGROUND

A growing number of studies have demonstrated that RPS6KA3 is a molecular etiology of Coffin–Lowry syndrome (CLS) (1, 2), an X-linked semidominant syndrome which was first reported by Coffin in 1966 and characterized by short stature, facial dysmorphism, severe-to-profound intellectual disability (ID), motor developmental delay, progressive skeletal deformities, dental disorders, hearing defect and other congenital deformity in males, with intellect ranges from normal to severely impaired in heterozygous females (3–6).

The protein responsible for RPS6KA3 is the ribosomal protein S6 kinase polypeptide 3 (RSK2), which is a serine/threonine kinase of a family of mitogen-activated protein kinases (4). The RSK2 protein is composed of two functional kinase domains that are activated in a sequential manner by a series of phosphorylations, such as PDK docking site (4). RSK2 has been found to be related with prompting cell proliferation, blocking cell differentiation and protecting cells from apoptosis (7, 8). To date, there have been reports of cases with different symptoms and with over 128 distinct mutations in the RSK2 gene split into 22 exons of chrXp22.2 (RPS6KA3) (4, 5). Recently, the mutations of RPS6KA3 have been found to be related with heart disease, osteosarcoma, foramen magnum compression, Drop episodes and so on (7, 9–12).

Worldwide, the speculated prevalence of CLS may be 1/50,000–1/100,000 and about 70–80% of probands have no family history while 20–30% have more than one affected family member as reported (13). As yet, there is no cure for this pathology. Here, we report a mutation of RPS6KA3 at Xp22 in a Chinese boy. The phenotypic appearance was typical, including growth and development retardation, as was described in the previous reports.

CASE PRESENTATION

The proband is a 12-months-old boy with typical facial dysmorphism, hearing defect and bony abnormality (**Table 1**). He was born after a normal pregnancy and delivered with birth weight of 2.9 kg (10th percentile) and birth length of 45 cm (3rd percentile) at 38 weeks, compared with the WHO Child Growth Standards in 2006. The facial appearance presents bulging forehead, prominent ears, widely spaced eyes, down slanted palpebral fissures, short nose with broad columella, thick

alae nasi and septum, thick and everted underlip (**Figures 1A,B**). The deciduous teeth were erupted at 8 months old (not delayed) (**Figure 1C**). The hands are short, fleshy, and with remarkably hyperextensible fingers that taper from wide to narrow with small terminal phalanges and nails (**Figures 1D, 2D**). But there was no deformity of his foramen magnum or spine column (**Figures 2B,C**). The weight at 12 months is 8.2 kg and height is 68.2 cm (<-3.17 z score, WHO). The bone metabolism and IGF-1 α is disturbance (Vit D 45.2 nmol/L, IGF-1 α < 25 ng/mL). He started sitting alone at 9 months and couldn't stand unaided until 12 months of age. At 12 months of age, his intelligence quotient (IQ) was 56 according to the Gesell Developmental Schedules. He had difficulty remaining seated or concentrating during task completion. His auditory threshold of auditory brainstem response (ABR) is >85 db and is diagnosed as a hearing disorder. The magnetic resonance imaging (MRI) showed the dilation of bilateral ventricles and less cerebral white matter (**Figure 2A**).

For genetic analysis, blood samples were obtained from the individual. The mother had given informed consent for her children. This research was approved by the bioethics committee for human gene analysis at the Zhejiang University.

TABLE 1 | The typical phenotype of Coffin–Lowry syndrome of proband.

1. Typical facial dysmorphism: bulging forehead, prominent ears, widely spaced eyes, down slanted palpebral fissures, short nose, everted underlip
2. Hearing defect: >85 db (both ears)
3. Hyperextensible fingers that taper from wide to narrow with small terminal phalanges and nails
4. Short stature: 68.2 cm at 1 year old
5. Mental retardation: IQ was 56
6. Hypoevolutism in motor development
7. Bone metabolism disturbance: Vit D 45.2 nmol/L
8. IGF-1 α < 25 ng/mL



FIGURE 1 | The distinctive facial dysmorphisms of a Coffin–Lowry syndrome boy. A side view of the patient's face (**B**), prominent ears, widely spaced eyes, down slanted palpebral fissures, short nose with broad columella, thick alae nasi and septum, thick and everted underlip (**A,C**). (**D**) The hands are short and with puffy tapered finger, small terminal phalanges, and nails.

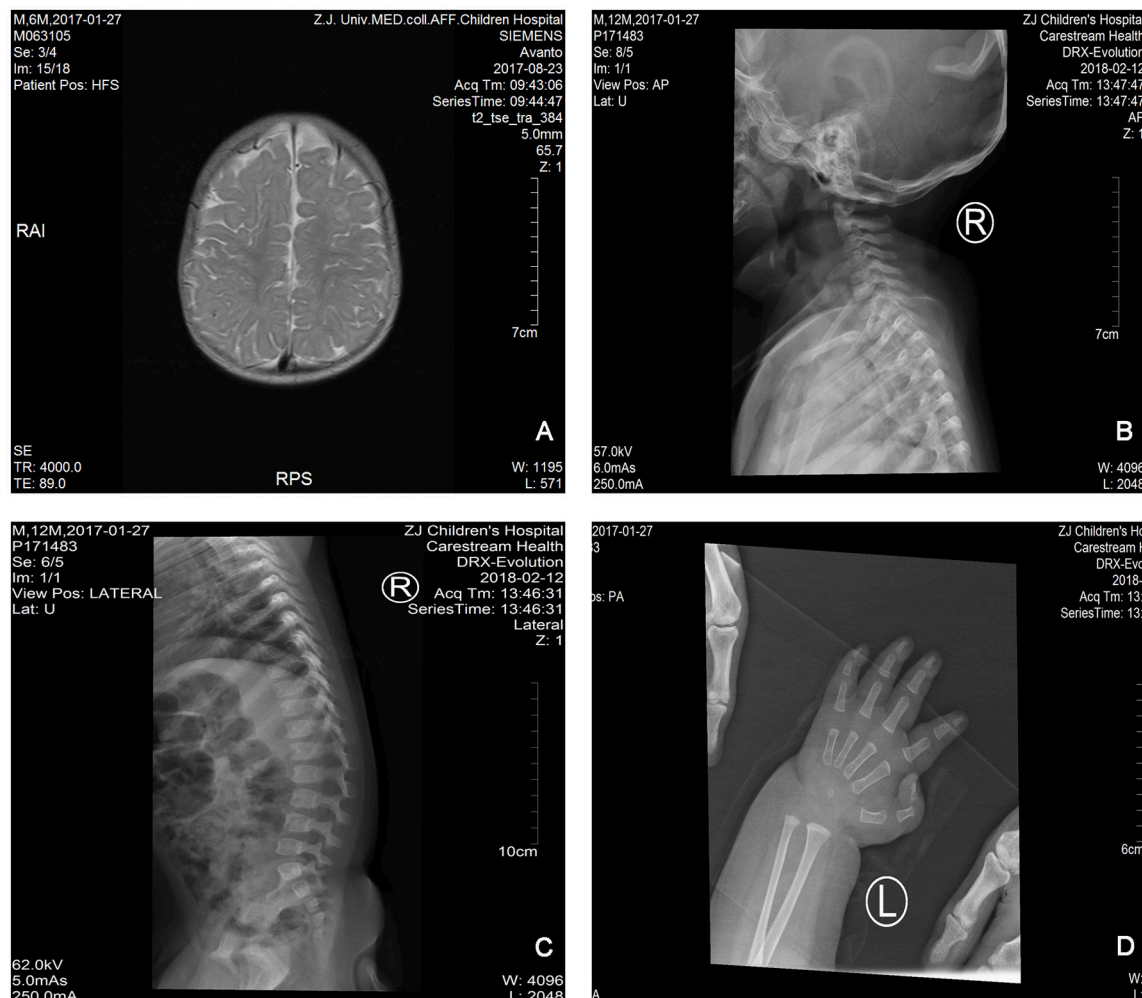


FIGURE 2 | The magnetic resonance imaging (MRI) and X-ray images of the Coffin–Lowry syndrome case. **(A)** The MRI of cerebral development; **(B,C)** Cervical and spinal canal X-ray photographs. **(D)** Nodular hyperplasia of finger tail end.

Next Generation Sequencing (NGS) analysis was performed using Agilent Human Genome panel (Agilent Technologies, Inc, Santa Clara, CA, USA). A c.2185 C > T at chrX20173554 was detected in the proband, which can cause p.Arg729Trp mutation and RSK2 instability (**Figure 3**). Subsequent analysis did not detect the same point mutation in other family members including mother, brother (**Supplementary Material**). This indicated that the micromutation is new but not inherited from the mother.

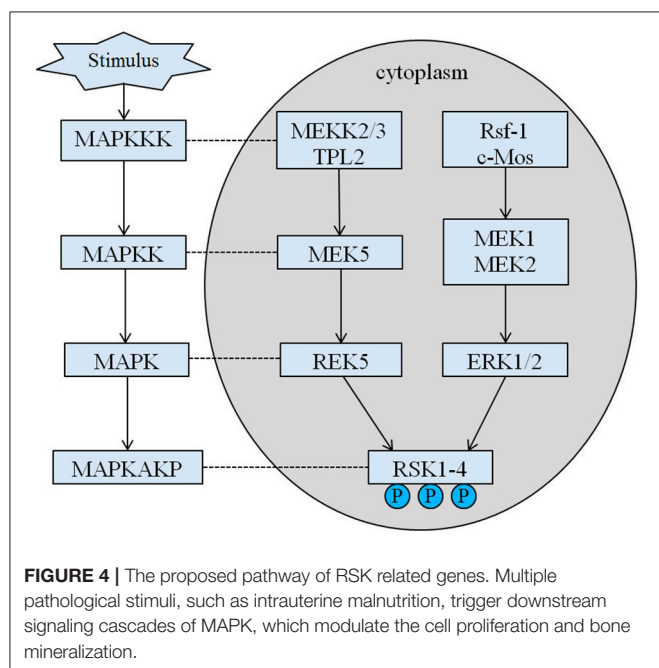
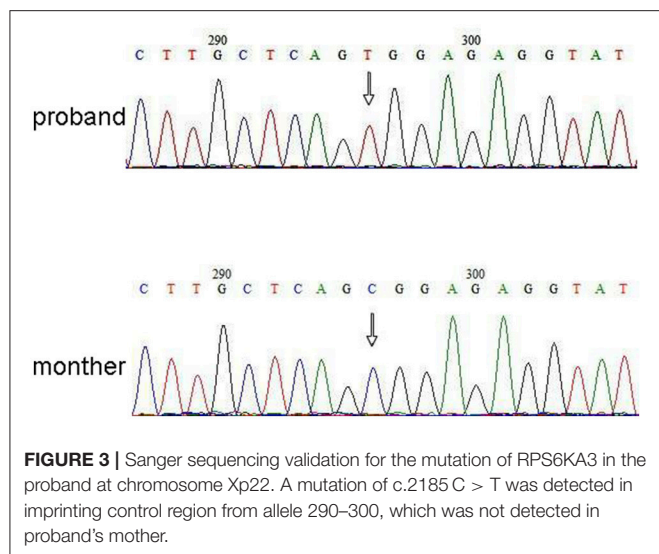
DISCUSSION

The case we present for genetics evaluation is a 12 months boy who was born after an uneventful pregnancy from healthy parents. However, his developmental age was delayed. Like other affected CLS patients, he had the typical phenotype observed in CLS including intellectual disability, retarded motor development, small stature, tapering fingers, hearing defect and characteristic facial features.

RPS6KA3 (OMIM 303600) is known to be mutated in patients with Coffin–Lowry syndrome. Exome c.2185 C > T variant at Xp22 was novel, segregated from the disease, and was predicted to be damaging using five different *in silico* software (SIFT, Polyphen, MutationTaster, FATHMM, LRT).

The inherited height of our proband is 167.5 cm because of his father's height is 173 cm while his mother's height is 150 cm (14). Whether the short stature of Coffin–Lowry syndrome could be treated with growth hormone analogs is still not clear. RSKs are serine/threonine kinases activated by ERK/MAPK and constitute four isoforms (RSK1–4) (15). RSKs act as downstream effectors of RTK/Ras/ERK signaling (**Figure 4**) (15–17). As the gene responsible for CLS, RSK2 play an important role in cell proliferation and migration. Full activation of RSK2 requires phosphorylation at multiple sites. Growth factors can stimulate ERK signal pathway to phosphorylated RSK2. As Ramos reported, ERK/RSK2 activating is involved in regulating cancer cell migrating (18).

What is more, RSK2 out of function is related with bone mineralization abnormal and also induces spine malformations



and calcifications of ligamentum flavum, such as thoracic lordosis, scoliosis, kyphosis, and degenerative disc disease (19, 20). Previous reports identified the ERK/ RSK2 as a regulator of bone formation *in vivo* (7). Indeed, as Marques et al. reported, RSK2-deficient mice recapitulated the progressive bone loss due to decreased bone mineralization by the osteoblasts observed in patients with CLS (21). Although the differentiation of these

cells *in vitro* was drastically blocked, the *in vivo* phenotype was clearly attributed to a cell autonomous decrease in the activity of the osteoblasts rather than a decrease in their numbers (7). Interestingly, RSK2 was recently shown to interact with TNF-RI to affected osteoblasts differentiation (18, 22). As we know, growth hormone can cause increased bone mineral density. The progress may aggravate calcifications of ligamentum flavum and skeletal deformity (23).

CONCLUDING REMARKS

The present study is the first to report a Chinese case of CLS with mutation of RPS6KA3, as well as distinctly growth retardation, multiple facial abnormalities, intellectual and motor disabilities. We speculate that the application of growth hormone analogs may increase the incidence and invasiveness of cancers as well as the risk of spine malformation. If its efficacy and safety are not confirmed, the application of growth hormone analogs on patients of CLS should be especially cautiously considered. Long-term studies with relevant outcome will be essential in the future clinical research.

PATIENT CONSENT

We obtained informed consent for publishing this case report and for using related images from patient's parents.

AUTHOR CONTRIBUTIONS

JS substantially contributed to the conception of this manuscript. YL, LZ, JZ, and DW contributed to the acquisition, analysis, or interpretation of data. YL drafted the manuscript. JS critically revised the manuscript for important intellectual content.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (no. 81501293 and 81773440). We acknowledge the family of the case described in this report that is aware of and approve this de-identified account.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2018.00430/full#supplementary-material>

Figure S1 | Sanger sequencing validation for the mutation of RPS6KA3 in the proband's brother.

REFERENCES

1. Stornetta RL, Zhu JJ. Ras and Rap signaling in synaptic plasticity and mental disorders. *Neuroscientist* (2011) 17:54–78. doi: 10.1177/1073858410365562
2. Jurkiewicz D, Jezela-Stanek A, Ciara E, Piekutowska-Abramczuk D, Kugajda M, Gajdulewicz M, et al. Four novel RSK2 mutations in

females with Coffin–Lowry syndrome. *Eur J Med Genet.* (2010) 53:268–73. doi: 10.1016/j.ejmg.2010.07.006

3. Abidi F, Jacquot S, Lassiter C, Trivier E, Hanauer A, Schwartz CE. Novel mutations in Rsk-2, the gene for Coffin–Lowry syndrome (CLS). *Eur J Hum Genet.* (1999) 7:20–26. doi: 10.1038/sj.ejhg.5200231

4. Frodin M, Gammeltoft S. Role and regulation of 90 kDa ribosomal S6 kinase (RSK) in signal transduction. *Mol Cell Endocrinol.* (1999) 151:65–77. doi: 10.1016/S0303-7207(99)00061-1
5. Delaunoy JP, Dubos A, Marques Pereira P, Hanauer A. Identification of novel mutations in the RSK2 gene (RPS6KA3) in patients with Coffin–Lowry syndrome. *Clin Genet.* (2006) 70:161–6. doi: 10.1111/j.1399-0004.2006.00660.x
6. Knish LV. Effect of vitamin E preparations used in pregnancy on the placenta. *Pediatr Akus Ginekol.* (1966) 2:51–5.
7. David JP, Mehic D, Bakiri L, Schilling AF, Mandic V, Priemel M, et al. Essential role of RSK2 in c-Fos-dependent osteosarcoma development. *J Clin Invest.* (2005) 115:664–72. doi: 10.1172/JCI200522877
8. Yang X, Matsuda K, Bialek P, Jacquot S, Masuoka HC, Schinke T, et al. ATF4 is a substrate of RSK2 and an essential regulator of osteoblast biology; implication for Coffin–Lowry Syndrome. *Cell* (2004) 117:387–98. doi: 10.1016/S0092-8674(04)00344-7
9. Koehne T, Jeschke A, Petermann F, Seitz S, Neven M, Peters S, et al. Rsk2, the kinase mutated in coffin-lowry syndrome, controls cementum formation. *J Dent Res.* (2016) 95:752–60. doi: 10.1177/0022034516634329
10. Gschwind M, Foletti G, Baumer A, Bottani A, Novy J. Recurrent nonconvulsive status epilepticus in a patient with coffin-lowry syndrome. *Mol Syndromol.* (2015) 6:91–5. doi: 10.1159/000430429
11. Stollberger C, Finsterer J. Consider the neuro-cardiac continuum of Coffin–Lowry syndrome! *Am J Med Genet A* (2015) 167A:1954–5. doi: 10.1002/ajmg.a.37099
12. Martinez HR, Niu MC, Sutton VR, Pignatelli R, Vatta M, Jefferies JL. Coffin–Lowry syndrome and left ventricular noncompaction cardiomyopathy with a restrictive pattern. *Am J Med Genet A* (2011) 155A:3030–4. doi: 10.1002/ajmg.a.33856
13. Delaunoy J, Abidi F, Zeniou M, Jacquot S, Merienne K, Pannetier S, et al. Mutations in the X-linked RSK2 gene (RPS6KA3) in patients with Coffin–Lowry syndrome. *Hum Mutat.* (2001) 17:103–16. doi: 10.1002/1098-1004(200102)17:2<103::AID-HUMU2>3.0.CO;2-N
14. Cameron N. BASIC programs for the assessment of skeletal maturity and the prediction of adult height using the Tanner–Whitehouse method. *Ann Hum Biol.* (1984) 11:261–4. doi: 10.1080/03014468400007151
15. Romeo Y, Zhang X, Roux PP. Regulation and function of the RSK family of protein kinases. *Biochem J.* (2012) 441:553–69. doi: 10.1042/BJ20110289
16. Kim M, Lee JH, Koh H, Lee SY, Jang C, Chung CJ, et al. Inhibition of ERK–MAP kinase signaling by RSK during Drosophila development. *EMBO J.* (2006) 25:3056–67. doi: 10.1038/sj.emboj.7601180
17. Beck K, Ehmann N, Andlauer TF, Ljaschenko D, Strecker K, Fischer M, et al. Loss of the Coffin–Lowry syndrome-associated gene RSK2 alters ERK activity, synaptic function and axonal transport in Drosophila motoneurons. *Dis Model Mech.* (2015) 8:1389–400. doi: 10.1242/dmm.021246
18. Shi GX, Yang WS, Jin L, Matter ML, Ramos JW. RSK2 drives cell motility by serine phosphorylation of LARG and activation of Rho GTPases. *Proc Natl Acad Sci USA.* (2018) 115:E190–9. doi: 10.1073/pnas.1708584115
19. Herrera-Soto JA, Santiago-Cornier A, Segal LS, Ramirez N, Tamai J. The musculoskeletal manifestations of the Coffin–Lowry syndrome. *J Pediatr Orthop.* (2007) 27:85–9. doi: 10.1097/01.bpo.0000187994.94515.9d
20. Welborn M, Farrell S, Knott P, Mayekar E, Mardjetko S. The natural history of spinal deformity in patients with Coffin–Lowry syndrome. *J Child Orthop.* (2018) 12:70–5. doi: 10.1302/1863-2548.12.170101
21. Pereira PM, Schneider A, Pannetier S, Heron D, Hanauer A. Coffin–Lowry syndrome. *Eur J Hum Genet.* (2010) 18:627–33. doi: 10.1038/ejhg.2009.189
22. Bohm C, Derer A, Axmann R, Hillienhoff U, Zaiss MM, Luther J, et al. RSK2 protects mice against TNF-induced bone loss. *J Cell Sci.* (2012) 125 (Pt. 9):2160–71. doi: 10.1242/jcs.096008
23. Tritos NA, Klibanski A. Effects of growth hormone on bone. *Prog Mol Biol Transl Sci.* (2016) 138:193–211. doi: 10.1016/bs.pmbts.2015.10.008

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer BP and handling Editor declared their shared affiliation at the time of review.

Copyright © 2019 Lv, Zhu, Zheng, Wu and Shao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Two Novel AGXT Mutations Cause the Infantile Form of Primary Hyperoxaluria Type I in a Chinese Family: Research on Missed Mutation

Xiulan Lu^{1,2†}, Weijian Chen^{1,3†}, Liping Li¹, Xinyuan Zhu¹, Caizhi Huang¹, Saijun Liu⁴, Yongjia Yang^{1*} and Yaowang Zhao^{1,5*}

¹ The Laboratory of Genetics and Metabolism, Hunan Children's Research Institute (HCRI), Hunan Children's Hospital, University of South China, Changsha, China, ² Pediatric Intensive Care Unit, Hunan Children's Hospital, University of South China, Changsha, China, ³ Department of Pathology, Hunan Children's Hospital, University of South China, Changsha, China, ⁴ BGI-China, Shenzhen, China, ⁵ Department of Urinary Surgery, Hunan Children's Hospital, University of South China, Changsha, China

OPEN ACCESS

Edited by:

Tielu Shi,

East China Normal University, China

Reviewed by:

Andre Laval Samson,
Walter and Eliza Hall Institute
of Medical Research, Australia
Sudheer Kumar Ravuri,
Steadman Philippon Research
Institute, United States
Qi Liu,
Tongji University, China

*Correspondence:

Yongjia Yang
yongjia727@aliyun.com
Yaowang Zhao
yw508@sina.com

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 06 August 2018

Accepted: 21 January 2019

Published: 06 February 2019

Citation:

Lu X, Chen W, Li L, Zhu X,
Huang C, Liu S, Yang Y and Zhao Y
(2019) Two Novel AGXT Mutations
Cause the Infantile Form of Primary
Hyperoxaluria Type I in a Chinese
Family: Research on Missed Mutation.
Front. Pharmacol. 10:85.
doi: 10.3389/fphar.2019.00085

Primary hyperoxaluria type 1 (PH1) is a rare metabolic disorder characterized by a defect in the liver-specific peroxisomal enzyme alanine-glyoxylate and serine-pyruvate aminotransferase (AGT). This disorder results in hyperoxaluria, recurrent urolithiasis, and nephrocalcinosis. Three forms of PH1 have been reported. Data on the infantile form of PH1 are currently limited in literature. Despite the fact that China is the most populated country in the world, only a few AGXT mutations have been reported in several Chinese PH1 patients. In the present study, we investigated a Chinese family in which two siblings are affected by the infantile form of PH1. Sanger sequencing was carried out on the proband, but the results were misleading. Two novel missense mutations (c.517T > C/p.Cys173Arg and c.667A > C/p.Ser223Arg) of the AGXT gene were successfully detected through whole-exome sequencing. These two mutations occurred in the highly conserved residues of the AGT. Four software programs predicted both mutations as the cause of the disease. A postmortem examination was performed and revealed the occurrence of global nephrocalcinosis on both kidneys. The crystals were collected and analyzed as calcium oxalate monohydrate. This study extends the knowledge on the clinical phenotype-genotype correlation of the AGXT mutation. That is, (i) two novel missense mutations were identified for the infantile form of PH1 and (ii) the same AGXT genotype caused the same infantile form of PH1 within the family.

Keywords: PH1 infantile type, AGXT, mutation, whole exome sequencing, nephrocalcinosis

INTRODUCTION

Rare diseases burden societies and families (Jia and Shi, 2017). The natural history of a rare disease is often poorly characterized owing to insufficient knowledge, misdiagnoses, missed diagnoses, and incurability (Ni and Shi, 2017). Primary hyperoxaluria type 1 (PH1) is an inherited metabolic disorder caused by alanine-glyoxylate and serine-pyruvate aminotransferase (AGT) deficiency

(Coulter-Mackie et al., 1993–2017). AGT is a hepatic peroxisomal enzyme that catalyzes the conversion of glyoxylate to glycine. When AGT activity is deficient, glyoxylate is converted to oxalate, which then forms insoluble calcium salt deposits in the kidney and other organs (Williams et al., 2009). Human PH1 can be classified into the following forms according to clinical severity: the infantile form with early renal insufficiency, the late-onset form with a good prognosis, and the most common form with recurrent urolithiasis or nephrocalcinosis that is usually accompanied by renal insufficiency (Jellouli et al., 2016). The infantile form is the most severe (Jellouli et al., 2016). The genetic basis for PH1 has been identified in at least 190 published *AGXT* gene (coding AGT) mutations (Coulter-Mackie et al., 1993–2017; Isiyel et al., 2016). However, data on the infantile form of PH1 are limited (less than 1/5 of PH1 patients have the infantile form) (Kurt-Sukur et al., 2015; Jellouli et al., 2016). Families with the infantile form of PH1 usually include only one affected patient (van Woerden et al., 2004; Kurt-Sukur et al., 2015; Jellouli et al., 2016; Cui et al., 2017; Soliman et al., 2017). This rate of occurrence makes the delineation of the clinical phenotype–genotype correlation of the *AGXT* mutation–infantile form of PH1 difficult. Meanwhile, China is the most populated country in the world, but cases of *AGXT* mutations in the Chinese population are rarely reported (Yuen et al., 2004; Li et al., 2014; Wang et al., 2016; Cui et al., 2017). The characteristics of the *AGXT* mutation spectrum in the Chinese population remain unclear. In this study, we investigated a Chinese family in which two siblings carried the c.517T > C/p.Cys173Arg and c.667A > C/p.Ser223Arg mutations of the *AGXT* gene. We discovered that the same *AGXT* genotype causes the same infantile form of PH1 within a family.

RESULTS

Clinical Data

Patient 1

A boy, aged 4 months and 7 days (Subject 16, **Figure 1A**) and suffering from recurrent diarrhea (7–8 times per day) of unknown etiology, was referred to our hospital. The boy was born in Central China (Hunan Province, Han Chinese) and was the first child of non-consanguineous parents. His birth weight was 3050 g after full-term gestation without any medical problem. When the boy was admitted to our hospital, his rectal temperature was 36.5°C, blood pressure was 130/90 mmHg, pulse rate was 163 beats/min, and breathing rate was 8 breaths/min. He had severe hyponatremia, metabolic acidosis, and anemia (**Table 1**). His urine analysis results showed proteinuria (**Supplementary Table S1**), and his renal ultrasonography revealed that both his kidneys were small and exhibited mildly increased echogenicity. The patient progressed rapidly to end-stage renal disease (ESRD) at the age of 4 months and 12 days. The patient died at 4 months and 17 days.

Patient 2

This child is the younger brother of patient 1 (Subject 17, **Figure 1A**). The boy also suffered from recurrent diarrhea (5–6

times per day) at the age of 4 months and 6 days. His birth weight was 3400 g at full-term gestation and presented no medical problem. His laboratory test results indicated abnormal liver and renal functions and anemia (**Table 1**). Urine analysis results showed proteinuria (**Supplementary Table S1**). ESRD developed at the age of 4 months and 15 days for the patient. Blood dialysis therapy was performed for 28 days. The patient died at the age of 5 months and 5 days. In brief, all the clinical data of the two patients were standardized and deposited into eRAM (Jia et al., 2018).

First Round of Sanger Sequencing of the *AGXT* Gene

Given the clinical findings with unknown ESRD, we suspected the occurrence of a metabolic disease. After examining the proband affected by PH1 (June 2008), we synthesized the primers for all the exons and exon–intron boundaries of the *AGXT* gene (**Supplementary Table S2**). Sanger sequencing was routinely performed on the coding regions and the exon–intron boundaries of the *AGXT* gene (NM_000030) in subjects 16, 9, and 10. We detected a missense mutation on exon 4 (c.517T > C/p.Cys173Arg; **Figure 1B**) and validated the mutation by Sanger sequencing using reverse primer. The mutation originated from the patients' healthy mother (**Figure 1A**). At this Cys173 amino acid position, the p.Cys173Tyr mutation associated with severely reduced catalytic activity was previously reported on a PH1 patient (Von Schnakenburg and Rumsby, 1998). Human PH1 is clearly caused by *AGXT* recessive mutation. However, no other pathogenic variant of the *AGXT* gene was detected in the trios-family despite the successful amplification and sequencing of all the exons and boundaries of *AGXT*. The PH1 diagnosis in the family was stalled because of another mutated *AGXT* allele, which was not detected in the family.

Next-Generation Sequencing for Four Family Members

On December 2016, subject 17 was admitted in our hospital because of clinical symptoms similar to those in subject 16. The high phenotypic similarity of the two infants prompted us to conduct next-generation sequencing on the family (**Figure 1A**; whole-exome sequencing on subjects 9, 10, 16, and 17). We envisage that next generation sequencing may generate new data for the diagnosis of the disease in the family (Shen, 2018). We hypothesized that a recessive metabolic disorder affected the family. Considering non-consanguineous mating, we mainly focused on the compound heterozygous mutations in the family. Common variants in public databases (the variant was neither found in ExAC nor 1000 Genomes) and in-house databases were filtered out. Given that both patients died of ESRD, we focused on the genes involved in renal disorders (**Supplementary Table S3**). We detected the two heterozygous variants (chr2:241810859T > C and chr2:241813466A > C) of the *AGXT* gene segregated with PH1 in the family (**Figure 1** and **Supplementary Figure S1**). Both *AGXT* variants were missense

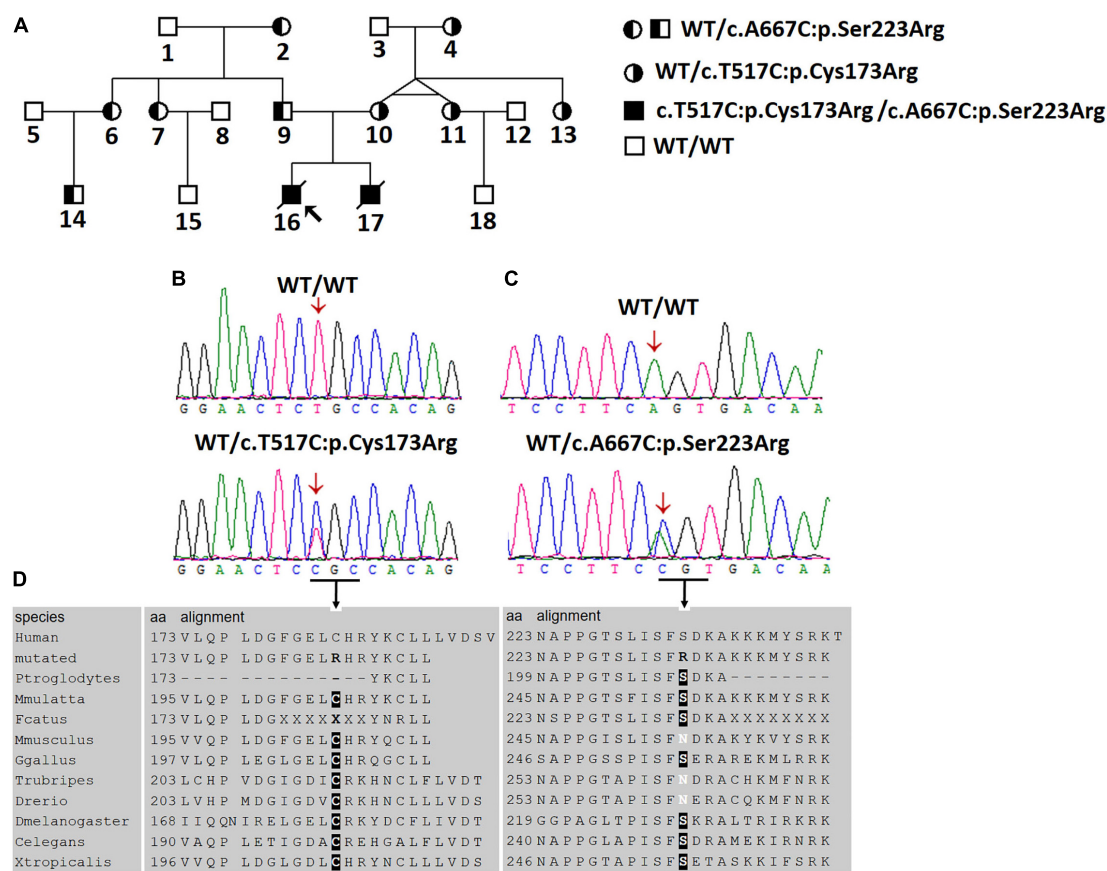


FIGURE 1 | Two novel AGXT mutations in a Chinese family with the infantile form of primary hyperoxaluria type I (PH1). **(A)** Investigated Chinese family with PH1. Subjects 16 and 17 suffered from the infantile form PH1. Subjects 4, 10, 11, and 13 were affected by benign urolithiasis at least once, and other family members were asymptomatic. **(B,C)** Sanger sequencing revealed the AGXT gene mutations, c.517T > C/p.Cys173Arg and c.667A > C/p.Ser223Arg. **(D)** Both mutations occurred on the conserved residues of the AGT protein.

variants and caused the disease according to the reports of four prediction software programs (Table 2).

The variant chr2:241810859T > C was identified in exon 4 of the AGXT gene and harbored a change in residue 173 from cysteine to arginine (Figure 1B). The mutation of AGXT-exon 4 (chr2:241810859T > C) was detected as mentioned above. By contrast, variant chr2:241813466A > C was identified in exon 6 of the AGXT gene and harbored a change in residue 223 from serine to arginine. The AGXT-exon 6 mutation (chr2:241813466A > C) was missed in the first round of Sanger sequencing (mentioned above; detailed data are presented in the Discussion section). After re-synthesis, a new pair of primers (2AGXT6F: 5'-CATCTCCCCTGCTATCGTGTAC-3'; 2AGXT6R: 5'-CCTCAGTCCTTCTGGTCAC-3', predicted PCR product size: 498 bp) was used for AGXT-exon 6, and the c.667A > C/p.Ser223Arg mutation (Figure 1C) was detected in the second round of Sanger sequencing.

AGT-Ma and AGT-Mi Detection

The wild-type AGXT gene can carry two polymorphic variants: the rs4426527 (p.Ile340Met) and the less frequent minor

haplotype (rs34116584, p.Pro11Leu). In this study, we utilized the results of next-generation sequencing (Supplementary Figure S1). However, we discovered that none of the family members carried any of the two variants. Therefore, both mutations are on the major haplotype.

Postmortem and Nephrocalcinosis

Subject 17 died of ESRD, and a postmortem examination was performed. No visible stone was observed inside both kidneys of the patient (Figures 2A,B). However, global nephrocalcinosis occurred in both kidneys, as observed under a polariscope (Figure 3). We obtained 5 g of the kidney tissue and placed the sample in a hot air oven for 2 h (Figures 4A,B). Crystals were collected for composition analysis (infrared spectroscopy test) and were identified as calcium oxalate monohydrate (Figure 4C).

DISCUSSION

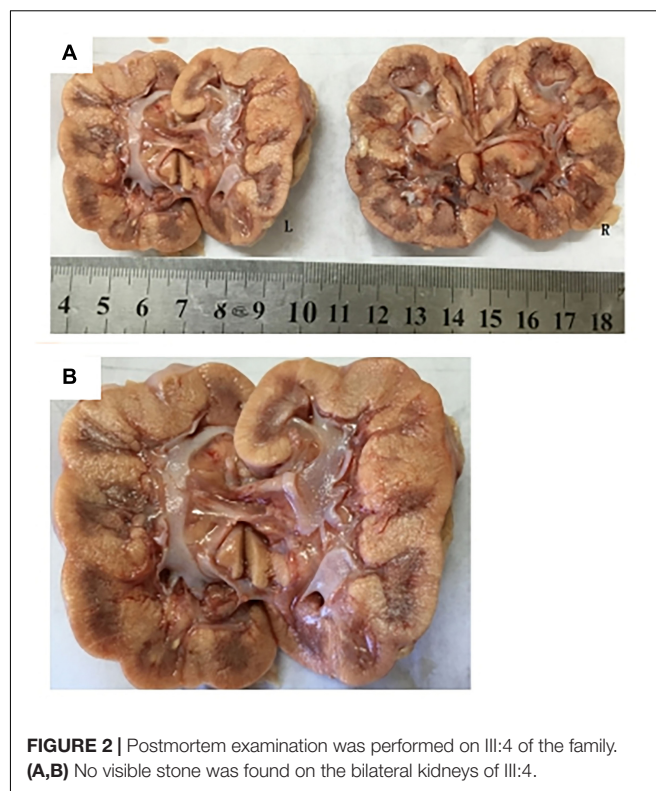
Human AGT protein is a fold-type I pyridoxal 5'-phosphate (PLP)-dependent enzyme. The AGT molecule has a homodimeric structure of 2 × 43 kD/392 amino acids

TABLE 1 | Biochemical features of two patients with infantile form of primary hyperoxaluria type 1.

		III:3	III:4	Normal range
Blood test + EPO	WBC ($\times 10^9/L$)	6.11	3.4	5–12
	N%	0.41	0.46	0.20–0.60
	L%	0.46	0.46	0.20–0.70
	RBC ($\times 10^{12}/L$)	2.68	2.12	3.5–5.5
	Hb (g/L)	58	55	110–160
	PLT ($\times 10^9/L$)	281	228	100–400
	EPO (mIU/mL)	1.85	1.76	2.59–18.50
Renal function	BUN (mmol/L)	23.64	27.24	1.8–8.2
+ Liver Function	Cr ($\mu\text{mol/L}$)	512.9	766	20–120
test + Myocardial	UA ($\mu\text{mol/L}$)	355.1	307	90–350
enzyme function	CK (IU/L)	774	380	38–174
	CK-MB (IU/L)	24	14	0–24
	LDH (IU/L)	358	185	0–450
	MB (IU/L)	487.92	560.2	0–90
	NT-ProBNP (pg/mL)	> 25000	25000	0–236
	TBIL ($\mu\text{mol/L}$)	3.5	2.71	3.40–17.0
	DBIL ($\mu\text{mol/L}$)	1.9	2.4	0–6.0
	IBIL ($\mu\text{mol/L}$)	1.6	0.31	3–17.0
	TP (g/L)	53.6	41.8	55–80
	ALB (g/L)	34.1	27.3	35–55
	GLO (g/L)	19.5	14.5	20–35
	ALT (IU/L)	38	49.6	0–40
Electrolyte analysis	AST (IU/L)	40	28.6	0–40
	K ⁺ (mmol/L)	5.26	5.52	3.5–5.5
	Na ⁺ (mmol/L)	114.3	116.91	135–145
	Cl ⁻ (mmol/L)	91.7	82.5	96–108
	Ca ²⁺ (mmol/L)	2.64	1.43	2.1–2.7
	Mg ²⁺ (mmol/L)	0.78	0.64	0.7–1.10
	P (mmol/L)	1.81	2.5	1.45–2.10
Arterial Blood	pH	7.26	7.28	7.35–7.45
gas + Lactic acid	HCO ₃ ⁻ (mmol/L)	9	10.8	22–27
	BE (mmol/L)	-16.6	-15.9	-3–3
	PCO ₂ (mmHg)	20	23	35–45
	PO ₂ (mmHg)	109	228	80–100
	Lac (mmol/L)	1.9	0.9	0.50–2.00

Urine test results are presented in **Supplementary Table S1**. The red values in the table mean the results that are beyond the normal range.

(**Supplementary Figure S2**; Williams et al., 2009). In this study, we identified two novel missense mutations (c.517T > C: p.Cys173Arg and c.667A > C: p.Ser223Arg) in the AGXT gene co-segregated with the infantile form of PH1 in a Chinese



family. Both mutations, c.517T > C and c.667A > C, were not recorded in 500 in-house exome data and did not appear in 191 matched controls. Moreover, both mutations were absent in the Gnom_AD, ExAc, and 1000 Genomes public databases. Mutations on the amino acid position Cys173 of the AGT protein was reported previously (Von Schnakenburg and Rumsby, 1998; van Woerden et al., 2004). van Woerden et al. identified the p.Cys173Ter/ IVS1-1G > A compound heterozygous mutation on a 14-year-old patient with PH1 (van Woerden et al., 2004). The same position of p.Cys173Tyr mutation was also reported in a patient with PH1, as included in the HGMD public database¹ (Von Schnakenburg and Rumsby, 1998).

In the crystal structure of AGT (**Supplementary Figure S2**), a change in Cys173 disrupts the alpha helix, which is on the exposed surface of the AGT protein; notably, Gly170, the most mutated residue on patients with PH1, is also on the alpha

¹<http://www.hgmd.cf.ac.uk>

TABLE 2 | Functional predictions of the two mutations of AGXT in this study.

Variant	Mutation taster ^c	Polyphen ^d	PROVEAN ^e	SIFT ^f
c.517T > C/p.Cys173Arg	Disease-causing prob: 0.9999999999999603 ^a	Probably damaging score: 1 (sensitivity: 0; specificity: 1)	Deleterious score: -10.79 ^b	DAMAGING
c.667A > C / p.Ser223Arg	Disease-causing prob: 0.999994390365302	Probably damaging score: 0.999 (sensitivity: 0.14; specificity: 0.99)	Deleterious score: -4.26	DAMAGING

^a, The probability value is the probability of the prediction, i.e., a value close to one indicates a high “security” of the prediction; ^b, PREDICTION (cutoff = -2.5); ^c, <http://www.mutationtaster.org/>; ^d, <http://genetics.bwh.harvard.edu/pph2/>; ^e, http://provean.jcvi.org/genome_submit_2.php; ^f, https://sift.bii.a-star.edu.sg/www/Extended_SIFT_chr_coords_submit.html.

helix (Von Schnakenburg and Rumsby, 1998; Williams et al., 2009). We noted that amino acids 201–221 of the AGT protein constitute the consensus sequence of the PLP cofactor binding site (Robbiano et al., 2010; Oppici et al., 2013). Without direct contact with a cognate monomer (**Supplementary Figure S2**), the p.Ser223Arg mutation considered in the present study was near the 201–221 PLP cofactor binding site and probably disrupted AGT-PLP binding.

Interestingly, we did not detect the exon 6 mutation of the *AGXT* gene (**Figures 5A–D**) in the family when AGXT6F/AGXT6R (for primer position, see **Supplementary Figure S3**) was used as the PCR primer. This finding is misleading for the diagnosis of PH1 in the family. However, the exon 6 mutation (c.667A > C/p.Ser223Arg) of the *AGXT* was detected through next-generation sequencing and was validated by the second round of Sanger sequencing (**Figure 5B**). The PCR and sequencing conditions for the second round of Sanger sequencing of *AGXT* exon 6 were as follows: the primers, 2AGXT6F/2AGXT6R (**Supplementary Figure S3**); predicted fragment length, 498 bp; and annealing temperature, 60°C. To determine why exon 6-c.667A > C was not detected in the first round of mutation screening, we sequenced the AGXT6F/AGXT6R-333 and 2AGXT6F/2AGXT6R-498 bp fragments by standard Sanger methods and used different sequencing primers. Accordingly, exon6-c.667A > C was not

detected when the AGXT6R was used for PCR as the reverse primer (**Table 3**).

After careful analysis, we found that the AGXT6R primer was not located on SNPs (rs117619103 and rs78178548; **Supplementary Figure S4**), which were linked to the exon 6-c.667A > C mutation, but located between the SNPs, as confirmed by T-A cloning-Sanger sequencing (**Supplementary Figure S5**). Probably, in the presence of these SNPs, the genomic DNA formed a complex conformation (as predicted *in silico*, **Supplementary Figure S6**), and this condition may have disrupted the amplification of the disease allele in the PCR.

To date, at least 190 *AGXT* mutations throughout the entire gene have been detected (Wang et al., 2016), and p.Gly170Arg, p.Phe152Ile, p.Ile244Thr, and c.33-34insC are four of the most common *AGXT* mutations (Williams and Rumsby, 2007; Fargue et al., 2013) in European and North American populations. This result suggests that *AGXT* mutations harbor “hot spots”. In East Asians, the p.Ser205Pro mutation is a PH1-specific mutation in Japanese patients (Kawai et al., 2012). Meanwhile, China, despite being the most populated country in the world, only have rare cases *AGXT* mutations in its population (Yuen et al., 2004; Li et al., 2014; Wang et al., 2016; Cui et al., 2017). To the best of our knowledge, only 11 *AGXT* gene mutations (**Figure 6**) on seven Chinese families have been

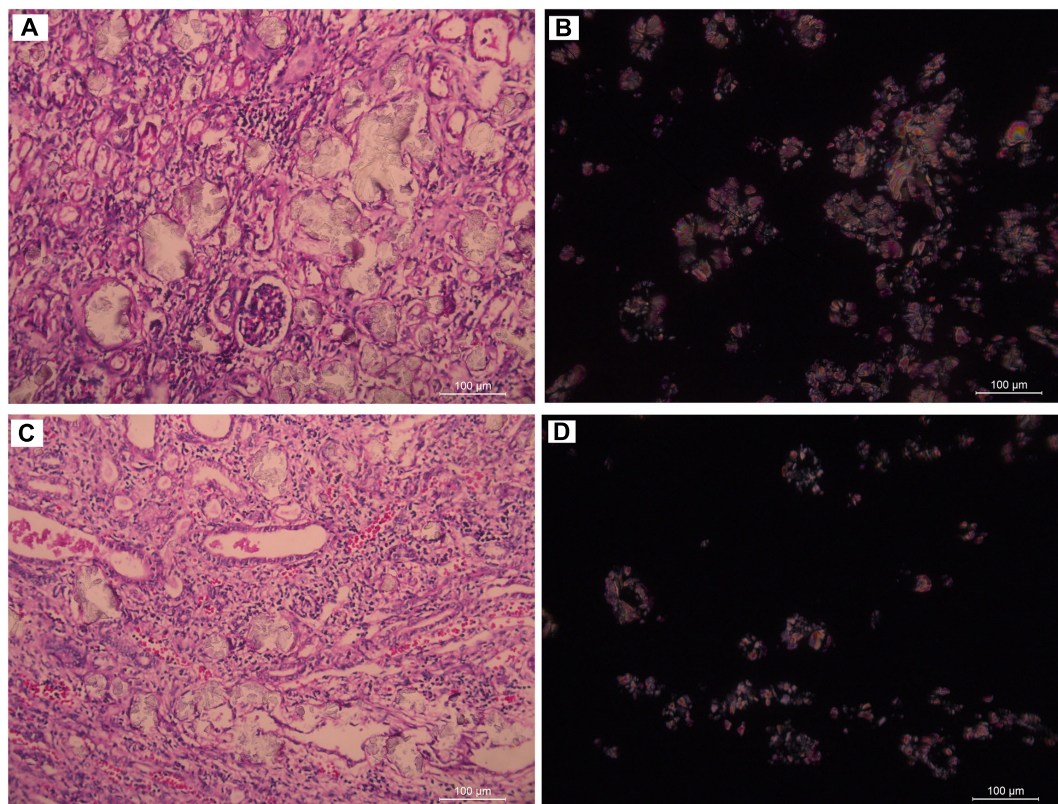


FIGURE 3 | Renal tissue of Ill:4 viewed under the polariscope. **(A,B)** medullary nephrocalcinosis, **(C,D)** cortical nephrocalcinosis. **(A,C)** Under Single polarization **(B,D)** Under crossed nicols.

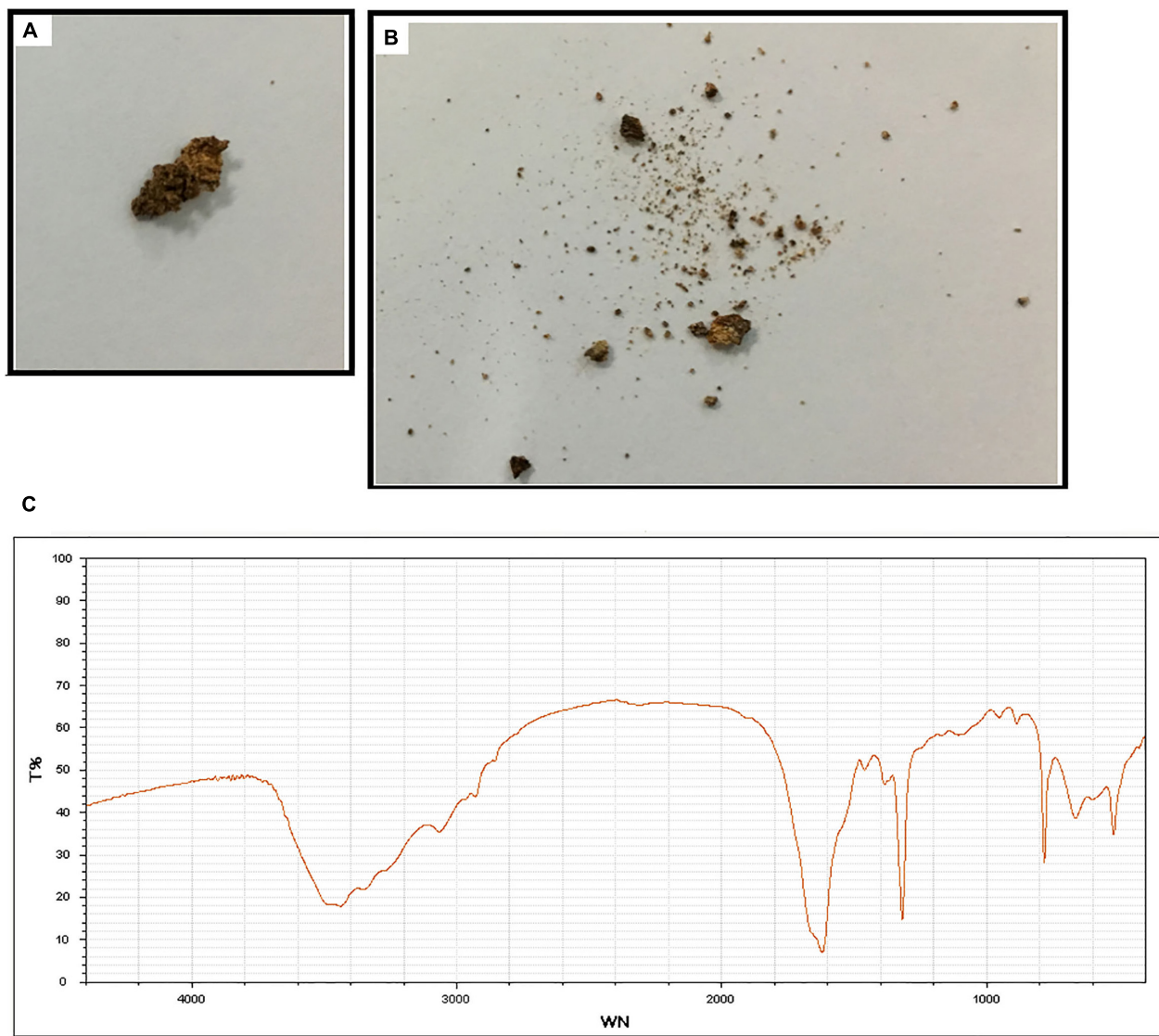


FIGURE 4 | Composition analysis of the renal calcareous sediments. **(A,B)** Renal tissue (5 g) dried at high temperature (180°C) for 2 h. **(C)** Infrared spectroscopy of the sediments revealed the crystal as calcium oxalate monohydrate. An automatic infrared spectrum analysis system, LILR-20 (approved by the Chinese FDA, No. 2008-2210004), was used in this study. T%, absorption frequency; WN, wavenumber.

reported in the country (Yuen et al., 2004; Li et al., 2014; Wang et al., 2016; Cui et al., 2017), and only one mutation carried the c.33-34insC hot spot AGXT mutation; the rest were not located on the hot spot (as mentioned above) nor on the Japanese p.Ser205 residue (Yuen et al., 2004; Li et al., 2014; Wang et al., 2016; Cui et al., 2017). Moreover, only two AGXT mutations, namely, c.2T > C/p.Met1? and c.824_825insAG/S275Rfs*38, have been reported twice in the Chinese populations.

Thus, this situation is consistent with the two missense mutations (of the present study) being unreported elsewhere. All data indicated that AGXT mutation occurs in the Chinese population and may represent a new spectrum of AGXT mutations.

MATERIALS AND METHODS

Ethics Statement

This study was approved by the ethics committee of the Hunan Children's Hospital, Changsha City, China. The procedure of the committee conformed with the principles of the 2008 edition of the declaration of Helsinki.

Before this study, written informed consent was obtained from the guardian of the human subject for the publication of this study.

Study Subjects

This study includes a Han, Chinese family (resides in the rural area of Hunan Province) with PH1 (**Figure 1**) and 496 ethnicity-

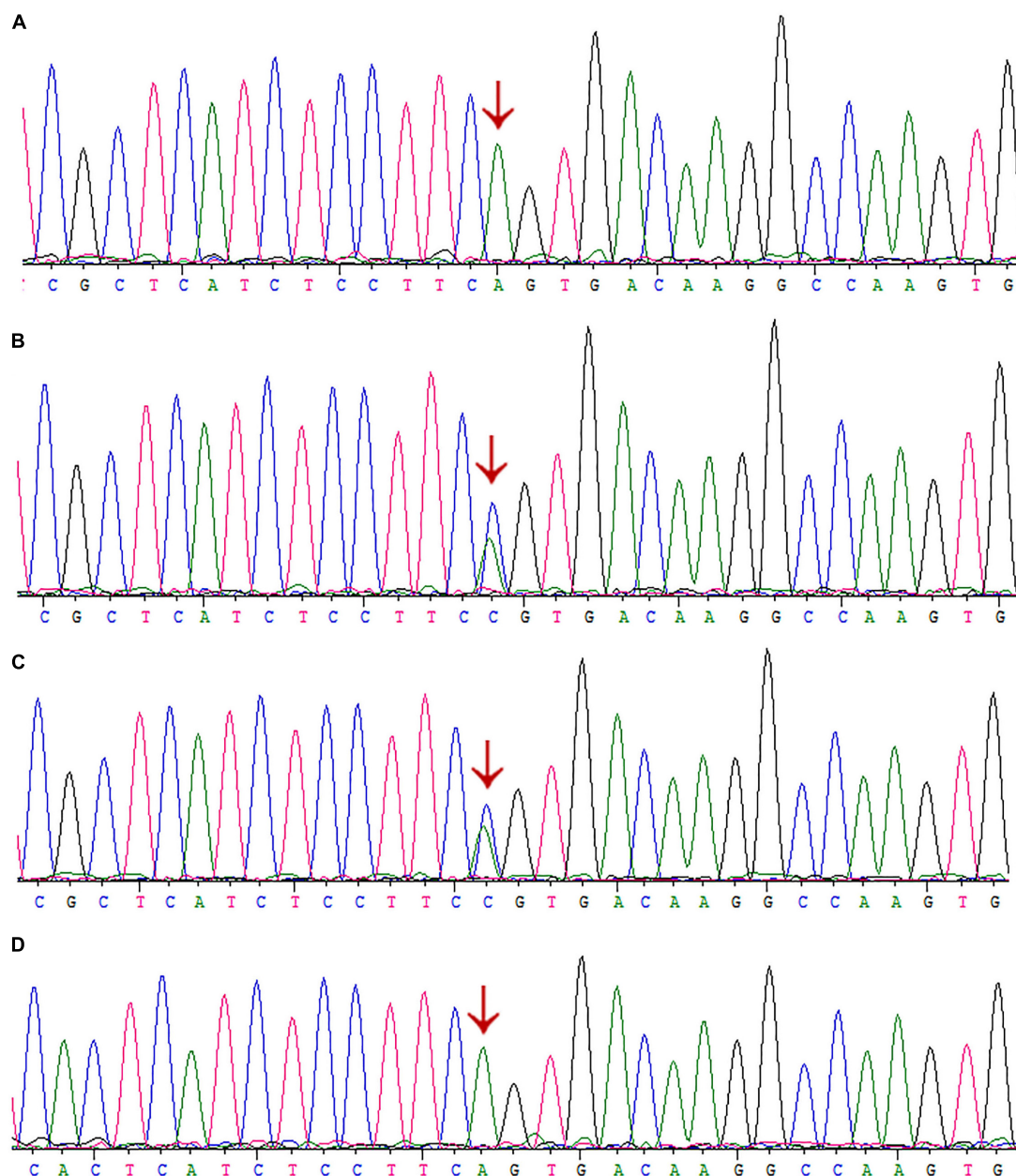


FIGURE 5 | Sanger sequencing of AGXT-exon6 c.667A>C/S223R mutation by different primers. **(A)** Exon 6 of III:3 was PCR-amplified using the primer AGXT6F/AGXT6R and Sanger sequenced using the AGXT6F as the sequencing primer. **(B)** Exon 6 of III:3 was PCR-amplified using the primer 2AGXT6F/2AGXT6R and Sanger sequenced using the 2AGXT6F as the sequencing primer. **(C)** 2AGXT6F/2AGXT6R-amplified fragments Sanger-sequenced by AGXT6F. **(D)** AGXT6F/AGXT6R-amplified fragments Sanger-sequenced by 2AGXT6F.

and region-matched controls (the controls were selected from patients who came to our laboratory for GTG banding but who were excluded for having renal disorders).

In the family, subjects 16 and 17 were affected by the infantile form of PH1, whereas subjects 4, 10, 11, and 13 were affected by benign urolithiasis at least once in that a kidney stone composition analysis indicated that all the crystals were calcium oxalate monohydrate, and other family members were asymptomatic.

For each subject, genomic DNA was extracted from peripheral blood (2–5 mL in heparin sodium tubes) through the phenol/trichloromethane method, as described by the standard protocol.

Sanger Sequencing

For mutation screening, the coding and intron-exon boundary regions of the AGXT gene (NM_000030, according to GRCh37/hg19) were amplified by PCR, using the primers

TABLE 3 | Results of mutation screening for AGXT exon-6-mutation (c.667A > C/p.Ser223Arg) by different primers (Primers sequence and positions are provided in see **Supplementary Figure S2**).

PCR Amplification			Sanger sequencing	
primer-F	primer-R	Size	primer	exon6-c.667A > C
AGXT6F	AGXT6R	333bp	AGXT6F	(-)
AGXT6F	AGXT6R	333bp	AGXT6R	(-)
2AGXT6F	2AGXT6R	498bp	2AGXT6F	(+)
2AGXT6F	2AGXT6R	498bp	2AGXT6R	(+)
AGXT6F	AGXT6R	333bp	2AGXT6F	(-)
2AGXT6F	2AGXT6R	498bp	AGXT6F	(+)
2AGXT6F	2AGXT6R	498bp	AGXT6R	(+)

synthesized by a local biotech company (BGI; Shenzhen, China) and designed with Primer³. The sequencing reaction (BigDye 3.1 Kit, Applied Biosystems, Waltham, MA, United States) of the purified PCR products was performed according to the recommended procedure. The labeled PCR fragments were purified through 70% alcohol precipitation and electrophoresed on an ABI-A3500 genetic analyzer (Applied Biosystems, Waltham, MA, United States). The primer sequences, PCR conditions, and sequencing are shown in **Supplementary Table S2**.

For the validation of the AGXT-exon6 mutation (c.667A > C: p.Ser223Arg), two pairs of primers were used as presented in **Supplementary Figure S3**.

Whole Exome Sequencing

The whole exome sequencing in the present study was performed according to the pipeline as reported elsewhere (Jin et al., 2018). In brief, two micrograms of genomic DNA for each sample (Subjects 16 and 17, **Figure 1**) was sheared into approximately 200-bp fragments. The fragments were enriched with ligation-mediated PCR. Exome capture (NimbleGen 2.1M HD array) was

performed according to the manufacturer's instructions (Roche NimbleGen, Inc., Madison, WI, United States). The captured library was sequenced on a HiSeq2000 sequencing platform (90 bp end reads, Illumina Inc., San Diego, CA, United States), and the Illumina base calling software V1.7 was used for the analysis of raw image files with default parameters. Clean reads were mapped to the reference human genome (hg19)³, by using the BWA (Burrows-Wheeler Aligner) program with, at most, two mismatches⁴. The alignment files (.bam) were generated with SAM tools⁵ and reads of low mapping quality (< Q30) were filtered out. Clonal duplicated reads that may have been derived from PCR artifacts were removed with Picard Tools using default parameters⁶. Short read alignment and annotation visualization were performed by using the Integrative Genomics Viewer⁷. The percentage of alignment of the clean read to the exome regions was obtained by using our custom Perl scripts on the base of the alignment files. Single nucleotide variants (SNVs) and indels were detected with a genome analysis tool kit⁸. All detected SNVs and indels were comprehensively annotated by ANNOVAR⁹, including function implication (gene region, functional effect, mRNA GenBank accession number, amino acid change, cytoband, and so on) and allele frequency, in dbSNP, 1000 Genomes¹⁰, ESP6500¹¹, and ExAc¹². Damaging missense mutations were predicted by SIFT¹³, PolyPhen-2¹⁴, and MutationTaster¹⁵.

³genome.ucsc.edu

⁴bio-bwa.sourceforge.net

⁵samtools.sourceforge.net

⁶broadinstitute.github.io/picard/

⁷http://www.broadinstitute.org/igv/

⁸www.broadinstitute.org/gatk

⁹annovar.openbioinformatics.org/en/latest/

¹⁰www.1000genomes.org

¹¹evs.gs.washington.edu/EVS/

¹²exac.broadinstitute.org/

¹³https://sift.bii.a-star.edu.sg/www/Extended_SIFT_chr_coords_submit.html

¹⁴genetics.bwh.harvard.edu/pph2/

¹⁵www.mutationtaster.org/

²http://frodo.wi.mit.edu

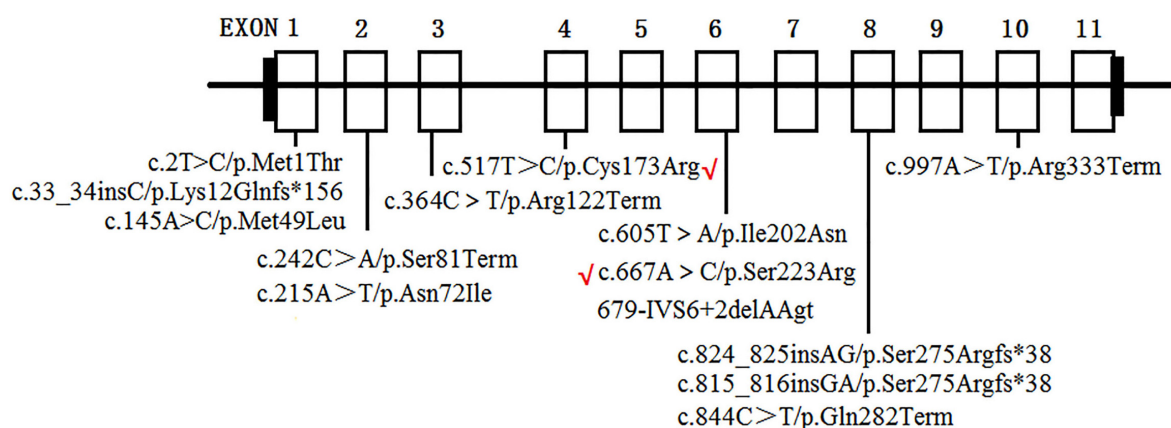


FIGURE 6 | AGXT mutations on Chinese population. Novel mutations identified in the present study are labeled by red "✓".

AUTHOR CONTRIBUTIONS

YY and YZ designed the research. XL, WC, LL, XZ, CH, SL, YY, and YZ performed the sample collection and the research. YY and XL analyzed the experimental data and wrote the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (31501017) and the Key laboratory fund of Hunan Province (2018tP1028 to XL).

REFERENCES

- Coulter-Mackie, M. B., White, C. T., Lange, D., and Chew, B. H. (1993–2017). “Primary Hyperoxaluria Type 1,” in *GeneReviews® [Internet]*, eds R. A. Pagon, M. P. Adam, H. H. Ardinger, S. E. Wallace, A. Amemiya, L. J. H. Bean, et al. (Seattle, WA: University of Washington).
- Cui, Y. J., Song, C. L., and Cheng, Y. B. (2017). Oliguria and acute renal dysfunction in a six-month-old infant. [Article in Chinese]. *Zhongguo Dang Dai Er Ke Za Zhi*. 19, 203–207.
- Fargue, S., Lewin, J., Rumsby, G., and Danpure, C. J. (2013). Four of the most common mutations in primary hyperoxaluria type 1 unmask the cryptic mitochondrial targeting sequence of alanine:glyoxylate aminotransferase encoded by the polymorphic minor allele. *J. Biol. Chem.* 288, 2475–2484. doi: 10.1074/jbc.M112.432617
- Isiyel, E., Ezgu, S. A., Caliskan, S., Akman, S., Akil, I., Tabel, Y., et al. (2016). Molecular analysis of the AGXT gene in patients suspected with hyperoxaluria type 1 and three novel mutations from turkey. *Mol. Genet. Metab.* 119, 311–316. doi: 10.1016/j.ymgme.2016.10.011
- Jellouli, M., Ferjani, M., Abidi, K., Zarrouk, C., Naija, O., Abdelmoula, J., et al. (2016). Primary hyperoxaluria in infants. *Saudi J. Kidney Dis. Transpl.* 27, 526–532. doi: 10.4103/1319-2442.182389
- Jia, J., An, Z., Ming, Y., Guo, Y., Li, W., Liang, Y., et al. (2018). eRAM: encyclopedia of rare disease annotations for precision medicine. *Nucleic Acids Res.* 46, D937–D943. doi: 10.1093/nar/gkx1062
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Jin, Y., Zhang, L., Ning, B., Hong, H., Xiao, W., Tong, W., et al. (2018). Application of genome analysis strategies in the clinical testing for pediatric diseases. *Pediatr. Investig.* 2, 72–81. doi: 10.1002/ped4.12044
- Kawai, C., Minatogawa, Y., Akiyoshi, H., Hirose, S., Suehiro, T., and Tone, S. (2012). A novel mutation of human liver alanine:glyoxylate aminotransferase causes primary hyperoxaluria type 1: immunohistochemical quantification and subcellular distribution. *Acta Histochem. Cytochem.* 45, 121–129. doi: 10.1267/ahc.11042
- Kurt-Sukur, E. D., Özçakar, Z. B., Fitöz, S., Yilmaz, S., Hoppe, B., and Yalcinkaya, F. (2015). Primary hyperoxaluria type 1: a cause for infantile renal failure and massive nephrocalcinosis. *Klin. Padiatr.* 227, 293–295. doi: 10.1055/s-0035-1554638
- Li, G. M., Xu, H., Shen, Q., Gong, Y. N., Fang, X. Y., Sun, L., et al. (2014). Mutational analysis of AGXT in two Chinese families with primary hyperoxaluria type 1. *BMC Nephrol.* 15:92. doi: 10.1186/1471-2369-15-92
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60, 681–685. doi: 10.1007/s11427-017-9100-1
- Oppici, E., Fodor, K., Paiardini, A., Williams, C., Voltattorni, C. B., Wilmanns, M., et al. (2013). Crystal structure of the S187F variant of human liver alanine:

ACKNOWLEDGMENTS

The authors are grateful to the families who participated in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00085/full#supplementary-material>

- glyoxylate [corrected] aminotransferase associated with primary hyperoxaluria type I and its functional implications. *Proteins* 81, 1457–1465. doi: 10.1002/prot.24300
- Robbiano, A., Freccer, V., Miertus, J., Zadro, C., Ulivi, S., Bevilacqua, E., et al. (2010). Modeling the effect of 3 missense AGXT mutations on dimerization of the AGT enzyme in primary hyperoxaluria type 1. *J. Nephrol.* 23, 667–676.
- Shen, Y. (2018). Next-generation sequencing based molecular testing is an equalizer for diagnostic service of rare genetic disorders in China. *Pediatr. Investig.* 2, 96–97. doi: 10.1002/ped4.12036
- Soliman, N. A., Nabhan, M. M., Abdelrahman, S. M., Abdelaziz, H., Helmy, R., Ghanim, K., et al. (2017). Clinical spectrum of primary hyperoxaluria type 1: experience of a tertiary center. *Nephrol. Ther.* 13, 176–182. doi: 10.1016/j.nephro.2016.08.002
- van Woerden, C. S., Groothoff, J. W., Wijburg, F. A., Annink, C., Wanders, R. J., and Waterham, H. R. (2004). Clinical implications of mutation analysis in primary hyperoxaluria type 1. *Kidney Int.* 66, 746–752. doi: 10.1111/j.1523-1755.2004.00796.x
- Von Schnakenburg, C., and Rumsby, G. (1998). Identification of new mutations in primary hyperoxaluria type 1 (PH1). *J. Nephrol.* 11(Suppl. 1), 15–17.
- Wang, C., Lu, J., Lang, Y., Liu, T., Wang, X., Zhao, X., et al. (2016). Two novel AGXT mutations identified in primary hyperoxaluria type-1 and distinct morphological and structural difference in kidney stones. *Sci. Rep.* 6:33652. doi: 10.1038/srep33652
- Williams, E., and Rumsby, G. (2007). Selected exonic sequencing of the AGXT gene provides a genetic diagnosis in 50% of patients with primary hyperoxaluria type 1. *Clin. Chem.* 53, 1216–1221. doi: 10.1373/clinchem.2006.084434
- Williams, E. L., Acquaviva, C., Amoroso, A., Chevalier, F., Coulter-Mackie, M., Monico, C. G., et al. (2009). Primary hyperoxaluria type 1: update and additional mutation analysis of the AGXT gene. *Hum. Mutat.* 30, 910–917. doi: 10.1002/humu.21021
- Yuen, Y. P., Lai, C. K., Tong, G. M., Wong, P. N., Wong, F. K., Mak, S. K., et al. (2004). Novel mutations of the AGXT gene causing primary hyperoxaluria type 1. *J. Nephrol.* 17, 436–440.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lu, Chen, Li, Zhu, Huang, Liu, Yang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Excess of Rare Missense Variants in Hearing Loss Genes in Sporadic Meniere Disease

Alvaro Gallego-Martinez¹, Teresa Requena¹, Pablo Roman-Naranjo¹ and Jose A. Lopez-Escamez^{1,2*} on behalf of the Meniere Disease Consortium (MeDiC)

¹ Otolaryngology and Neurotology Group CTS 495, Department of Genomic Medicine, Centre for Genomics and Oncological Research (GENYO), Pfizer, University of Granada, Andalusian Regional Government, Granada, Spain, ² Department of Otolaryngology, Instituto de Investigación Biosanitaria (Ibs. GRANADA), Hospital Universitario Virgen de las Nieves, Universidad de Granada, Granada, Spain

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Sally Dawson,
University College London,
United Kingdom
Francisco Javier del Castillo,
Hospital Universitario Ramón y Cajal,
Spain
Zhining Wen,
Sichuan University, China

*Correspondence:

Jose A. Lopez-Escamez
antonio.lopezescamez@genyo.es

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 06 September 2018

Accepted: 28 January 2019

Published: 15 February 2019

Citation:

Gallego-Martinez A, Requena T,
Roman-Naranjo P and
Lopez-Escamez JA (2019) Excess of
Rare Missense Variants in Hearing
Loss Genes in Sporadic Meniere
Disease. *Front. Genet.* 10:76.
doi: 10.3389/fgene.2019.00076

Meniere's disease (MD) is a clinical spectrum of rare disorders characterized by vertigo attacks, associated with sensorineural hearing loss (SNHL) and tinnitus involving low to medium frequencies. Although it shows familial aggregation with incomplete phenotypic forms and variable expressivity, most cases are considered sporadic. The aim of this study was to investigate the burden for rare variation in SNHL genes in patients with sporadic MD. We conducted a targeted-sequencing study including SNHL and familial MD genes in 890 MD patients to compare the frequency of rare variants in cases using three independent public datasets as controls. Patients with sporadic MD showed a significant enrichment of missense variants in SNHL genes that was not found in the controls. The list of genes includes *GJB2*, *USH1G*, *SLC26A4*, *ESRRB*, and *CLDN14*. A rare synonymous variant with unknown significance was found in the *MARVELD2* gene in several unrelated patients with MD. There is a burden of rare variation in certain SNHL genes in sporadic MD. Furthermore, the interaction of common and rare variants in SNHL genes may have an additive effect on MD phenotype. This study will contribute to design a gene panel for the genetic diagnosis of MD.

Keywords: SNHL, Meniere's disease, vertigo, variant aggregation, Spanish population

INTRODUCTION

Meniere's disease (MD, MIM 156000) is a chronic disorder of the inner ear characterized by episodes of vertigo, associated with low to middle frequency sensorineural hearing loss (SNHL), tinnitus and aural fullness (Lopez-Escamez et al., 2015). The disorder produces an accumulation of endolymph in the membranous labyrinth, and it may affect both ears in 25–40% of patients (termed bilateral MD) and most of cases are considered sporadic (Paparella and Griebie, 1984). However, heterogeneity in the phenotype is observed and some patients may have co-morbid conditions such as migraine or systemic autoimmune disorders (Gazquez et al., 2011; Caulley et al., 2017). This phenotypic spectrum can make the clinical diagnosis challenging considering that some of the symptoms overlap with other vestibular disorders such vestibular migraine (VM) or autoimmune inner ear disease (AIED) (Hietikko et al., 2011; Lempert et al., 2012; Mijovic et al., 2013; Requena et al., 2014b).

Epidemiological evidence showing a genetic contribution in MD is based on familial aggregation studies with a high siblings recurrence risk ratio ($\lambda_s = 16$ –48) (Requena et al., 2014a) and the

description of multiple familial cases in European and Asian descendant populations (Arweiler-Harbeck et al., 2011; Hietikko et al., 2013). Exome sequencing has identified private variants in the *FAM136A*, *DTNA*, *PRKCB*, *SEMA3D*, and *DPT* genes in 4 families with autosomal dominant MD, showing incomplete penetrance and variable expressivity (Requena et al., 2015; Martín-Sierra et al., 2016, 2017). Moreover, some relatives in familial MD show partial syndromes, either with SNHL or episodic vertigo, increasing the granularity in the phenotype in a given family. However, the genetic contribution of familial genes in sporadic cases has not been investigated and the occurrence of recessive and novel variants is not known. More than 110 genes and ≈6,000 variants have been related to hereditary non-syndromic hearing loss, making gene sequencing panels an essential tool for genetic diagnosis of hearing loss (The Molecular Otolaryngology and Renal Research Laboratories, The University of Iowa, 0000; Sloan-Heggen et al., 2016). Those genes include 45 genes associated to autosomal dominant SNHL and 71 genes related to recessive hearing loss (Shearer et al., 1999; Van Camp and Smith, 2018).

Targeted gene sequencing panels have been demonstrated to be an excellent tool for molecular diagnosis of rare variants in known genes with allelic heterogeneity (Brownstein et al., 2011; Lionel et al., 2018), as well as in sporadic cases of hearing impairment in specific populations (Gu et al., 2015; Dallol et al., 2016). So, we selected SNHL genes and designed a custom gene panel to search for rare and novel variants in sporadic MD.

In the present study, we describe the genetic variation found in a custom exon-sequencing panel of 69 genes in a large cohort of MD sporadic cases. We report that certain genes such as *GJB2*, *USH1G*, *SLC26A4*, and *CLDN14* show an excess of missense variants in sporadic MD cases when compared to controls in the Iberian population, suggesting that several rare variants in these genes may contribute to the SNHL phenotype in sporadic MD.

MATERIALS AND METHODS

Editorial Policies and Ethical Considerations

This study protocol was approved by the Institutional Review Board for Clinical Research (MS/2014/02), and a written informed consent to donate biological samples was obtained from all subjects.

Sample Selection

A total of 890 Spanish and Portuguese patients with MD were recruited. All patients were diagnosed by neurotology experienced clinicians from the Meniere's Disease Consortium (MeDiC), according to the diagnostic criteria for MD formulated by the International Classification Committee for Vestibular Disorders of the Barany Society in 2015 (Lopez-Escamez et al., 2015). Among them, 830 were considered sporadic MD cases while 60 were familial MD cases. Details from the selected cases are described in **Table 1**. As controls, 40 healthy individuals were selected from the same population.

TABLE 1 | Participant subjects in this study.

Diagnosed	N	Sex	%	Location	N
SMD	830	Male	40	North	500
				South	320
FMD	60	Female	60	Center	90
Control	40			Portugal	20

Number of individuals and geographical distribution of the selected cases and controls for targeted-gene sequencing. SMD, sporadic Meniere disease; FMD, familial Meniere disease.

Selection of Target Genes

Target genes were selected from a literature search attending to human phenotype (hearing profile, comorbid vestibular symptoms) and phenotype observations in mouse and zebrafish models. Most of them were selected from HearingLoss.org website gene list for monogenic SNHL. Additional genes were added because they have been previously found in familial MD (Requena et al., 2015; Martín-Sierra et al., 2016, 2017), or allelic variations associated with hearing outcome in MD had been described, such as *NFKB1* or *TLR10* genes (Requena et al., 2013; Cabrera et al., 2014). Mitochondrial genes were added since maternal inheritance is suspected in several families with MD (Requena et al., 2014b). Relevant information about the location, size, bibliography, and other characteristics about each gene included in the panel is presented in **Tables S1, S2**.

The custom panel (Panel ID: 39351-1430751809) were designed by the SureDesign webtool (Agilent) to cover the exons and 50 bp in the flanking regions (5' and 3' UTR). This allowed the sequencing around 533.380 kb with more than 98.46% coverage.

Sample Pooling

Enrichment technology allows the selective amplification of targeted sequences and DNA sample pooling, reducing the costs of reagents and increasing sample size. We decided to pool patient samples according to their geographical origin. Each pool consisted of 10 DNA samples from the same hospital for a total of 93 pools (930 samples).

DNA concentration and quality were measured on each sample using two methods: Qubit dsDNA BR Assay kit (ThermoFisher Scientific) and Nanodrop 2000C (ThermoFisher Scientific). All samples had quality ratios ranging 1.8 and 2.0 in 280/260 and 1.6 to 2.0 in 260/230.

Libraries Preparation Protocol

The HaloPlex Target Enrichment System (Agilent, Santa Clara, CA) was used to prepare the DNA libraries, according to the manufacturer protocol. Validation of the protocol and library performance was analyzed with a 2,100 Bioanalyzer High Sensitivity DNA Assay kit. Expected concentrations were between 1 and 10 ng/ul. Higher concentrations than 10 ng/ul were diluted 1:10 in 10 mM TRIS, 1 mM EDTA. Targeted-sequencing was performed in an Illumina Nextseq500 platform.

Data Generation Pipelines

Raw data downloaded and sequencing adapters were trimmed following manufacturer indications. The requested depth of coverage for the sequencing panel was 250X. The minimum coverage considered was 30X mean depth for nuclear genes, however mitochondrial sequences reached higher coverages with the enrichment technology. Bioinformatic analyses were performed according to the Good Practices recommended by Genome Analysis ToolKit (<https://software.broadinstitute.org/gatk/>). Mitochondrial genes were analyzed using the same pipeline.

Two methods were used to find differences in how UnifiedGenotyper and HaplotypeCaller (the old and the most recent tools for variant calling in GATK suite) address sequenced pools. Both custom pipelines use BWA-mem aligner and GATK suite tools following the GATK protocol for Variant Calling against GRCh37/hg19 human reference genome. Left normalization for multi-allelic variants were addressed by separated. Calling was made in the first pipeline with UnifiedGenotyper modifying number of chromosomes per sample (per pool, there are 20 chromosomes). The second pipeline used HaplotypeCaller, which cannot allow the same approach, but can automatically address high number of calls with a different approach. Variants with read depth (RD) <10 and genotype quality (GQ) <20 were excluded in all the calling pipelines following recommended hard filtering steps by GATK suite.

A third caller tool, VarScan, was used to filter and annotate quality strand data per variant to compare its output with

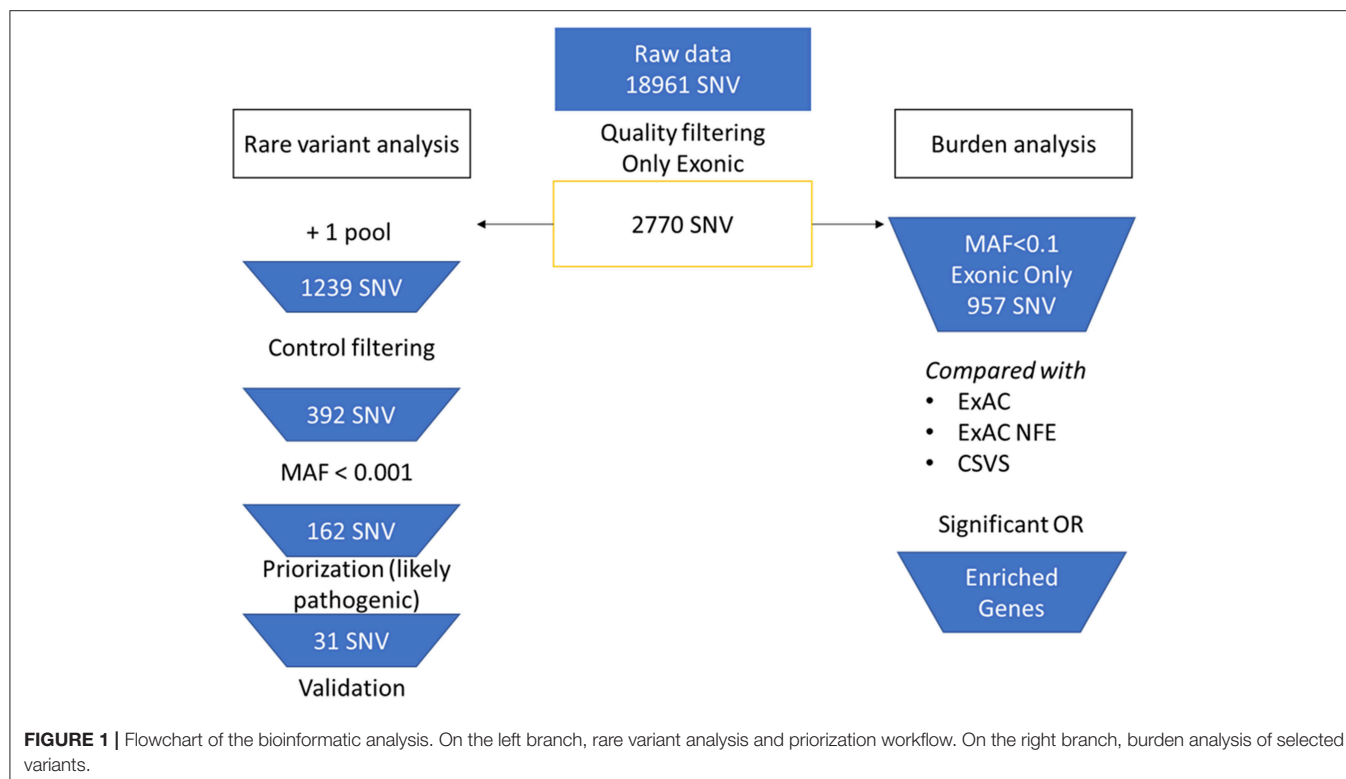
GATK-based callers. VarScan allows the variant filtering using the information obtained according to each strand polarity. The method retrieves those variants that were only called in one strand, but not in the reverse strand, leading to false positive calls. This step was used as internal quality control to avoid strand bias usually generated in Haloplex data, as it has been reported in other studies (Collet et al., 2015).

Positive Control SNV Validation

Positive control testing was addressed using samples from patients with familial MD with known variants on certain genes. These individuals come from previous familial studies with independently validated variants by Sanger sequencing. Known variants were also sequenced and validated by Sanger (Table S3). Coverage and mapping quality after each pipeline were annotated and measured. Representative chromatographs from validated SNVs are detailed in Figure S1.

Selection and Prioritization of Pathogenic SNV

In order to obtain more information of each SNV, we annotated the merged files using the ANNOVAR tool. Minor allele frequencies (MAF) were retrieved for each candidate variant from gnomAD and ExAC database (total individuals and non-Finnish European (NFE) individuals). Since the estimated prevalence of sporadic MD in Spain is 0.75/1,000 individuals (Morales Angulo et al., 2003), we selected variants with MAF <0.001 for single rare variant analysis and prioritized them according to Combined Annotation Dependent Depletion



(CADD) phred score. For burden analysis of common and rare variants, we chose a higher MAF value <0.1 . The Collaborative Spanish Variant Server (CVCS) database including 1,644 unrelated individuals was also used for annotation of exonic rare variants and to retrieve MAF in Spanish population (dataset fully accessible from <http://csvs.babelomics.org/>) (Dopazo et al., 2016).

KGGseq suite (grass.cgs.hku.hk/limx/kggseq) was used for the selection of rare variants to prioritize the most pathogenic variants according to the integrated model trained algorithm with known pathogenic variants and neutral control variants.

Enrichment analysis for each gene was made with all the exonic variants found with a MAF <0.1 . This analysis required to divide the total amount of variants into three groups: those described in global ExAC population, those found in NFE population, and finally those included in CVCS Spanish population. These three reference datasets were used for enrichment analysis comparison.

Validation of Candidate Pathogenic SNV

Candidate SNV were visually revised in the BAM files using Integrative Genomics Viewer (<https://software.broadinstitute.org/software/igv/>) and validated in the different pools where they were called using Sanger sequencing.

Population Statistics

Statistical analysis was performed with IBM SPSS v.20 program, Microsoft Excel suite tools, and diverse python and java encoded public scripts. Due to the overrepresentation of Spanish population in our dataset, most of the selected variants were filtered through exome sequencing data from Spanish controls of CSVS database. The MAF was calculated for each variant in our dataset and rare and previously unreported variants on MD patients were identified in our gene panel. Odds ratios with 95% confidence interval were calculated for each variant using MAF obtained from Spanish population ($N = 1,579$), ExAC ($N = 60,706$), and ExAC NFE ($N = 33,370$) populations as controls.

Gene burden analysis was addressed using 2×2 contingency tables counting total exonic alternate allele counts per gene in our cases against total and NFE controls in ExAC and CSVS controls. Odds ratios with 95% confidence intervals were calculated using Fisher's exact test and obtaining one-sided p -values. P -values were also corrected for multiple testing by the total amount of variants found for each gene following Bonferroni approach.

Position of Variants in Significant Enriched Genes

Several models were generated for rare variant-enriched domains in significant enriched genes by using the INSIDER modeling tool (Meyer et al., 2018). The selected variants per gene are detailed in results. Prediction values were annotated with their calculated p values.

RESULTS

Single Rare Variant Analysis

We achieved an average capture efficiency rate (percentage of total on-target reads in total sequenced reads) of 69% on the target regions above 30X (minimum depth considered for quality filtering). The mean coverage percentage can be found in **Table S4**. A total of 2,770 SNV in nuclear genes were selected from the raw merged dataset (18,961 SNVs) after filtering by quality controls. The analysis workflow is summarized in **Figure 1**. For rare variants analysis, SNV that were found in more than one pool were selected, remaining 1,239 variants. After that, we filtered by variants observed in the control pools, leaving only 392 exonic SNV in cases (278 missense, 111 synonymous, 2 stopgain, and 1 stoploss).

A final set of 162 SNV with a MAF <0.001 were retrieved (143 missense, 18 synonymous, 1 stoploss, 1 stopgain). All the exonic variants were annotated and scored using different prioritization tools. Of them, 136 SNVs were not previously described in any population database and we considered them as potential novel variants.

After prioritizing the exonic variants by CADD phred, 31 rare variants remained (**Table 2**). Six of them were validated by Sanger sequencing in more than two individuals in the following genes: *GJB2*, *ESRRB*, *USH1G*, *SLC26A4* (**Table S3**). The rest of the variants were considered benign or likely benign since they did not reach the pathogenicity threshold predicted for KGGSeq. However, a novel synonymous variant in the *MARVELD2* gene was found and validated in three unrelated individuals.

The minor allelic frequencies in SNV of the 24 mitochondrial genes included in the panel were compared with the reference data obtained from MITOMap through its automated mtDNA sequence analysis system Mitomaster (Ruiz-Pesini et al., 2007). However, the candidate variants observed do not belong to the genes targeted in the mitochondrial genome, since they were not validated by Sanger sequencing. We did not find any SNV associated with MD (data not shown).

Gene Burden Analysis

To analyse the interaction of multiple variants, we considered SNV with a MAF <0.1 for the gene burden analysis. A total of 957 exonic variants were retrieved and their frequencies were compared with the global and NFE frequencies from ExAC, and with the Spanish population frequencies from CSVS.

A gene burden analysis using our gene set was performed using these three reference datasets. After Bonferroni correction, some genes showed a significant enrichment of rare variants in the three comparisons, making them candidate genes to be selected for a diagnosis panel for MD (**Table 3**). Moreover, 6 genes (*FAM136A*, *ADD1*, *SLC12A2*, *POU4F3*, *RDX*, and *PRKCB*) presented some novel variants that were validated by Sanger, but they have not been described in global ExAC or CSVS datasets. Although these previously unreported variants could not be sequenced in all the parents of these patients, we considered them as potential *de novo* variants.

A second variant analysis using the missense variants described in CSVS Spanish population database was made

TABLE 2 | Prioritized rare SNVs found in the single rare variant analysis for sporadic MD cases.

Chr.	Position	Ref	Alt	Gene	Exonic function	dbSNP	ExAC MAF	gnomAD MAF	Cohort MAF	CADD phred score
chr1	6,488,328	C	T	<i>ESPN</i>	Non-synonymous	rs143577178	0.0005	0.0005	0.004819	35
chr4	6,303,574	G	T	<i>WFS1</i>	Synonymous		–	–	0.016867	–
chr4	6,303,946	C	A	<i>WFS1</i>	Non-synonymous		–	–	0.004819	22.1
chr4	6,304,133	G	C	<i>WFS1</i>	Non-synonymous			4.08E-06	0.010843	21
chr4	38,775,706	G	T	<i>TLR10</i>	Non-synonymous		–	–	0.003614	20.7
chr4	38,776,070	C	A	<i>TLR10</i>	Non-synonymous		–	–	0.024096	14.27
chr5	68,715,821	G	T	<i>MARVELD2</i>	Non-synonymous		–	–	0.003614	23.3
chr7	107,336,408	A	C	<i>SLC26A4</i>	Non-synonymous	rs200511789	0.0004	0.0004	0.003614	24.3
chr10	14,563,299	G	T	<i>FAM107B</i>	Non-synonymous		–	–	0.003614	24.9
chr11	76,885,923	G	A	<i>MYO7A</i>	Non-synonymous	rs781991817	0.0002	0.0003	0.004819	34
chr11	76,892,617	G	C	<i>MYO7A</i>	Non-synonymous	rs200641606	0.0007	0.0007	0.003614	25.5
chr12	65,672,602	C	T	<i>MSRB3</i>	Synonymous	rs149757878	0.0002	0.0005	0.004819	–
chr13	20,763,612	C	A	<i>GJB2</i>	Non-synonymous	rs72474224	–	–	0.007229	23.8
chr14	76,957,891	G	A	<i>ESRRB</i>	Non-synonymous	rs201344770	0.0003	0.0002	0.008434	23.8
chr14	76,966,336	G	A	<i>ESRRB</i>	Non-synonymous	rs200237229	0.0007	0.0005	0.003614	1.198
chr14	76,966,347	C	T	<i>ESRRB</i>	Non-synonymous	rs201448899	0.001	0.0007	0.008434	15.41
chr16	24,046,852	C	T	<i>PRKCB</i>	Synonymous	rs115645964	0.0003	0.0003	0.003614	–
chr17	72,915,919	C	T	<i>USH1G</i>	Non-synonymous	rs151242039	0.0006	0.0008	0.003614	8.91
chr17	72,916,543	T	G	<i>USH1G</i>	Non-synonymous	–	–	0.004819	14.31	–
chr17	79,478,028	G	A	<i>ACTG1</i>	Synonymous	rs187127467	0.0002	0.0003	0.008434	–
chr18	32,462,094	G	T	<i>DTNA</i>	Non-synonymous	rs533568822	2.47E-05	3.66E-05	0.003614	25
chr19	50,784,974	A	C	<i>MYH14</i>	Non-synonymous		–	–	0.003614	20.3
chr22	38,119,197	G	T	<i>TRIOBP</i>	Non-synonymous		–	–	0.003614	17.9
chr22	38,119,969	C	A	<i>TRIOBP</i>	Non-synonymous		–	–	0.010843	15
chr22	38,119,977	G	T	<i>TRIOBP</i>	Non-synonymous		–	–	0.003614	0.022
chr22	38,120,106	A	C	<i>TRIOBP</i>	Non-synonymous		–	–	0.003614	2.968
chr22	38,120,116	C	A	<i>TRIOBP</i>	Non-synonymous		–	–	0.024096	16.55
chr22	38,120,263	C	A	<i>TRIOBP</i>	Non-synonymous		–	–	0.020482	22.3
chr22	38,120,302	C	A	<i>TRIOBP</i>	Non-synonymous		–	–	0.003614	23.8
chr22	38,168,691	G	C	<i>TRIOBP</i>	Non-synonymous	rs373236761	0.0001	0.0002	0.003614	26.6

Minor allele frequency for each SNV is detailed as annotated by ExAC and gnomAD (exomes). Pathogenicity prediction is detailed according to CADD phred score.

(Table 4). Eighteen genes showed an excess of missense variants (a total of 46 variants, detailed in Table S5). Of note, five genes causing autosomal recessive SNHL showed the highest accumulation of missense variants when they were compared with NFE and Spanish population datasets: *SLC26A4*, *GJB2*, *CLDN14*, *ESRRB*, and *USH1G*. The variants in these five genes were validated through Sanger sequencing and considered Spanish population-specific variants.

Excess of Rare Variants in Hearing Loss Genes in Familial Cases

We used exome sequencing datasets from familial MD cases previously reported to search for rare variants identified in our panel in the sporadic cases. Although no single missense variant was found segregated in all the cases in the same family, we found several rare missense variants in at least one case per family in genes such as *GJB2*, *GRHL2*, *TRIOBP*, *RDX*, *KCNQ4*, *WFS1*, and *ADD1*. These MD families show phenotypic differences in terms of age of onset, hearing profile and disease progression

and the presence of rare variants can be addressed as potential modulators of the phenotype in each familial case (Table 5).

Effect of Rare Variant Interaction

We selected exonic variants from the gene burden analysis to analyze their potential additive effect at the protein-protein interaction interfaces by the tool INSIDER for our selected five genes. However, protein interfaces for *ESRRB*, *CLDN14* and *SLC26A4* genes could not be loaded and processed on the database (lacking predicted interfaces on ÉCLAIR database or crystalized protein structures on Protein Data Bank (PDB) database). Of note, most relevant affected interaction is observed in the self-interaction *GJB2-GJB2* by the known variants observed in the burden analysis (significant spatial clustering with 4 SNV, $p = 0.0009$) rs111033218:G>C (p.Phe83Leu), rs80338945:A>G(p.Leu90Pro), rs374625633:T>C(p.Ile30Val), and rs2274084:C>T(p.Val27Ile) (Figure 2A).

Other interactions of interest were founded between the *USH1G*—*USH1C* genes, but the involved variants were not located in the known interaction surface of *USH1G* (Figure 2B).

TABLE 3 | Gene burden analysis 1.

Gene	# variants	Odds ratio ExAC ^a	P value	P corrected ^b	Odds ratio ExAC NFE ^a	P value	P corrected ^b	Odds ratio CSVS ^a	P value	P corrected ^b
MYH14	50	13.54 (5.85–31.37)	1.20E-09	6.01E-08	56.94 (25.13–128.98)	<1.00E-15	<1.00E-15	38.52 (16.95–87.56)	<1.00E-15	<1.00E-15
MYO7A	43	3.11 (1.73–5.57)	1.41E-04	6.08E-03	4.64 (2.85–8.13)	7.72E-08	3.32E-06	3.85 (2.18–6.81)	3.60E-06	1.55E-04
WFS1	36	1.24 (0.6–2.57)	5.67E-01	1	1.84 (0.94–3.62)	7.70E-02	1	3.28 (1.76–6.1)	1.83E-04	6.57E-03
ADD1	29	249.89 (96.17–649.29)	<1.00E-15	<1.00E-15	318.03 (122.45–826)	<1.00E-15	<1.00E-15	747.23 (287.95–1939.08)	<1.00E-15	<1.00E-15
WHRN	28	24.63 (7.96–76.17)	2.68E-08	7.51E-07	39.97 (13.03–122.58)	1.12E-10	3.13E-09	62.91 (20.62–191.98)	3.44E-13	9.64E-12
TPRN	26	36.63 (10.61–126.49)	1.24E-08	3.22E-07	46.93 (13.64–161.49)	1.03E-09	2.68E-08	81.9 (23.93–280.26)	2.24E-12	5.81E-11
USH1G	24	4.04 (1.91–8.51)	2.46E-04	5.90E-03	18.38 (9.26–36.48)	<1.00E-15	<1.00E-15	8.12 (4–16.49)	6.57E-09	1.58E-07
USH1C	22	5.12 (2.58–10.17)	3.13E-06	6.88E-05	10.68 (5.54–20.59)	1.54E-12	3.38E-11	5.12 (2.58–10.18)	3.06E-06	6.73E-05
P2RX2	22	20.34 (9.75–42.44)	<1.00E-15	<1.00E-15	23.74 (11.41–49.41)	<1.00E-15	<1.00E-15	23.51 (11.29–48.94)	<1.00E-15	<1.00E-15
ESPN	19	27.88 (17.06–45.55)	<1.00E-15	<1.00E-15	24.35 (14.88–39.84)	<1.00E-15	<1.00E-15	17.39 (10.59–28.56)	<1.00E-15	<1.00E-15
RDX	17	94.37 (36.06–247)	<1.00E-15	<1.00E-15	57.38 (21.85–150.66)	<1.00E-15	<1.00E-15	75.26 (28.72–197.23)	<1.00E-15	<1.00E-15
TJP2	17	2.78 (1.59–4.88)	3.58E-04	6.09E-03	7.99 (4.79–13.32)	<1.00E-15	<1.00E-15	2.92 (1.67–5.11)	1.67E-04	2.85E-03
SLC26A4	15	1.79 (1.02–3.15)	4.31E-02	6.46E-01	4.14 (2.5–6.86)	3.37E-08	5.06E-07	3.65 (2.19–6.09)	6.62E-07	9.93E-06
ESRRB	14	12.54 (7.26–21.63)	<1.00E-15	<1.00E-15	10.67 (6.16–18.48)	<1.00E-15	<1.00E-15	7.54 (4.31–13.18)	1.41E-12	1.98E-11
DTNA	13	6.07 (4.68–7.87)	<1.00E-15	<1.00E-15	4.87 (3.74–6.35)	<1.00E-15	<1.00E-15	4.78 (3.67–6.23)	<1.00E-15	<1.00E-15
PRKCB	13	27.62 (18.65–40.91)	<1.00E-15	<1.00E-15	17.6 (11.84–26.16)	<1.00E-15	<1.00E-15	40.79 (27.61–60.28)	<1.00E-15	<1.00E-15
SLC12A2	12	21.91 (15.95–30.1)	<1.00E-15	<1.00E-15	55.14 (40.31–75.43)	<1.00E-15	<1.00E-15	79.32 (58.04–108.41)	<1.00E-15	<1.00E-15
KCNQ1	11	7.93 (5.92–10.63)	<1.00E-15	<1.00E-15	13.75 (10.33–18.29)	<1.00E-15	<1.00E-15	6.96 (5.18–9.34)	<1.00E-15	<1.00E-15
NFKB1	10	1.87 (1.3–2.71)	7.81E-04	7.81E-03	1.92 (1.33–2.77)	4.69E-04	4.69E-03	3.79 (2.72–5.29)	4.44E-15	4.44E-14
DPT	10	1.88 (1.2–2.92)	5.36E-03	5.36E-02	8.03 (5.5–11.74)	<1.00E-15	<1.00E-15	4.38 (2.94–6.51)	3.01E-13	3.01E-12
GJB2	9	0.59 (0.31–1.12)	1.06E-01	9.51E-01	2.81 (1.77–4.44)	1.04E-05	9.32E-05	2.5 (1.57–3.98)	1.21E-04	1.09E-03
CEACAM16	9	1.25 (1.04–1.49)	1.56E-02	1.40E-01	1.33 (1.12–1.59)	1.48E-03	1.33E-02	2.45 (2.09–2.87)	<1.00E-15	<1.00E-15
FAM136A	8	31.18 (27.22–35.71)	<1.00E-15	<1.00E-15	38.36 (33.5–43.92)	<1.00E-15	<1.00E-15	14.71 (12.81–16.89)	<1.00E-15	<1.00E-15
GRHL2	8	11.72 (9.35–14.7)	<1.00E-15	<1.00E-15	9.28 (7.39–11.66)	<1.00E-15	<1.00E-15	6.45 (5.11–8.15)	<1.00E-15	<1.00E-15
EYA4	7	36.65 (28.83–46.59)	<1.00E-15	<1.00E-15	25.31 (19.88–32.22)	<1.00E-15	<1.00E-15	7.68 (5.97–9.88)	<1.00E-15	<1.00E-15
COCH	7	3.41 (2.54–4.58)	<1.00E-15	<1.00E-15	3.39 (2.52–4.56)	<1.00E-15	<1.00E-15	3.95 (2.95–5.28)	<1.00E-15	<1.00E-15
CCDC50	5	1.25 (0.99–1.59)	6.22E-02	3.11E-01	1.62 (1.3–2.03)	2.51E-05	1.25E-04	3.22 (2.63–3.94)	<1.00E-15	<1.00E-15
KCNJ10	5	4.38 (3.9–4.93)	<1.00E-15	<1.00E-15	3.14 (2.78–3.55)	<1.00E-15	<1.00E-15	2.3 (2.02–2.61)	<1.00E-15	<1.00E-15
SEMA3D	4	1.56 (1.03–2.38)	3.60E-02	1.44E-01	1.13 (0.72–1.77)	5.84E-01	1	3.79 (2.63–5.47)	1.09E-12	4.35E-12
CLDN14	4	4.51 (2.56–7.94)	1.74E-07	6.97E-07	22.86 (13.56–38.55)	<1.00E-15	<1.00E-15	4.72 (2.69–8.29)	6.61E-08	2.65E-07
POU4F3	4	7.31 (6.53–8.18)	<1.00E-15	<1.00E-15	4.92 (4.38–5.62)	<1.00E-15	<1.00E-15	9.2 (8.23–10.28)	<1.00E-15	<1.00E-15
MSRB3	4	13.35 (11.34–15.71)	<1.00E-15	<1.00E-15	10.85 (9.21–12.79)	<1.00E-15	<1.00E-15	4.06 (3.4–4.83)	<1.00E-15	<1.00E-15

^aOdds ratios were calculated in the 95% confidence interval.^bP-values were corrected with Bonferroni method.

List of 29 genes showing a significant excess of missense exonic variants in patients with sporadic MD, according to the MAF observed in global ExAC population (N = 60,706), non-Finnish European ExAC population (NFE) (N = 33,370) and Spanish population from CSVS (N = 15,79).

TABLE 4 | Gene burden analysis 2.

Gene	Number variants	Percentage variants retained	Odds ratio ExAC ^a	P value	P corrected ^b	Odds ratio ExAC NFE ^a	P value	P corrected ^b	Odds ratio CSVS ^a	P value	P corrected ^b
GJB2	6	80	0.5 (0.27–0.93)	2.91E-02	1.75E-01	3.2 (2.12–4.83)	2.75E-08	1.65E-07	2.06 (1.33–3.19)	1.14E-03	6.85E-03
SEMA3D	2	50	1.1 (0.76–1.61)	6.13E-01	1	0.8 (0.53–1.21)	2.85E-01	5.70E-01	2.67 (1.94–3.68)	2.03E-09	4.06E-09
CLDN14	2	50	4.47 (2.55–7.83)	1.67E-07	3.35E-07	23.18 (13.81–38.9)	<1.00E-15	<1.00E-15	4.64 (2.65–8.11)	7.47E-08	1.49E-07
SLC26A4	6	40	1.18 (0.72–1.93)	5.04E-01	1	2.88 (1.89–4.38)	8.13E-07	4.88E-06	2.33 (1.51–3.59)	1.23E-04	7.37E-04
NFKB1	3	30	1.37 (0.99–1.91)	5.92E-02	1.78E-01	1.43 (1.03–1.98)	3.37E-02	1.01E-01	2.73 (2.03–3.66)	2.21E-11	6.62E-11
POU4F3	1	25	1.56 (1.47–1.66)	<1.00E-15	<1.00E-15	1.06 (1–1.14)	6.16E-02	6.16E-02	1.84 (1.73–1.95)	<1.00E-15	<1.00E-15
ESRRB	3	21	4.41 (3.31–5.89)	<1.00E-15	<1.00E-15	3.39 (2.52–4.55)	<1.00E-15	<1.00E-15	1.84 (1.33–2.54)	2.04E-04	6.12E-04
USH1G	5	21	2.51 (1.38–4.56)	2.66E-03	1.33E-02	20.27 (12.06–34.06)	<1.00E-15	<1.00E-15	4.67 (2.68–8.17)	6.11E-08	3.05E-07
CCDC50	1	20	0.54 (0.44–0.66)	7.23E-10	7.23E-10	0.7 (0.59–0.84)	1.41E-04	1.41E-04	1.38 (1.19–1.61)	3.38E-05	3.38E-05
P2RX2	3	14	2.57 (1.93–3.42)	8.71E-11	2.61E-10	3.14 (2.38–4.14)	<1.00E-15	<1.00E-15	2.67 (2.01–3.55)	1.12E-11	3.36E-11
FAM136A	1	13	140.36 (131.25–150.11)	<1.00E-15	<1.00E-15	76.75 (71.75–82.09)	<1.00E-15	<1.00E-15	3.68 (3.41–3.97)	<1.00E-15	<1.00E-15
RDX	2	12	4.49 (3.59–5.61)	<1.00E-15	<1.00E-15	2.69 (2.12–3.4)	<1.00E-15	<1.00E-15	3.33 (2.65–4.2)	<1.00E-15	<1.00E-15
TPRN	3	12	2.24 (1.7–2.96)	1.46E-08	4.37E-08	6.55 (5.1–8.4)	<1.00E-15	<1.00E-15	2.95 (2.26–3.86)	<1.00E-15	<1.00E-15
ESPN	2	11	10.69 (9–12.69)	0.00E+00	<1.00E-15	10.26 (8.64–12.19)	<1.00E-15	<1.00E-15	2.01 (1.64–2.46)	9.24E-12	1.85E-11
SLC12A2	1	8	281.71 (246.32–322.17)	<1.00E-15	<1.00E-15	308.08 (269.39–352.33)	<1.00E-15	<1.00E-15	14.77 (12.86–16.96)	<1.00E-15	<1.00E-15
PRKCB	1	8	15.1 (13.03–17.48)	<1.00E-15	<1.00E-15	10.5 (9.05–12.19)	<1.00E-15	<1.00E-15	5.54 (4.75–6.46)	<1.00E-15	<1.00E-15
MYH14	3	6	14.44 (10.42–20.03)	<1.00E-15	<1.00E-15	28.12 (20.38–38.79)	<1.00E-15	<1.00E-15	5.86 (4.16–8.25)	<1.00E-15	<1.00E-15
ADD1	1	3	563.42 (492.71–644.27)	<1.00E-15	<1.00E-15	308.08 (269.39–352.33)	<1.00E-15	<1.00E-15	14.77 (12.86–16.96)	<1.00E-15	<1.00E-15

^aOdds ratios were calculated in the 95% confidence interval.^bP-values were corrected with Bonferroni method.

List of 18 genes showing a significant excess of previously reported missense exonic variants in patients with sporadic MD, according to the MAF observed in CSVS Spanish database (N = 1579), compared with global ExAC population (N = 60,706) and non-Finnish European ExAC population (N = 33,370). In bold, selected genes with higher percentage of variants retained (>20%) and significant OR on Spanish and NFE populations. Selected SNV can be consulted in Table S5.

TABLE 5 | Missense variants found in familial MD cases.

Cases	Family	SNV	Gene	RefSeq	Described in	Pathogenicity	MAF (gnomAD)	MAF (ExAC)
All	1	chr2:70527974 G>A	FAM136A	NM_032822.2	(Requena et al., 2015)	Pathogenic		
1	1	chr4:6303197 G>A	WFS1	NM_006005.3	No	Conflicting interpretations of pathogenicity	0.0041	0.0045
2	1	chr13:20763264 C>T	GJB2	NM_004004.5	No	Benign/likely benign	0.0094	0.0106
All	1	chr18:32462094 G>T	DTNA	NM_001198938.1	(Requena et al., 2015)	Likely pathogenic	3.66E-05	2.47E-05
2	2	chr1:41296865 G>A	KCNQ4	NM_004700.3	No	Unknown significance	2.13E-05	3.49E-05
All	2	chr1:168665849 G>A	DPT	NM_001937.4	(Martín-Sierra et al., 2017)	Likely pathogenic	2.03E-05	2.5E-05
2	2	chr4:2900221 A>G	ADD1	NM_014189.3	No	Unknown significance	8.12E-06	8.24E-06
All	3	chr7:84642128 C>T	SEMA3D	NM_152754.2	(Martín-Sierra et al., 2017)	Pathogenic		
All	3	chr8:102555482 G>T	GRHL2	NM_024915.3	No	Unknown significance		
All	4	chr16:23999898 G>T	PRKCB	NM_002738.6	Martín-Sierra et al., 2016	Pathogenic		
1	4	chr22:38119405 C>T	TRIOBP	NM_001039141.2	No	Likely benign	4.06E-05	2.49E-05
1	4	chr3:191098660 A>G	CCDC50	NM_178335.2	No	Benign	0.0061	0.0065
1	4	chr11:110134833 T>C	RDX	NM_002906.3	No	Unknown significance		

Variants were retrieved from familial cases segregating a partial phenotype in different families.

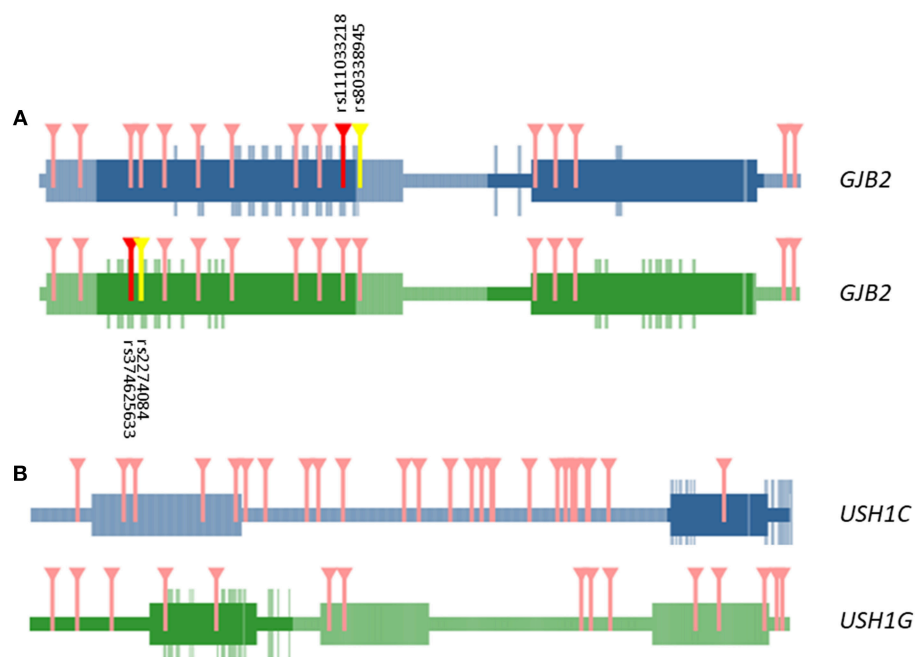


FIGURE 2 | Representation of domains and interactive interface in GJB2–GJB2 (A) and USH1G–USH1C (B) interaction. Marked in darker color boxes, interactive surfaces of the protein. Aminoacids in the interactive surface of the protein are highlighted in the same color. Variants that affect the interaction regions between both proteins are marked in red and yellow. Only in GJB2–GJB2 self-interaction missense mutations are relevant in the interaction (dbSNP ids detailed in black). The rest of the variants affecting aminoacids tested in both interactions that are out in the interactive surface region are marked in pink.

To assess if the SNHL genes showing enrichment of missense variants were located in genomic regions with a higher recombination rates, we retrieved recombination rates from deCODE genetics maps for the *ESRRB*, *GJB2*, *USH1G*, *CLDN14*, and *SLC26A4* genes and calculated linkage disequilibrium correlations for candidate missense variants in these five genes.

USH1G and *ESRRB* genes have the highest recombination rates and they seem to be in genomic regions considered as hotspots (Table S6). However, most of the rare missense variants found were not clustered and showed a scattered distribution along the different exons with a low recombination rate (Table S7).

DISCUSSION

This study shows that patients with sporadic MD have an enrichment of few rare variants in certain hearing loss genes such as *GJB2*, *SLC26A4*, or *USH1G*. This excess of missense variants in some genes may increase the risk to develop hearing loss in MD and may contribute to explain the heterogeneity observed in the phenotype (Francioli et al., 2015). To understand the relevance of population frequencies in our cohort, we performed the association analysis between variants observed in MD cases against their respective frequencies on a healthy population for each gene of the panel (Lek et al., 2016). From the total amount of variants, we selected rare coding variants for all the targeted genes. We applied a stronger filter for the selection of missense variants by choosing previously described variants for each gene significantly overrepresented in MD cases.

Since many missense variants were not found in the Spanish population from CSVS (Dopazo et al., 2016), a third comparison limited to previously reported variants in CSVS database was carried out. We followed this conservative approach to reduce false positive findings in the readings. This third restrictive gene analysis was limited to 132 variants observed at least once in the Spanish reference population.

From the final analysis, we found that some genes such as *SLC26A4*, *ESRRB*, *CLDN14*, *GJB2*, and *USH1G* retained the higher number of missense variants among Spanish MD patients. We also found one novel synonymous variant in the *MARVELD2* gene in 3 unrelated patients. Besides from its functional implications, it may also generate a cryptic splice site. However, more testing is needed to confirm this finding.

Multiallelic Model for MD

The excess of missense variants in SNHL genes may point to core gene for hearing loss in MD. Our hypothesis is that common cis-regulatory variants and rare variants in one or more genes will contribute to the phenotype in MD. The model will need the additive effect of at least a common and a rare variant in the same gene in a given individual (Castel et al., 2018). In the simplest bi-allelic hypothesis, we will have:

$$Ind\ 1 = cv\ a + rv\ z\ (geneA)$$

$$Ind\ 2 = cv\ b + rv\ y\ (geneB)$$

$$Ind\ 3 = cv\ a + rv\ x\ (geneA)$$

$$Ind\ 4 = cv\ b + rv\ w\ (geneB)$$

Where *cv* is a common variant and *rv* represents a rare variant; however, this model could be more complex for a single gene:

$$Ind\ 1 = cv\ a + cv\ c + rv\ z\ (geneA)$$

$$Ind\ 2 = cv\ b + cv\ d + rv\ z\ (geneB)$$

So, several rare variants will be targeting the core genes (*rv z*, *rv x* for gene A; *rv y*, *rv w* for gene B) and common variants in the same genes will explain variable expressivity of the MD phenotype. Finally, in a more complex scenario, it could involve several genes (oligogenic multiallelic hypothesis):

$$Ind\ n = cv\ a + rv\ z\ (gene\ A) + cv\ b + rv\ y\ (gene\ B)$$

$$+ \dots + cv\ n + rv\ m\ (gene\ N)$$

Gene Panel for Familial MD

The Genomics England project (<https://www.genomicsengland.co.uk/>) has designed gene panels for the diagnosis of many genetic disorders including familial MD (<https://panelapp.genomicsengland.co.uk/panels/394/>). This panel is in an early stage of development because it only considers 130 genes with limited evidence to few families. The results of this study can be used to improve the design of panels for the diagnosis of MD.

For the design of our panel, we chose a total of 69 genes. Most of the genes were selected according to the hearing loss profile (low frequency or pantonal hearing loss). However, more than 90 genes have been related to hearing loss, so more hearing loss genes could be involved in the phenotype (fully accessible from Hereditary Hearing Loss Homepage: <http://hereditaryhearingloss.org/>). Genetic evidence of hearing loss has been obtained from linkage analyses until the emergence of NGS techniques (Shearer et al., 1999), that have facilitated the clinical development of genetic diagnosis in hearing loss. Custom panels and microarrays have been the flags of a new age of discovery of novel and rare variants for genetic diagnostic of hearing loss (Brownstein et al., 2011; Shearer and Smith, 2015).

Our panel was designed considering hearing loss as the main symptom shared by all patients with MD, since the vestibular phenotype and other associated co-morbidities such as migraine or autoimmune disorders are more variable. To improve the diagnostic yield of MD and to decrease this granularity in the phenotype, it will be recommendable to select sporadic patients with an early age of onset for future studies.

Rare Missense Variants in Hearing Loss Genes in Sporadic MD

The frequency of hearing loss related genes is population-specific (Sloan-Heggen et al., 2016). Herein, we present a study for MD patients in the Spanish population. As a part of the study, we consider a panel of genes related to hearing loss and other symptoms. Besides from the validated variants in singletons, only a few rare variants such as *ESRRB* rs201448899:C>T, *MARVELD2* rs369265136:G>A, *SLC26A4* rs200511789:A>C, and *USH1G* rs151242039:C>T have been validated in more than one sporadic case in the entire cohort. All these genes had been previously considered as pathogenic for hearing loss, but they have never been involved with MD.

ESRRB encodes the estrogen-related receptor beta, also known as nuclear receptor subfamily 3, group B, member 2 or NR3B2. This gene encodes for a protein like the estrogen receptor but with a different and unknown role. Mutations in the mouse ortholog have been involved in the placental development and autosomal recessive SNHL (Collin et al., 2008; Weber et al., 2014).

MARVELD2 encodes a protein found in the tight junctions, between epithelial cells. The encoded protein seems to forge barriers between epithelial cells such the ones in the organ of Corti. Defects in this gene are associated with DFNB49 (Mašindová et al., 2015).

SLC26A4 gene encodes pendrin, a protein extensively studied in hearing loss. Its alteration is one of the most common causes of syndromic deafness and autosomal recessive SNHL. It is also associated with enlarged vestibular aqueduct syndrome (EVAS) [(Yang et al., 2007), 36].

USH1G is a gene translating to a protein that contains three ankyrin domains, a class I PDZ-binding motif and a sterile alpha motif. This protein is well-known to interact with harmonin (*USH1C*) in the stereocilia of hair cells, a protein associated with Usher syndrome type 1C (Weil et al., 2003). This protein plays a role in the development and maintenance of the auditory and visual systems and functions in the cohesion of hair bundles formed by inner ear sensory cells. Alterations in the integrity of the protein seem to be the cause of Usher syndrome type 1G (Weil et al., 2003; Miyasaka et al., 2016).

However, *ESRRB* rs201448899:C>T has been observed in more Spanish controls than in global or NFE in ExAC. This increased frequency on the Iberian population when compared with other known largest frequencies as NFE, suggests that this is a population specific variant rather than a MD disease variant. Only the *MARVELD2* rs369265136:G>A variant remains as a proper novel related to MD cases. However, the functional effect of a synonymous variant is unknown and functional studies will be required to decipher the relevancy of this variant in MD cases in the future.

Burden Analysis of Rare Missense Variants in Sporadic MD

Our results demonstrate a burden of rare missense variants in few SNHL genes, including *GJB2*, *ESRRB*, *CLDN14*, *SLC26A4*, and *USH1G*. We speculate that the additive effect of several missense variants in the same gene could interact with the same or other genes at the protein level resulting in the hearing loss phenotype.

Population analysis was addressed in order to obtain a better image of our cohort. Despite the limitation that represents the small number of genes considered in our panel, we have found a significant increase of missense variants on several hearing loss genes in the Iberian population (Table 4). These findings suggest the involvement of multiple missense variants in the same gene and may explain several clinical findings in MD. So, incomplete phenotype found in relatives of patients with familial MD or even the variable expressivity observed could be explained on the differences found in multiple rare variants with additive effect among individuals of the same family (Requena et al., 2015; Martín-Sierra et al., 2016, 2017). In addition, some sporadic cases where a single rare variant with unknown significance cannot explain the phenotype could be singletons individuals with low frequency variants probably following a compound heterozygous recessive pattern of inheritance. Our results start to decipher the complex interaction between rare and ultrarare variations (MAF <0.0001) with common variants in the same or different genes in sporadic MD, adding more evidence to understand the genetic architecture of MD. However, one of the limitations of this study is the lack of availability of a replication cohort with different ethnicity in which to validate these findings.

Another limitation of our dataset is that the method used for resequencing mitochondrial genes may not be able to distinguish mitochondrial from nuclear sequences, as capture panels such as those based in the *Haloplex* technology may sequence all mitochondrial genome fragment replicas that are dispersed throughout the nuclear genome. Hence, variants observed may

not belong to the genes targeted in the mitochondrial genome, but to their pseudogenes in the nuclear genome.

Several hypotheses could explain the excess of missense variants in SNHL genes in MD. First, the variable expressivity of SNHL in MD phenotype, could be the result of additive effect of low frequency or rare variants in the same gene. The combination of low frequency variants in the same gene can be a rare situation, as rare as the disease. As much changes are added to the protein, its integrity could be affected, showing a suboptimal functioning and finally, a loss of function. In our case, *GJB2*, that forms a hexamer with a transmembrane channel function, has been determined as possible affected by these changes in their interactions. Previous studies have determined how certain changes in the monomer can affect to the development of the hexamer hemichannel (Bicego et al., 2006; Jara et al., 2012). Here, bioinformatics models show how the interaction of low frequency variants found in MD patients can impact the interaction between two connexins monomers, but this effect could be amplified in a model including the 6 connexins that form the connexon. However, this hypothesis is difficult to reconcile with the fact that for some small genes such as *GJB2*, complex alleles with several point mutations are exceedingly rare.

A second hypothesis points to the interaction of common and rare variants in one or several genes in the disease phenotype, following its complex disease definition (Becker, 2004; Mitchell, 2012). So, the excess of rare variants will be targeting core genes for hearing loss in MD. In this case, high significant genes in our study could be added to the panel of candidate targets of the disease, although a single variant could not be enough to explain the disease phenotype. So, the interaction between cis-regulatory variants with rare variant in some of our candidate genes and other, a priori, not related SNHL genes could be relevant in the expressivity. *USH1G* interacts with *USH1C*, a known gene involved in Usher syndrome. *USH1G* has been observed to have a minor role in Usher syndrome in Spanish population (Aller et al., 2007), but not in MD, even though they share similar hearing loss profile. Although no one of the missense variants in *USH1G* were in an interaction domain, this could be of interest when considering interaction between different proteins as a main factor to develop a mild phenotype. This hypothesis was reinforced through the data found in familial cases. For instance, the variant rs748718975 in *DPT* gene was only associated with the SNHL phenotype in the family where it was described, but these cases showed different characteristics in the age of onset or hearing loss outcome. These differences between the cases can be explained with other variants found in *KCNQ4* (rs574794136:G>A) and *ADD1* (rs372777117:A>G) genes, although these variants were previously described as variants of unknown significance. So, the variant rs574794136:G>A was found in two sisters with MD, but not in the third one, that was carrier of rs372777117:A>G. This excess of rare variants in certain genes observed in familial cases could explain the differences in expressivity in a given family.

This panel was made as an early screening diagnostic panel. Here we have found that certain SNHL gene variants can be related to MD in the Iberian population and the results show that multiple rare allelic variants in the same gene should be consider

as likely pathogenic. Although there are large differences in the coverage for some genes between the MD panel and the WES databases, these are not the ones with excess of missense variants. Our results will contribute to design a novel gene panel for the genetic diagnosis of MD.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

TR and JL-E as project managers conceived the main idea of the work. Sample preparation and protocol were carry out by TR and AG-M. AG-M performed the bioinformatics and statistical analysis. Validations of tested SNV were done by AG-M and PR-N. AG-M, and JL-E took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis, and manuscript.

FUNDING

This study was funded by FPS-PI0496-2014 and EF-0247-2017 from Consejería de Salud, Spain, 2016-MeniereSociety Grant, UK and Luxembourg National Research Fund (INTER/Mobility/17/11772209).

ACKNOWLEDGMENTS

We acknowledge to all members of the Meniere disease Consortium (MeDiC), a network of clinical and research

centers contributing to the study of Meniere disease. List of participants in MeDiC: Juan Carlos Amor-Dorado (Hospital Can Misses Ibiza, Spain), Ismael Aran (Complejo Hospitalario de Pontevedra, Spain), Angel Batuecas-Caletrío (Hospital Universitario Salamanca, Spain), Jesus Benitez (Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain), Jesus Fraile (Hospital Miguel Servet, Zaragoza, Spain), Ana Garcia-Arumí (Hospital Universitario Vall d'Hebron, Barcelona, Spain), Rocio Gonzalez-A (Hospital Universitario Marqués de Valdecilla, Santander, Spain), Juan M. Espinosa-Sanchez (Hospital Virgen de las Nieves, Granada, Spain), Raquel Manrique Huarte, Nicolas Perez-Fernandez (Clinica Universidad de Navarra, Spain), Pedro Marques (Centro Hospitalar de São João, Porto, Portugal), EM-S, Ricardo Sanz (Hospital Universitario de Getafe, Madrid, Spain), Manuel Oliva Dominguez (Hospital Costa del Sol Marbella, Spain), PP (Hospital Cabueñes, Asturias, Spain), HP-G, VP-G (Hospital La Fe, Valencia, Spain), SS-P, AS-V (Complejo Hospitalario Universitario, Santiago de Compostela, Spain), MT (Instituto Antolí Candela, Madrid, Spain), Roberto Teggi (San Raffaele Scientific Institute, Milan, Italy), and GT (Complejo Hospitalario Badajoz, Spain). We also acknowledge Joaquin Dopazo (Clinical Bioinformatics Area, Hospital Virgen del Rocio, Sevilla, Spain) to facilitate us the access to CSVS datasets. AGM is a PhD student in the Biomedicine program at University of Granada and this work is part of his doctoral thesis. A preprint version of this work has been deposited at Biorxiv repository (Gallego-Martinez et al., 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00076/full#supplementary-material>

REFERENCES

- Aller, E., Jaijo, T., Beneyto, M., Nájera, C., Morera, C., Pérez-Garrigues, H., et al. (2007). Screening of the USH1G gene among Spanish Patients with Usher Syndrome. Lack of mutations and evidence of a minor role in the pathogenesis of the syndrome. *Ophthalm. Genet.* 28, 151–155. doi: 10.1080/13816810701537374
- Arweiler-Harbeck, D., Horsthemke, B., Jahnke, K., and Hennies, H. C. (2011). Genetic aspects of Familial Ménière's disease. *Otol. Neurotol.* 32, 695–700. doi: 10.1097/MAO.0b013e318216074a
- Becker, K. G. (2004). The common variants/multiple disease hypothesis of common complex genetic disorders. *Med. Hypotheses* 62, 309–317. doi: 10.1016/S0306-9877(03)00332-3
- Bicego, M., Beltramello, M., Melchionda, S., Carella, M., Piazza, V., Zelante, L., et al. (2006). Pathogenetic role of the deafness-related M34T mutation of Cx26. *Hum. Mol. Genet.* 15, 2569–2587. doi: 10.1093/hmg/ddl184
- Brownstein, Z., Friedman, L. M., Shahin, H., Oron-Karni, V., Kol, N., Abu Rayyan, A., et al. (2011). Targeted genomic capture and massively parallel sequencing to identify genes for hereditary hearing loss in middle Eastern Families. *Genome Biol.* 12:R89. doi: 10.1186/gb-2011-12-9-r89
- Cabrera, S., Sanchez, E., Requena, T., Martinez-Bueno, M., Benitez, J., Perez, N., et al. (2014). Intronic variants in the NFKB1 gene may influence hearing forecast in patients with unilateral sensorineural hearing loss in Meniere's disease. *PLoS ONE* 9:e112171. doi: 10.1371/journal.pone.0112171
- Castel, S. E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., et al. (2018). Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.* 50, 1327–1334. doi: 10.1038/s41588-018-0192-y
- Caulley, L., Quimby, A., Karsh, J., Ahrari, A., Tse, D., and Kontorinis, G. (2017). *Autoimmune Arthritis in Ménière's Disease: A Systematic Review of the Literature*. *Semin. Arthritis Rheum.* 48, 141–147. doi: 10.1016/j.semarthrit.2017.11.008
- Collet, A., Tarabeux, J., Girard, E., Enghien, C. D. D., Golmard, L., and Deshaies, V., Lermine, A., et al. (2015). Pros and cons of haloplex enrichment in cancer predisposition genetic diagnosis. *Genetics* 2, 263–280. doi: 10.3934/genet.2015.4.263
- Collin, R. W., Kalay, E., Tariq, M., Peters, T., van der Zwaag, B., Venselaar, H., et al. (2008). Mutations of ESRRB encoding estrogen-related receptor beta cause autosomal-recessive nonsyndromic hearing impairment DFNB35. *Am. J. Hum. Genet.* 82, 125–138. doi: 10.1016/j.ajhg.2007.09.008
- Dalol, A., Daghistani, K., Elaimi, A., Al-Wazani, W. A., Bamanie, A., Safiah, M., et al. (2016). Utilization of amplicon-based targeted sequencing panel for the massively parallel sequencing of sporadic hearing impairment patients from Saudi Arabia. *BMC Med. Genet.* 17(Suppl. 1):67. doi: 10.1186/s12881-016-0329-8

- Dopazo, J., Amadoz, A., Bleda, M., Garcia-Alonso, L., Alemán, A., García-García, F., et al. (2016). 267 Spanish exomes reveal population-specific differences in disease-related genetic variation. *Mol. Biol. Evol.* 33, 1205–1218. doi: 10.1093/molbev/msw005
- Francioli, L. C., Polak, P. P., Koren, A., Menelaou, A., Chun, S., Renkens, I., et al. (2015). Genome-wide patterns and properties of *de novo* mutations in humans. *Nat. Genet.* 47, 822–826. doi: 10.1038/ng.3292
- Gallego-Martinez, A., Requena, T., Roman-Naranjo, P., and J. A., Lopez-Escamez (2018). Excess of rare missense variants in hearing loss genes in sporadic meniere disease. *BioRxiv [preprint]* 393322. doi: 10.1101/393322
- Gazquez, I., Soto-Varela, A., Aran, I., Santos, S., Batuecas, A., Trinidad, G., et al. (2011). High prevalence of systemic autoimmune diseases in patients with Meniere's disease. *PLoS ONE* 6:e26759. doi: 10.1371/journal.pone.0026759
- Gu, X., Guo, L., Ji, H., Sun, S., Chai, R., Wang, L., et al. (2015). Genetic testing for sporadic hearing loss using targeted massively parallel sequencing identifies 10 novel mutations. *Clin. Genet.* 87, 588–593. doi: 10.1111/cge.12431
- Hietikko, E., Kotimäki, J., Kentala, E., Klockars, T., Sorri, M., and Männikkö, M. (2011). Finnish Familial Meniere disease is not linked to chromosome 12p12.3, and anticipation and cosegregation with migraine are not common findings. *Genet. Med.* 13, 415–420. doi: 10.1097/GIM.0b013e3182091a41
- Hietikko, E., Kotimäki, J., Sorri, M., and Männikkö, M. (2013). High incidence of Meniere-like symptoms in relatives of meniere patients in the areas of oulu university hospital and kainuu central hospital in Finland. *Eur. J. Med. Genet.* 56, 279–285. doi: 10.1016/j.ejmg.2013.03.010
- Jara, O., Acuña, R., García, I. E., Maripillán, J., Figueroa, V., Sáez, J. C., et al. (2012). Critical role of the first transmembrane domain of Cx26 in regulating oligomerization and function. *Mol. Biol. Cell* 23, 3299–3311. doi: 10.1091/mbc.e11-12-1058
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536 (7616): 285–291. doi: 10.1038/nature19057
- Lempert, T., Olesen, J., Furman, J., Waterston, J., Seemungal, B., Carey, J., et al. (2012). Vestibular migraine: diagnostic criteria. *J. Vestib. Res.* 22, 167–172. doi: 10.3233/VES-2012-0453
- Lionel, A. C., Costain, G., Monfared, N., Walker, S., Reuter, M. S., Hosseini, S. M., et al. (2018). Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome sequencing as a first-tier genetic test. *Genet. Med.* 20, 435–443. doi: 10.1038/gim.2017.119
- Lopez-Escamez, J. A., Carey, J., Chung, W. H., Goebel, J. A., Magnusson, M., Mandalà, M., et al. (2015). Diagnostic criteria for Meniere's disease. *J. Vestib. Res.* 25, 1–7. doi: 10.3233/VES-150549
- Martín-Sierra, C., Gallego-Martinez, A., Requena, T., Frejo, L., Batuecas-Caletrío, A., and Lopez-Escamez, J. A. (2017). Variable expressivity and genetic heterogeneity involving DPT and SEMA3D genes in autosomal dominant familial Meniere's disease. *Eur. J. Hum. Genet.* 25, 200–207. doi: 10.1038/ejhg.2016.154
- Martín-Sierra, C., Requena, T., Frejo, L., Price, S. D., Gallego-Martinez, A., Batuecas-Caletrío, A., et al. (2016). A novel missense variant in PRKCB segregates low-frequency hearing loss in an autosomal dominant family with Meniere's disease. *Hum. Mol. Genet.* 25, 3407–3415. doi: 10.1093/hmg/ddw183
- Mašindová, I., Šoltýsová, A., Varga, L., Mátyás, P., and Ficek, A., Hučková, M., et al. (2015). MARVELD2 (DFNB49) mutations in the hearing impaired central european roma population - prevalence, clinical impact and the common origin. *PLoS ONE* 10:e0124232. doi: 10.1371/journal.pone.0124232
- Meyer, M. J., Beltrán, J. F., Liang, S., Fragoza, R., Rumack, A., Liang, J., et al. (2018). Interactome INSIDER: a structural interactome browser for genomic studies. *Nat. Methods* 15, 107–114. doi: 10.1038/nmeth.4540
- Mijovic, T., Zeitouni, A., and Colmegna, I. (2013). Autoimmune sensorineural hearing loss: the otology–rheumatology interface. *Rheumatology* 52, 780–789. doi: 10.1093/rheumatology/ket009
- Mitchell, K. J. (2012). What is complex about complex disorders? *Genome Biol.* 13:237. doi: 10.1186/gb-2012-13-1-237
- Miyasaka, Y., Shitara, H., Suzuki, S., Yoshimoto, S., Seki, Y., Ohshiba, Y., et al. (2016). Heterozygous mutation of Ush1g/Sans in mice causes early-onset progressive hearing loss, which is recovered by reconstituting the strain-specific mutation in Cdh23. *Hum. Mol. Genet.* 25, 2045–2059. doi: 10.1093/hmg/ddw078
- Morales Angulo, C., Gómez Castellanos, R., García Mantilla, J., Bezos Capelastegui, J. T., and Carrera, F. (2003). Epidemiology of Meniere's disease in Cantabria. *Acta Otorrinolaringol. Esp.* 54, 601–605. doi: 10.1016/S0001-6519(03)78456-8
- Paparella, M. M., and Griebie, M. S. (1984). Bilaterality of Meniere's disease. *Acta Oto Laryngol.* 97, 233–37.
- Requena, T., Cabrera, S., Martín-Sierra, C., Price, S. D., Lysakowski, A., and Lopez-Escamez, J. A. (2015). Identification of two novel mutations in FAM136A and DTNA genes in autosomal-dominant familial Meniere's disease. *Hum. Mol. Genet.* 24, 1119–1126. doi: 10.1093/hmg/ddu524
- Requena, T., Espinosa-Sanchez, J. M., Cabrera, S., Trinidad, G., Soto-Varela, A., Santos-Perez, S., et al. (2014a). Familial clustering and genetic heterogeneity in Meniere's disease. *Clin. Genet.* 85, 245–252. doi: 10.1111/cge.12150
- Requena, T., Espinosa-Sanchez, J. M., and Lopez-Escamez, J. A. (2014b). Genetics of dizziness: cerebellar and vestibular disorders. *Curr. Opin. Neurol.* 27, 98–104. doi: 10.1097/WCO.0000000000000053
- Requena, T., Gazquez, I., Moreno, A., Batuecas, A., Aran, I., Soto-Varela, A., et al. (2013). Allelic variants in TLR10 Gene may influence bilateral affection and clinical course of Meniere's disease. *Immunogenetics* 65, 345–355. doi: 10.1007/s00251-013-0683-z
- Ruiz-Pesini, E., Lott, M. T., Procaccio, V., Poole, J. C., Brandon, M. C., Mishmar, D., et al. (2007). An enhanced MITOMAP with a global MtDNA mutational phylogeny. *Nucleic Acids Res.* 35, D823–D828. doi: 10.1093/nar/gkl927
- Shearer, A. E., Hildebrand, M. S., and Smith, R. J. H. (1999). "Hereditary hearing loss and deafness overview," in *GeneReviews®*, eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, and A. Amemiya (Seattle, WA: University of Washington). Available online at: <http://www.ncbi.nlm.nih.gov/books/NBK1434/> (Accessed December 20, 2018).
- Shearer, A. E., and Smith, R. J. (2015). Massively parallel sequencing for genetic diagnosis of hearing loss: the new standard of care. *Otolaryngol. Head Neck Surg.* 153, 175–182. doi: 10.1177/0194599815591156
- Sloan-Heggen, C. M., Bierer, A. O., Shearer, A. E., Kolbe, D. L., Nishimura, C. J., Frees, K. L., et al. (2016). Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum. Genet.* 135, 441–450. doi: 10.1007/s00439-016-1648-8
- The Molecular Otolaryngology and Renal Research Laboratories, The University of Iowa. *Deafness Variation Database*. Available online at: <http://deafnessvariationdatabase.org/> (Accessed June 30, 2018).
- Van Camp, G., and Smith, R. J. H. (2018). *Hereditary Hearing Loss Homepage*. Available online at: <https://hereditaryhearingloss.org> (Accessed December 20, 2018).
- Weber, M. L., Hsin, H. Y., Kalay, E., Brožková, D. S., Shimizu, T., Bayram, M., et al. (2014). Role of Estrogen Related Receptor Beta (ESRRB) in DFNB35 hearing impairment and dental decay. *BMC Med. Genet.* 15:81. doi: 10.1186/1471-2350-15-81
- Weil, D., El-Amraoui, A., Masmoudi, S., Mustapha, M., Kikkawa, Y., Lainé, S., et al. (2003). Usher syndrome type I G (USH1G) is caused by mutations in the gene encoding SANS, a protein that associates with the USH1C protein, harmonin. *Hum. Mol. Genet.* 12, 463–471. doi: 10.1093/hmg/ddg051
- Yang, T., Vidarsson, H., Rodrigo-Blomqvist, S., Rosengren, S. S., Enerback, S., and Smith, R. J. (2007). Transcriptional control of SLC26A4 is involved in pendred syndrome and nonsyndromic enlargement of vestibular aqueduct (DFNB4). *Am. J. Hum. Genet.* 80, 1055–1063. doi: 10.1086/518314

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gallego-Martinez, Requena, Roman-Naranjo and Lopez-Escamez. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Next Generation Sequencing and Animal Models Reveal *SLC9A3R1* as a New Gene Involved in Human Age-Related Hearing Loss

Giorgia Giotto^{1,2*}, Anna Morgan^{1,2}, Navaneethakrishnan Krishnamoorthy^{3,4}, Massimiliano Cocca², Marco Brumat^{1,2}, Sissy Bassani^{1,2}, Martina La Bianca², Mariateresa Di Stazio^{1,2} and Paolo Gasparini^{1,2}

¹ Department of Medicine, Surgery and Health Sciences, University of Trieste, Trieste, Italy, ² Institute for Maternal and Child Health – IRCCS “Burlo Garofolo”, Trieste, Italy, ³ Sidra Medical and Research Center, Doha, Qatar, ⁴ Heart Science Centre, National Heart and Lung Institute, Imperial College London, London, United Kingdom

OPEN ACCESS

Edited by:

Mike Mikailov,
U.S. Food and Drug Administration,
United States

Reviewed by:

Theodora Katsila,
University of Patras, Greece
GuangJun Zhang,
Purdue University, United States

*Correspondence:

Giorgia Giotto
giorgia.giotto@burlo.trieste.it

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 21 August 2018

Accepted: 11 February 2019

Published: 26 February 2019

Citation:

Giotto G, Morgan A, Krishnamoorthy N, Cocca M, Brumat M, Bassani S, La Bianca M, Di Stazio M and Gasparini P (2019) Next Generation Sequencing and Animal Models Reveal *SLC9A3R1* as a New Gene Involved in Human Age-Related Hearing Loss. *Front. Genet.* 10:142. doi: 10.3389/fgene.2019.00142

Age-related hearing loss (ARHL) is the most common sensory impairment in the elderly affecting millions of people worldwide. To shed light on the genetics of ARHL, a large cohort of 464 Italian patients has been deeply characterized at clinical and molecular level. In particular, 46 candidate genes, selected on the basis of genome-wide association studies (GWAS), animal models and literature updates, were analyzed by targeted re-sequencing. After filtering and prioritization steps, *SLC9A3R1* has been identified as a strong candidate and then validated by “*in vitro*” and “*in vivo*” studies. Briefly, a rare (MAF: 2.886e-5) missense variant c.539G > A, p.(R180Q) was detected in two unrelated male patients affected by ARHL characterized by a severe to profound high-frequency hearing loss. The variant, predicted as damaging, was not present in healthy matched controls. Protein modeling confirmed the pathogenic effect of p.(R180Q) variant on protein’s structure leading to a change in the total number of hydrogen bonds. *In situ* hybridization showed *slc9a3r1* expression in zebrafish inner ear. A zebrafish knock-in model, generated by CRISPR-Cas9 technology, revealed a reduced auditory response at all frequencies in *slc9a3r1*^{R180Q/R180Q} mutants compared to *slc9a3r1*^{+/+} and *slc9a3r1*^{+/R180Q} animals. Moreover, a significant reduction (5.8%) in the total volume of the saccular otolith (which is responsible for sound detection) was observed in *slc9a3r1*^{R180Q/R180Q} compared to *slc9a3r1*^{+/+} ($P = 0.0014$), while the utricular otolith, necessary for balance, was not affected in agreement with the human phenotype. Overall, these data strongly support the role of *SLC9A3R1* gene in the pathogenesis of ARHL opening new perspectives in terms of diagnosis, prevention and treatment.

Keywords: hearing loss, new gene discovery, zebrafish model, CRISPR-Cas9, next-generation sequencing

INTRODUCTION

Age-related Hearing loss (ARHL) is the predominant sensory impairment in the elderly (Huang and Tang, 2010; Kidd Iii and Bao, 2012), affecting millions of people worldwide (National Health and Nutrition Examination Survey [NHANES], 2015). Moreover, projections suggest that, in the next decade, the number of patients will double, largely due to an increased lifespan (Bowl and Dawson, 2015).

Age-related Hearing loss is characterized by bilateral and progressive hearing loss that usually starts at the high frequencies (Huang and Tang, 2010) causing communication difficulties associated with cognitive decline, social isolation and depression (Bowl and Dawson, 2015). It is a complex disease in which genetic and environmental risk factors (i.e., noise, smoking, alcohol, etc.) interplay (Bovo et al., 2011; Ohgami et al., 2013; Vuckovic et al., 2013, 2014). To date, it is still unclear whether ARHL is caused by (1) rare Mendelian gene variants with large effect size or (2) multiple variants each contributing to the disease. Despite significant research efforts carried out during the last decade, the genetics risk factors involved in ARHL are still mainly unknown. To date, only few ARHL susceptibility genes have been detected by genome-wide association studies (GWAS) (Friedman et al., 2009; Van Laer et al., 2010; Giroto et al., 2011, 2014; Wolber et al., 2014; Vuckovic et al., 2015), and thus there is a strong need to plan for new research activities aimed at understanding the molecular mechanisms underlying this diseases and to define possible targets for therapeutic and preventive plans.

The use of animal models is a powerful tool for the discovery and/or the validation of human disease-genes. In particular, zebrafish has become an attractive model for the study of the development and function of the vertebrate inner ear (Whitfield et al., 2002). Although the zebrafish ear does not contain an equivalent of the mammalian cochlea, many features (e.g., the organization and morphology of the supporting cells and hair cells, the mechanical stimulation of the ear, etc.) are conserved and analogous with other species (Haddon and Lewis, 1996; Nicolson, 2005; Abbas and Whitfield, 2009; Baxendale and Whitfield, 2016). Moreover, it has been shown that approximately 70% of human genes have, at least, one zebrafish ortholog (Howe et al., 2013) and a number of genes required for the hearing function in zebrafish have been also associated with auditory defects in mammals (i.e., mice and humans) further supporting their role across different species (Coimbra et al., 2002).

Starting from these considerations, we developed a combined strategy, based on: (1) Targeted Re-Sequencing (TRS) of a panel of 46 ARHL candidate genes (Morgan et al., 2018) in 464 patients, followed by (2) zebrafish models of the most interesting variants identified.

Here, we report the results obtained using this approach demonstrating the presence of a *SLC9A3R1* (also named Na^+/H^+ Exchange Regulatory Cofactor, *NHERF1*) pathogenic variant in two unrelated ARHL patients and finally that this variant has deleterious effects in a zebrafish knock-in (KI) model. These findings, together with a previous study demonstrating the involvement of *Slc9a3r1* in hearing loss in

mouse (Kamiya et al., 2014), allow us to propose *SLC9A3R1* as an extremely promising new ARHL candidate gene.

MATERIALS AND METHODS

Ethics Statement

Human

The study was reviewed and approved by the Ethics Committee of the Burlo Garofolo children's hospital in Trieste (Italy) (2007 242/07). Written informed consent was obtained from each participant to the study and all the research was carried out according to the ethical standards defined by the Helsinki declaration.

Zebrafish

Zebrafish procedures were performed in accordance with Spanish and European laws, guidelines and policies for animal (Real Decreto 1201/05 (BOE 252, October 21, 2005 and European Directive 2010/63/EU on the protection of animals used for scientific purposes of October 20, 2010). This project was approved by the institutional Ethical Committee for Animal Experimentation of the PRBB, where ZeClinics conducts all experimental work (approval number CEA-OH/9421/2).

Patients Recruitment

A large cohort of 464 ARHL patients coming from inbred (Friuli Venezia Giulia-FVG- Cohort, Carlantino Cohort) and outbred (Milan, Naples, Trieste) Italian populations was analyzed by TRS. The cohort consists of 258 males and 206 females, all aged over 50 and with high frequencies bilateral HL developed around the 5th decade of life [high-frequency pure tone average (PTAH) > 40dB]. No disorders in the external and middle ear, as well as any vestibular problem or syndromic features were present.

Furthermore, 350 healthy individuals (all aged ≥ 50 y.o with PTAH ≤ 25 dB) coming from five out of six villages of the FVG cohort (Resia, Sauris, Clauzetto, San Martino del Carso, Illegio) were used as an internal control.

Targeted Re-sequencing

A total number of 46 ARHL-candidate genes, including *SLC9A3R1*, were sequenced using Ion Torrent PGMTM (Life-Technologies). Genes were selected according to data from (a) GWAS meta-analyses on isolated and outbred populations of European and Asian ancestry (Giroto et al., 2011, 2014; Wolber et al., 2014; Vuckovic et al., 2015); (b) literature updates and (c) animal models (Morgan et al., 2018).

Briefly, 10 ng of genomic DNA were used to construct DNA libraries using Ion AmpliSeq Library Kit 2.0 (Life Technologies). Template Ion Sphere Particles were prepared using Ion PGM Template OT2 200 kit and a single end 200 base-read sequencing run was carried out using Ion PGM sequencing 200 kit v2 (Life Technologies), on Ion Torrent PGMTM (Life Technologies). Ten indexed patients' libraries were sequenced simultaneously on each Ion 318 Chip. Sequencing data were then analyzed according to the Ion Torrent SuiteTM v3.6. The annotated

SNVs\INDELS were evaluated according to several *in silico* predictor tools (SIFT, Polyphen2, MutationTaster, LRT) (Ng and Henikoff, 2003; Chun and Fay, 2009; Schwarz et al., 2010; Adzhubei et al., 2013) and conservation across species (PhyloP) (Pollard et al., 2010). Variants were classified as ultra-rare (MAF < 0.001), rare (MAF < 0.01) or common (MAF > 0.01) based on the frequencies reported NCBI dbSNP build142¹ as well as in 1000 Genomes Project², NHLBI Exome Sequencing Project (ESP) Exome Variant Server³, ExAC Browser⁴ and gnomAD browser⁵. Finally, those variants most likely to be disease causing (i.e., rare and ultra-rare variants predicted as damaging by all *in silico* predictor tools) were analyzed by Sanger sequencing and tested in controls.

SNP Genotyping

SLC9A3R1 variant was checked in 350 healthy controls using a TaqMan SNP genotyping assay (Assay ID: C_164690872_10, Thermo Fisher Scientific). Reactions were performed according to the manufacturer's instructions. Data were analyzed with the Taqman Genotyping Software (Thermo Fisher Scientific).

Mutational Protein Modeling

The 3D structure of PDZ2 domain of the NHERF1 protein was retrieved from protein data bank [ID: 2KRG, (Bhattacharya et al., 2010)]. This structure was considered as wild type (WT) and used in discovery studio [DS, (Accelrys Inc., San Diego, CA, United States)] to produce a mutant model (R180Q) as previously described (Krishnamoorthy et al., 2011).

Molecular Dynamics Simulations

The 3D structures of WT and mutant were used in the Groningen machine for chemical simulations (GROMACS) to perform molecular dynamics (MD) simulations. The atoms of the systems were prepared with GROMOS96 force field (van Gunsteren, 1996; Van Der Spoel et al., 2005; Hess et al., 2008). To solvate the protein, the SPC3 water model was used within a cubic box sized 1.5 nm (Berendsen et al., 1981). The counter ions were added to neutralize the systems and the periodic boundary conditions were applied in all directions. The prepared systems consist of 30202 and 30204 atoms in total. The LINCS (Hess et al., 1997) algorithm was used to constrain all the bond lengths and the SETTLE (Miyamoto and Kollman, 1992) algorithm was applied to constrain the geometry of the water molecules. A twin range cut-off was used: 0.8 nm for Van der Waals and 1.4 nm for electrostatic interactions, for managing long range interactions. The steepest descent algorithm was applied to energy minimize the systems with the tolerance of 2000 Kj/mol/nm. Subsequently, these structures were pre-equilibrated with 100 ps simulation before performing the production MD simulation for 25 ns with a time-step of 2 fs at constant temperature (300 K), pressure (1 atm) and number of particles, without any position restraints

(Berendsen et al., 1984). The structures were collected at regular interval for every 100 ps to trace the trajectories and the tools in PyMOL⁶, DS and GROMACS were utilized for analyzing the structures and their molecular interactions.

Cluster Analysis and Surface Map

To represent the structure from the MD simulations, each trajectory with 2500 structures were classified into clusters based on their structural deviations. A structure from the top ranked cluster was chosen for representation as it was frequently occurring conformation. The representative structures were further used in DS for mapping the surface and to analyze the modifications on the binding surface (potential sites for binding partners).

In vitro Molecular Cloning

The impact of the identified mutation on mRNA and protein levels was tested by transient transfection in HEK 293 cells using expression clones containing either the wild-type (WT) or the mutant cDNA. cDNAs were cloned into a pCMV6-Entry vector (Origene, Rockville, MD, United States), Myc-tagged.

The calcium phosphate transfection method was used (Kingston et al., 2003). Forty-eight hours after transfection total cell proteins and RNAs were prepared and analyzed by Western Blot (WB) and quantitative Real Time PCR (qRT-PCR), respectively.

Western Blot Analysis

For protein analysis, HEK 293 cells were lysed in IPLS buffer (50 mM Tris-HCL pH7.5, 120 mM NaCl, 0.5 mM EDTA and 0.5% Nonidet P-40) supplemented with proteases inhibitors (Roche). After sonication and pre-clearing, protein lysate concentration was determined by Bradford Assay (Bio-Rad). An 8% polyacrylamide gel was used for protein electrophoresis. After blotting, membranes were blocked with 5% skim milk in Tris-buffered saline, 0.1% Tween 20 (TBST) and then incubated with primary c-Myc Antibody 9E10 monoclonal (Santa Cruz) overnight. Secondary antibodies [anti-mouse antibody (Santa Cruz)] were diluted in blocking buffer and incubated with the membranes for 45 min at room temperature. Proteins were detected with the ECL detection kit (GE Health Care Bio-Sciences).

Housekeeping proteins (e.g., β -actin or Hsp90) were used as an internal control for protein loading as well as for reference in the WB analysis.

Quantitative Real-Time PCR (qRT-PCR)

RNA was extracted from cell pellets using High Pure RNA isolation Kit (Roche). Total RNA (1 μ g) was reverse transcribed to cDNA using Transcriptor First Strand cDNA Synthesis kit (Roche). qRT-PCR was performed using standard PCR conditions in a 7900HT Fast Real Time PCR System (Applied Biosystems) (i.e., 95°C for 10', 40 cycles of 95°C for 15'' and 60°C for 1', followed by a dissociation stage of 95°C for 15'', 60°C for 15'' and 95°C for 15'') with Power SYBR Green PCR Master Mix

¹<https://www.ncbi.nlm.nih.gov/projects/SNP/>

²<http://www.1000genomes.org/>

³<http://evs.gs.washington.edu>

⁴<http://exac.broadinstitute.org/about>

⁵<http://gnomad.broadinstitute.org/>

⁶www.pymol.org

(Thermo Fisher Scientific). Gene-specific primers were designed by using Primer3Web software⁷. All experiments were performed in biological triplicate. Expression levels have been standardized to Neo gene expression and all data have been analyzed using the $2^{-\Delta\Delta CT}$ Livak Method (Livak and Schmittgen, 2001).

Zebrafish Husbandry

Zebrafish KI lines used in this study were generated, analyzed phenotypically and maintained at ZeClinics (Barcelona, Spain). Two different zebrafish transgenic lines have been used as a background for the KI lines generation:

- (1) Tg(brn3c:mGFP), which is a stable reporter line where a GFP tagged to the membrane is expressed under the control of the brn3c promoter (Xiao et al., 2005), thus specifically labeling a subset of retinal ganglion cells, hair cells of the inner ear and the lateral line neuromasts (Di Donato et al., 2013).
- (2) Tg(isl3:GFP), which is a stable reporter line where a GFP is expressed in the cytoplasm of sensory neurons under the control of the isl3 promoter (Pittman et al., 2008).

Embryos were maintained in petri dishes at 28.5 Celsius. Developmental stages were evaluated as hours and days post-fertilization (hpf and dpf). All zebrafish experimental protocols were approved by the Generalitat de Catalunya.

Gene Expression in Zebrafish Larvae

Gene expression in Zebrafish larvae (5 dpf) was performed by whole mount *in situ* hybridization (ISH). At this stage, inner ear is developed to allow functional hearing, which implies hair cells involved in that function are mature. Synthesis of antisense RNA and whole-mount *in situ* hybridization were performed as previously described (Thisse et al., 2004).

Specific riboprobes intended to recognize *slc9a3r1* mRNA were designed (Slc9a3r1_Fw: ATGTCCAGCGACCTCAGGCC; Slc9a3r1_Rv_T7: TAATACGACTCACTATAGGGCTCCAGTCC ATCTGCGGAGCTC).

cDNAs were amplified by PCR, using Expand High Fidelity PCR System (Roche), from a custom Zebrafish cDNAs library obtained by RT-PCR from (SuperScript III Rev Transcript kit, Invitrogen) a mRNA pool coming from 5 dpf Zebrafish larvae [Trizol mRNA extraction protocol, Trizol (Sigma-Aldrich)]. A T7 sequence linker in reverse primers was included to directly use the synthesized PCR products as templates to amplify the reverse digoxigenin-labeled riboprobe to be used for ISH.

Once dissected, embryos were fixed in 4% paraformaldehyde (PFA) over night (O/N), and then dehydrate with increasing concentrations of Metanol 0,1% Tween20 in PBS (PBT) for long-term storage (25, 50, and 75%). Embryos were then rehydrated and treated with proteinase K (Sigma-Aldrich). Afterwards embryos were incubated with a hybridization mix [i.e., 50% Formamide, 5× saline-sodium citrate (SSC), 0.1% Tween 20, Citric acid to adjust HM to pH 6.0 (460 µl of 1 M citric acid for 50 ml of mix), 50 µg/ml heparin, 500 µg/ml

tRNA] containing the riboprobe (dilution: 1:100) and then with antibody against digoxigenin [Peroxidase-conjugated anti-DIG Fab (Roche), dilution 1:1000]. Nitroblue Tetrazolium (NBT: 3,5 µl of 50 mg/ml NBT solution in 1 ml of alkaline Tris buffer) was used with the alkaline phosphatase substrate 5-Bromo- 4-Chloro-3-Indolyl Phosphate (BCIP: 4,5 µl of 50 mg/ml BCIP solution in 1 ml of alkaline Tris buffer).

In order to check gene expression in hair cells, marked with GFP through its activation by the Brn3c promoter, secondary antibody against GFP was used (rabbit anti-GFP, Torrey Pinnes; 1:400).

Stained embryos were processed for imaging through 2 different methods:

- (1) Whole embryos were imaged with a bright field stereoscope to determine the overall expression pattern.
- (2) Transversal sections were acquired through a cryostat (20 µm width) to determine precisely the mRNA location within the inner ear across three different anterioposterior positions. Gene expression pattern was compared with the position of hair cells, marked with GFP through its activation by the Brn3c promoter. Sections were imaged with Leica DM5 light/fluorescent microscope at 40× magnification.

Generation of *slc9a3r1* Knock-In (KI) Zebrafish Line

To generate the KI, we designed a strategy where double-strand breaks (DSBs), generated by CRISPR technology, are repaired through homologous recombination (HR) DNA repair mechanism (Cong et al., 2013). HR has been enhanced through the co-injection of Rad51 protein (Takayama et al., 2017) and providing a single-stranded oligonucleotide (ssOligo) that contains two homologous arms flanking the desired mutation.

Synthesis of *slc9a3r1* sgRNA

After assessing Slc9a3r1 protein conservation with its human orthologous, single guide RNAs (sgRNAs) were synthesized *in vitro* from double strand DNA that contained a T7 promoter region, a specific CRISPR recognition site (20N spacer) and a guide sequence that allows recognition by CAS9. This double strand DNA was synthesized by the partial annealing of a constant oligo (80 bps) and a gene-specific oligo (60 bps). After annealing, DNA was fill-in by T4 polymerase (Gene-specific oligo: TAATACGACTCACTATA-N20-GTTTTAGAGCTAGAAATAGCAAG; Constant oligo: AA AAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGG ACTAGCCTTATTTAACTTGCTATTCTAGCTCTAAAC).

CRISPR Injection and Screening

Tg(brn3c:mGFP) or Tg(isl3:GFP) embryos were injected with a mix of sgRNA, Cas9 (mRNA), ssOligo and Rad51 human Protein. 20–25 embryos of 48 hpf were selected and pooled for genomic DNA extraction. Genotyping was performed using specific diagnostic primers, a pair of primers to amplify the surrounding area (SLC9A3R1_Fw/SLC9A3R1_Rv = 379 bps; SLC9A3R1_Fw: CCCTTTGTAGGACTCGAAGAACGAG; SLC

⁷<http://bioinfo.ut.ee/primer3/>

9A3R1_Rv: TGTTCCTCAAATAAGCCAGAGCAGAAC) and one primer to discriminate the mutated region (SLC9A3R1_KI_Fw/SLC9A3R1_Rv = 278 bps; SLC9A3R1_KI_Fw: CTTCTTGCAGAAATGGTTGTTG) (Figure 1C).

The remaining injected animals were grown until sexual maturity. F1 animals were Sanger sequenced to identify positive targeted mutants generated through CRISPR KI. These animals were incrossed to obtain F2 homozygous larvae to perform their phenotypical analysis.

Genotyping

Genomic DNA of single larvae from the different experiment was extracted using Extract-N-Amp™ Tissue PCR Kit (Sigma). For the *Slc9a3r1* KI a first PCR reaction was performed using the primers SLC9A3R1_Fw and SLC9A3R1_Rv and the following conditions: 94°C 2 min, 25× (94°C 15 s, 58°C 30 s, 72°C 1:10 min), 72°C 7 min, 4°C hold. Then, a nested PCR reaction was performed using the primers SLC9A3R1_Fw, SLC9A3R1_Rv, SLC9A3R1_INT_MUT_Fw (5'-CTCATCCACCGCCCTGATG-3'), SLC9A3R1_INT_WT_Rv (5'-CCGGCCAGTATATCAAGCTGTC-3') and the same thermal protocol.

slc9a3r1 KI Sound Response

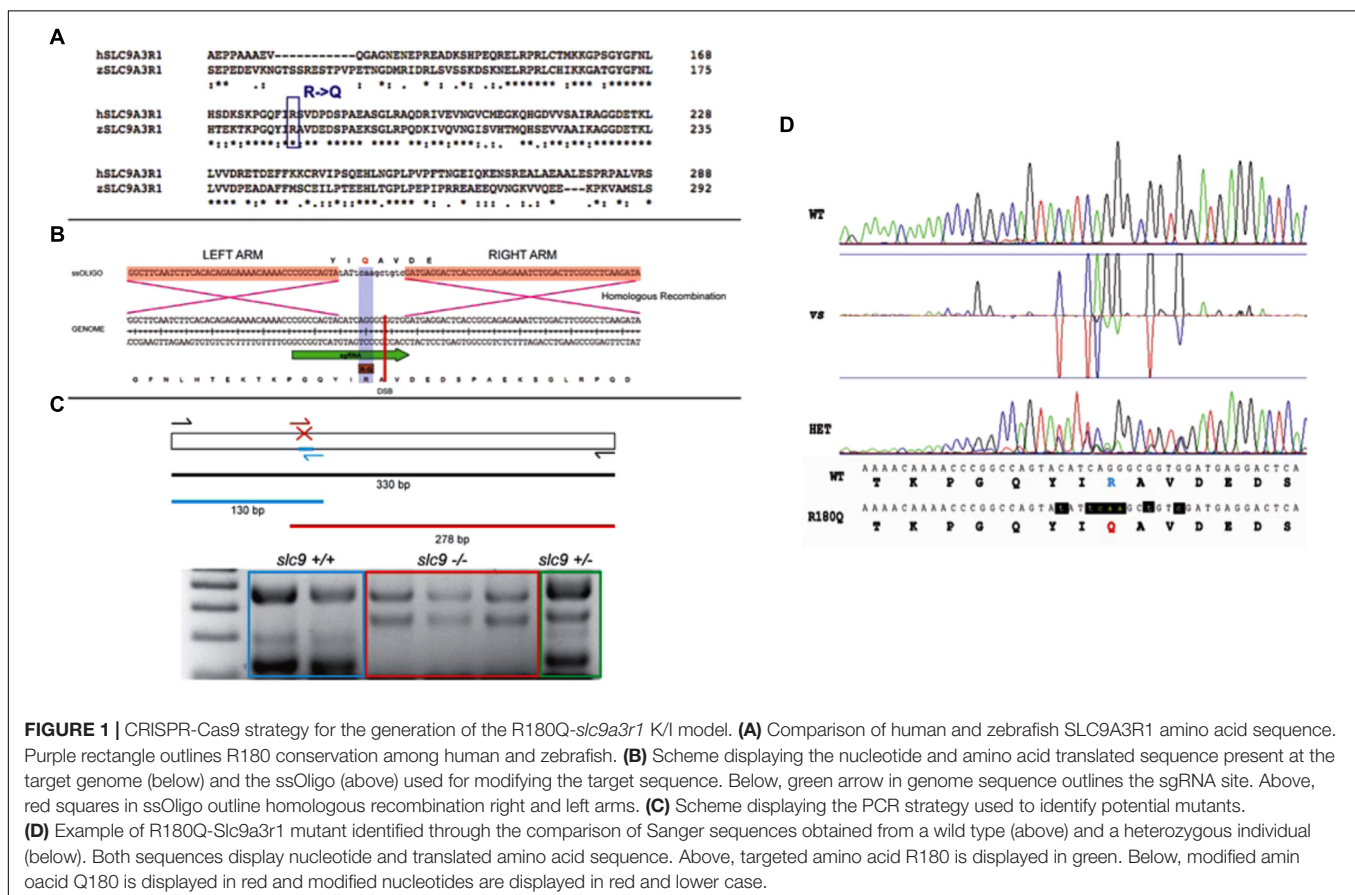
Six day-old larvae obtained by pairwise mating of adult tg(brn3c:mGFP;R180Q-*slc9a3r1*^{+/R180Q}) were tested to

determine their ability to respond to sound stimuli at different frequencies, measuring the acoustic startle response (Bhandiwad et al., 2013). The EthoVision XT 12 software and the DanioVision device from Noldus Information Technologies (Wageningen, Netherlands) were used. This closed system consists of a camera placed above a chamber with circulating water and a temperature sensor that is set at 28°C. A 96-wells plate is placed in the chamber, which can provide different stimuli (light/dark environment, tapping, sound) controlled by the software. For each sound frequency tested, larvae were left for 10 min in dark for acclimation, 2 min with the light on (natural locomotor behavior of zebrafish is active in dark and immobile in light) and then the sound stimulus was provided. Finally, a tapping was given to the 96-wells plate in order to verify the absence of general locomotor defects. The system provides tapping stimuli that range from 1 to 8 arbitrary units (AU). An intensity of 3 AU was chosen for the experiments (1 and 2 were too weak to induce any response in normal larvae and higher intensities could over stimulate larvae and mask possible subtle effects).

The final readout to assess the capacity to respond to sound was the percentage of stimuli responding 6 day-old larvae.

Four different sound stimuli (300, 325, 350, and 375 Hz) that showed the highest percentage of responding larvae were tested.

Then genomic DNA was extracted from single larvae to identify their specific genotype.



To further verify the normal locomotion activity of *slc9a3r1*^{R180Q/R180Q} larvae, we analyze their movement during the entire trial and compare it with the wild-type.

Hair Cell Characterization

Hair Cell Number

In order to test the number of functional hair cells in the saccular macula (the sensory patch responsible for hearing) *tg(brn3c:mGFP;R180Q-slc9a3r1*^{+/R180Q}) adult carrier were incrossed and 6 dpf progeny were intra-ear injected with FM1-43FX, a vital dye that penetrates in only mature and fully functional hair cells through calcium channels. In particular, 6 dpf larvae were anesthetized using Tricaine (20 μ M) and 1 nl of FM1-43FX (stock solution 300 μ M) was microinjected into the inner ear lumen (intra-ear). Once injected, larvae were fixed with 4% PFA 2 h at room temperature and then incubated with PBS-Triton 2% O/N, in order to dissolve the otoliths. The larvae were then embedded on their lateral sides in 1% low melting point agarose and the hair cells expressing membrane GFP (mGFP) and that internalized the FM1-43X were imaged using a SP8 Leica confocal microscope and Z-stacks spanning the entire saccular macula were taken (one z-plane imaged every 1 μ m). Raw data was analyzed with FIJI software (Schindelin et al., 2012).

Confocal imaging of the saccular macula of *slc9a3r1*^{+/+}, *slc9a3r1*^{+/R180Q} and *slc9a3r1*^{R180Q/R180Q} was done and images were analyzed. An in-house FIJI-based macro allows us to 3D reconstruct the entire saccular macula, measure the GFP (*brn3c*) and DsRed (FM1-43FX) signals and estimate the total number and the number of mature hair cells.

Hair Cells Morphology and Polarity

To test sensory hair cells morphology and orientation, 6 dpf larvae from *tg(brn3c:mGFP;R180Q-slc9a3r1*^{+/R180Q}) adult pairwise mating were cryosectioned and an immunohistochemistry against the acetylated-tubulin, which labels the kinocilium, was performed. This staining, together with the membrane GFP signal help to detect alteration in hair cell morphology or in the orientation of the hair cell bundles.

Six dpf larvae were fix with 4% PFA 2 h at room temperature and then incubated with PBS-Triton 2% O/N in order to dissolve the otoliths. After several washes with 0.1%-Tween-20 PBS, larvae heads were incubated 1 h in 15% sucrose (in PBS) then in 15% sucrose/7.5% gelatin and placed in cryomold in the desired orientation (transversally for hair cell morphology, laterally for hair cell bundle orientation), while tails were used for genotyping. Blocks were frozen in 2-Methylbutane for tissue preservation and cryosectioned at 30 μ m on a Leica CM 1510-1 cryostat. Sections were collected on Superfrost slides. Then the inner ear containing slides were blocked with blocking solution [0.1%Tween-20 in PBS (PBT), 2% bovine serum albumin (BSA) and 10% goat serum] for 1.5 h at RT. Mouse anti-acetylated-tubulin (Sigma; 1:1000) and rabbit anti-GFP (Torrey Pines; 1:400) were incubated overnight at 4°C in blocking solution. After washing with PBT for the whole day, anti-rabbit Alexa488 and anti-mouse Alexa648 (Invitrogen; 1:400) were incubated overnight at 4°C in blocking solution. Slides were then mounted with mowiol and imaged using a SP8 Leica confocal microscope and Z-stacks spanning

the entire saccular macula were taken (one z-plane imaged every 1 μ m). Raw data were analyzed with FIJI software in order to generate 3D reconstruction of the saccular macula or analyze the hair cell bundle orientation. To quantify and generate the graphs showing the ciliary orientation a Python-based in-house program was used.

For every single cell, we determine the orientation based on the position of the kinocilia and plot every cell on an Angle's plot (Inoue et al., 2013) based on their polarization and finally generate a Polar plot to represent the percentage of the cells oriented in the dorsal, ventral, anterior, and posterior direction.

Stato-Acoustic Ganglion Characterization

The number of neurons in the posterior stato-acoustic ganglion (SAG) (which innervate the saccular macula) was tested.

Adults *Tg(isl3:mGFP;R180Q-slc9a3r1*^{+/R180Q}) were incrossed to obtain progeny. Six dpf larvae were fixed with 4% PFA 2 h at room temperature and then incubated with PBS-Triton 2% O/N in order to dissolve the otoliths. The larvae were then embedded on their sides in 1% low melting point agarose and the sensory neurons expressing GFP were imaged using a SP8 Leica confocal microscope and Z-stacks spanning the entire posterior SAG were taken (one z-plane imaged every 1 μ m). Raw data were analyzed with FIJI software (Schindelin et al., 2012).

Using a FIJI-based in-house macro the entire posterior SAG was 3D reconstructed the number of the sensory neurons was calculated.

Otolith Analysis

To test the size of the otolith, a crystalline structure contacted by the hair cell ciliary array and essential to drive the movement of the hair cell bundle, we incrossed *tg(brn3c:mGFP;R180Q-slc9a3r1*^{+/R180Q}) adult carriers and 6 dpf progeny was anesthetized using Tricaine (20 μ M), placed on their side, and imaged with Leica M165 Stereoscope before genotyping.

The images were analyzed with FIJI software and the area of the posterior otolith was measured.

Statistical Analysis

Behavioral and otolith differences between *slc9a3r1*^{+/+}, *slc9a3r1*^{+/R180Q}, and *slc9a3r1*^{R180Q/R180Q} were evaluated with an unpaired *T*-test with Welch's correction.

The significance threshold was set at 0.05.

RESULTS

A total of 464 ARHL Italian patients were screened with a TRS panel of 46 ARHL candidate genes (Morgan et al., 2018). A mean of 33.5 megabases of raw sequence data were available for each subject. The coverage, on 95% of the targeted region, was at least 20-folds, with a 270-fold mean-depth total coverage. On average 333 single nucleotide variants (SNVs) and small insertions/deletions (INDELs) were called for each patient. After applying the filtering procedure described in the Section "Materials and Methods," a missense variant at

the heterozygous state, c.539G > A, p.(R180Q) in *SLC9A3R1* (NM_004252.4) gene was detected in two unrelated ARHL male patients (582130, aged 79, and 593486, aged 63). It affects the PDZ2 domain of NHERF1 protein (**Figure 2A**) and all *in silico* predictor tools classified it as damaging (**Supplementary Figure S1B**). Both subjects come from the same isolated village located in North Eastern Italy (Friuli Venezia Giulia-FVG cohort) and they were recruited thanks to a huge project on isolated community in Italy (INGI consortium) based on a volunteer study. Both patients show severe to profound high-frequency hearing loss resembling a sensory ARHL phenotype and the age of onset was 48 y.o. (582130) and 51 y.o. (593486) (**Supplementary Figures S1A,B**) without vestibular signs or symptoms. Additional clinical data include: (1) patient 582130: normal BMI, high pressure treated with drugs. During his life he was affected by Tuberculosis (TBC), lung cancer and dislipidemia. Habits: no smoking and limited alcohol consumption; (2) patient 593486: overweight, high pressure treated with drugs, asthma treated with drugs, sarcoidosis. Habits: no smoking and no alcohol consumption. No other clinical features were detected in the patients nor exposure to noise.

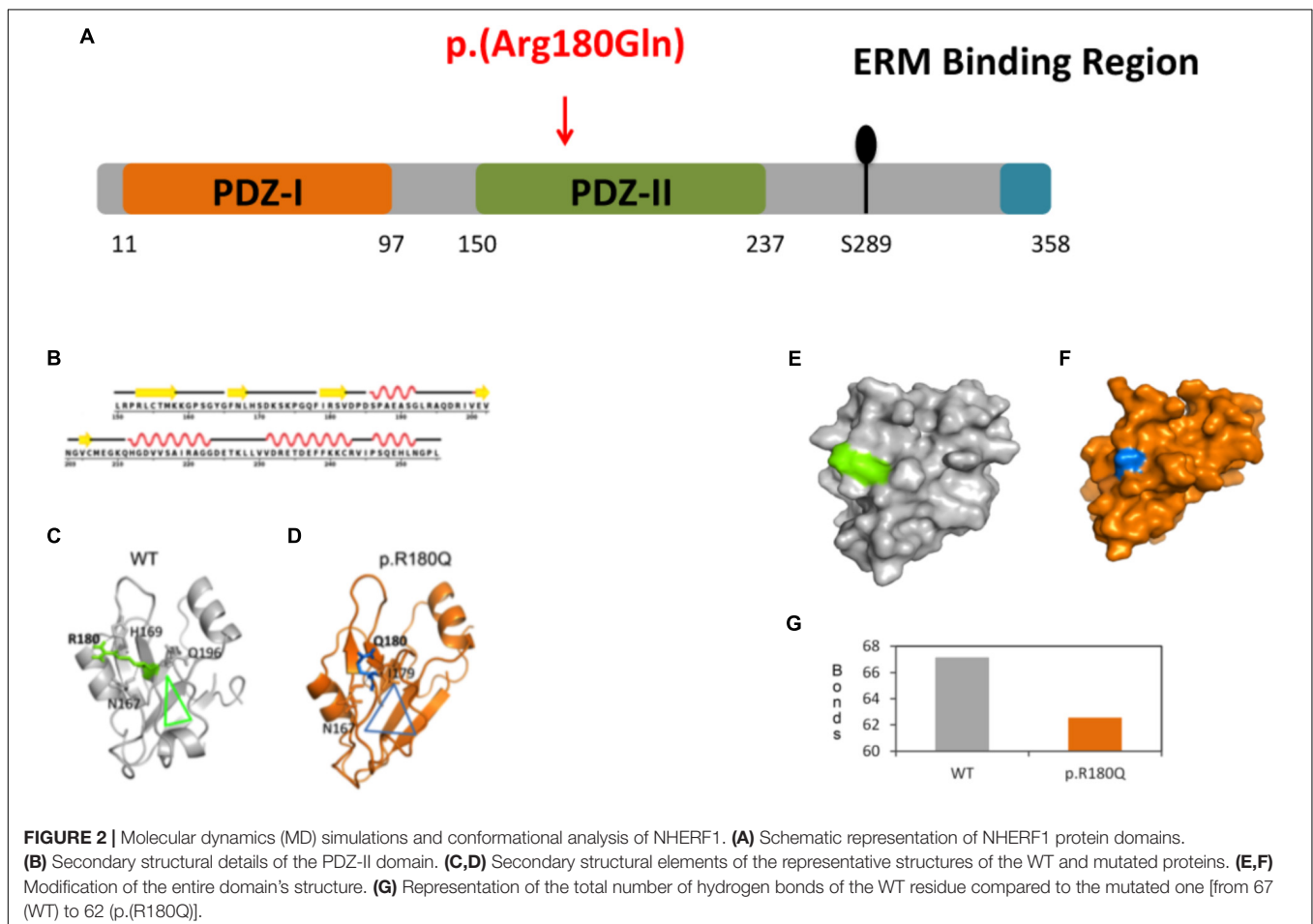
The c.539G > A, p.(R180Q) variant has a frequency of 0.002 in our cohort of 464 ARHL patients and it is described in

gnomAD browser with a minor allele frequency (MAF) of 2.886e-5 (rs146832150) (date of access 21/08/2018).

We confirmed the absence of the variant in all the available relatives of the patients (2 healthy daughters of 582130) as well as any other ARHL patient and control from the same small community (15 ARHL cases and 37 healthy controls). Moreover, 350 additional controls coming from the other five villages of the FVG cohort have been tested, confirming the absence of the *SLC9A3R1* variant.

Protein Simulations and Conformational Analysis

Protein modeling and molecular dynamics (MD) simulations were used to test the impact of the mutation at molecular level. The secondary structural details of the PDZ2 domain (**Figure 2B**) show that residue R180 is located in the surface of a well-structured β sheet (**Figure 2C**) underlying its importance for the protein structure and binding activity. MD simulations revealed that secondary structural elements of both the wild-type (WT) and the mutant are stable during simulations. However, a major change was observed in the side chain of Q180 compared to R180, which further affects their hydrogen bonding partners (R180: H169, Q196 and N167 and Q180:



I179 and N167) in the nearby region (Figures 2C,D). The outcome is a series of local network changes leading to rearrangements of neighboring loops of the mutational spot (Figures 2C,D, see the triangles for the impact) and following modifications on the surface of the domain itself whose entire conformation is significantly modified (Figures 2E,F). This finding correlates with the change in the total number of hydrogen bonds from 67 (WT) to 62 (p.(R180Q)) (Figure 2G). Overall, the modeling suggests a significant conformational change that can alter the normal function of the domain and the protein itself.

In vitro Experiments: qRT-PCR, Western Blot, Immunostaining

In order to understand the effect of the mutated allele on RNA and protein stability, HEK 293 cells have been transfected with expression vectors containing either the WT or the mutant cDNA. qRT-PCR and WB analysis 48 h after the transient transfection did not reveal any difference in the

levels of both mRNA and protein between WT and mutant (Supplementary Figure S2).

Moreover, immunostaining assay on transfected HeLa cells revealed no differences in the cellular localization of the mutated protein compared to the WT (Supplementary Figure S2).

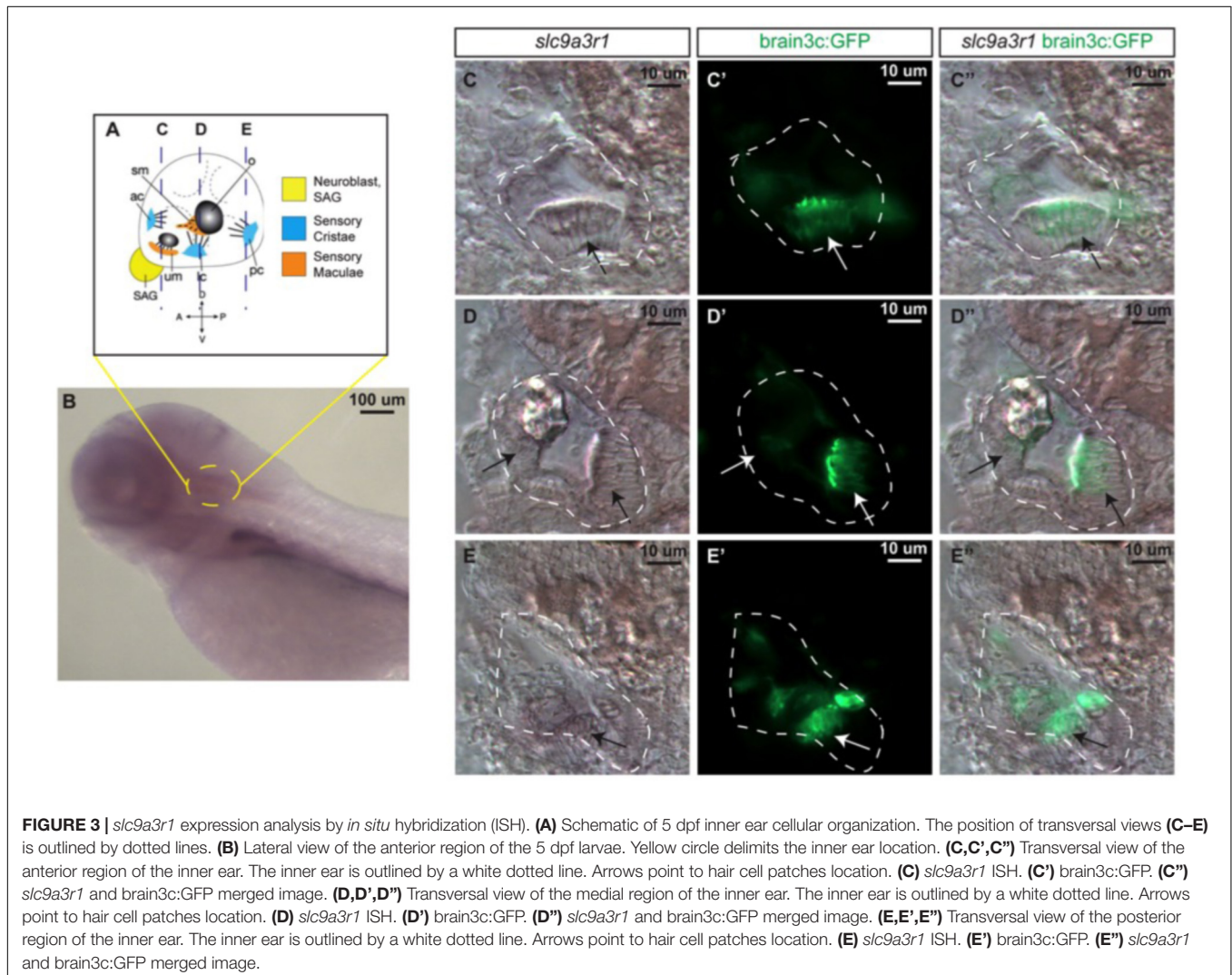
In vivo Experiments: Zebrafish KI Model

In situ Hybridization

Prior to the generation of the KI zebrafish model, *slc9a3r1* expression was tested in 5 dpf zebrafish larvae, by whole mount *in situ* hybridization (Figure 3). *slc9a3r1* mRNA was detected broadly across the larvae, but its expression was enriched in the hematopoietic lineage, liver primordium (Figure 3B) and in the inner ear (Figures 3C–E).

KI Hearing Phenotype

CRISPR/Cas9 technology has been used to generate the KI model (R180Q- *slc9a3r1*), in which has been introduced a precise nucleotide modification mimicking the human mutation (Figures 1A–D). Overall 135 six days post



fertilization (dpf) larvae obtained by pairwise mating of adult tg (brn3c:mGFP;R180Q-*slc9a3r1*^{+/R180Q}), were tested to determine their capacity to respond to sound stimuli at different frequencies (300, 325, 350, and 375). In general, for all stimuli, the response of *slc9a3r1*^{R180Q/R180Q} KI larvae ($n = 31$) was lower as compared to that of heterozygous *slc9a3r1*^{+/R180Q} ($n = 47$) and WT animals ($n = 57$). Interestingly, a statistical significant difference in sound perception at 325 Hz was definitely evident between the WT and the homozygous mutants ($P = 0.0132$), and a worsening hearing perception was also noticed in the heterozygous animals (Figure 4A) while no variation in the ability to respond to mechanical stimuli was observed in comparison to the wild type siblings.

Behavioural Test

Locomotion was studied after light/darkness stimuli to verify that response differences to sound stimuli among WT and mutant siblings were not promoted by general defects in locomotive behavior.

Wild type Zebrafish larvae show a relatively high swimming activity in darkness when compared to the activity in light (de Esch et al., 2012). No differences have been observed for the *slc9a3r1*^{+/R180Q} and *slc9a3r1*^{R180Q/R180Q} models (see Figure 4B). The graph shows normal behavior and response to visual stimuli of homozygous *slc9a3r1* mutants compared to heterozygous and WT, demonstrating that the impairment observed in sound response is specific to hearing function and not to defects in central nervous system.

To further investigate the role of this gene/mutation, several functional tests in the KI model have been performed and, among them, a significant reduced otolith size has been detected.

Measurement of the Otolith Size

Despite fish vestibular-acoustic system presents several functional elements that are very similar to the human ones (e.g., hair cells, supporting cells, etc.), they do not have a counterpart of the cochlea. Moreover, in fish inner ear the otoliths are essential structures for balance and hearing. Otoliths are crystalline structure contacted by the hair cell ciliary array and are essential to drive the movement of the hair cells bundle and start the electrical signaling in case of sound (Riley and Phillips, 2003). The alteration of the size, shape or the absence of this structure could have striking effects on the hearing ability (Inoue et al., 2013). For this reason, we looked for any difference in the otoliths of WT compared to that of the mutated animals.

Results shows a 5,8% reduction of the otolith size in *slc9a3r1*^{+/R180Q} compared to WT ($P = 0.0014$) and an even more increased 10.2% reduction in the size of the otolith of *slc9a3r1*^{R180Q/R180Q} compare to WT ($P < 0.0001$) (Figures 5A,B). These results thus explain the hearing impairment observed in the *slc9a3r1*^{R180Q/R180Q} KI larvae.

Other functional tests demonstrated no alterations in the hair cells number and morphology as well as in the count of neurons, as reported below.

Hair Cells Number and Morphology

The reduced capacity of *slc9a3r1* mutant larvae to respond to acoustic stimuli could have several explanations such as: (a) a reduced number or a defective maturation of the sensory hair cells leading to not functional hair cells in the saccular macula, (b) an abnormal morphology or an altered orientation of the sensory hair cells. As regards to (a), 9 WT animals compared to 10 heterozygous mutants and 6 homozygous mutants were tested. No differences in the total number (35.79 ± 6.11 cells for WT, 38.32 ± 5.11 cells for R180Q/R180Q, 43.18 ± 6.53 for \pm total cells) or fully mature hair cells (31.16 ± 5.22 cells for WT, 32.26 ± 3.25 cells for R180Q/R180Q, 35.56 ± 3.27 for $+/R180Q$ total cells) among *slc9a3r1*^{R180Q/R180Q}, *slc9a3r1*^{+/R180Q} mutants and their Wt siblings were observed (Figures 6A,B). As regards to (b) 6 dpf larvae (*slc9a3r1*^{+/+} $n = 9$, *slc9a3r1*^{+/R180Q} $n = 10$, *slc9a3r1*^{R180Q/R180Q} $n = 6$) were cryosectioned and an immunohistochemistry against the acetylated-tubulin, that labels the kinocilium, was performed. Data analysis revealed no differences in hair cell body morphology or size, hair cell bundle organization or kinocilia/stereocilia length suggesting that the KI do not alter the hair cell morphology (Figure 6C) but induce a reduced response to sound stimuli due to other events. Moreover, no statistically significant differences were observed in the orientation of the hair cells within the sensory patch (see Figure 6D), suggesting that the hearing impairment in KI is not due to hair cell altered polarization. All together these findings indicate the KI hearing impairment is not caused by alterations in the hair cell population.

Count of Neurons

Since the hair cells of *slc9a3r1* mutants do not show any defect, the number of neurons in the posterior stato-acoustic ganglion (SAG), which innervate the saccular macula and could be related to a hearing loss, was checked.

In 6 dpf larvae (3 *slc9a3r1*^{+/+}, 6 *slc9a3r1*^{R180Q/R180Q}, 6 *slc9a3r1*^{+/R180Q} animals used) few differences (although not statistically significant) in the volume of the total SAG ($42.439 \mu\text{m}^3$ for *slc9a3r1*^{+/+}, $35.454 \mu\text{m}^3$ for *slc9a3r1*^{R180Q/R180Q}, $40.237 \mu\text{m}^3$ for *slc9a3r1*^{+/R180Q}), and no alterations in the volume of the posterior SAG ($13.676 \mu\text{m}^3$ for *slc9a3r1*^{+/+}, $11.404 \mu\text{m}^3$ for *slc9a3r1*^{R180Q/R180Q}, $12.709 \mu\text{m}^3$ for *slc9a3r1*^{+/R180Q}) were detected (Figures 7A,B), thus suggesting that the hearing impairment is not due to a reduction of neurons in this structure.

DISCUSSION

Despite significant research efforts carried out, the genetics risk factors involved in ARHL are still mainly unknown and only few ARHL susceptibility genes have been detected so far.

Zebrafish mutants can show morphological and functional defects similar to those of other species (Ernest et al., 2000) and their vestibular-acoustic system is very similar to that of human beings, thus representing a cost and time-saving alternative to mouse model allowing a rapid assessment of inner ear

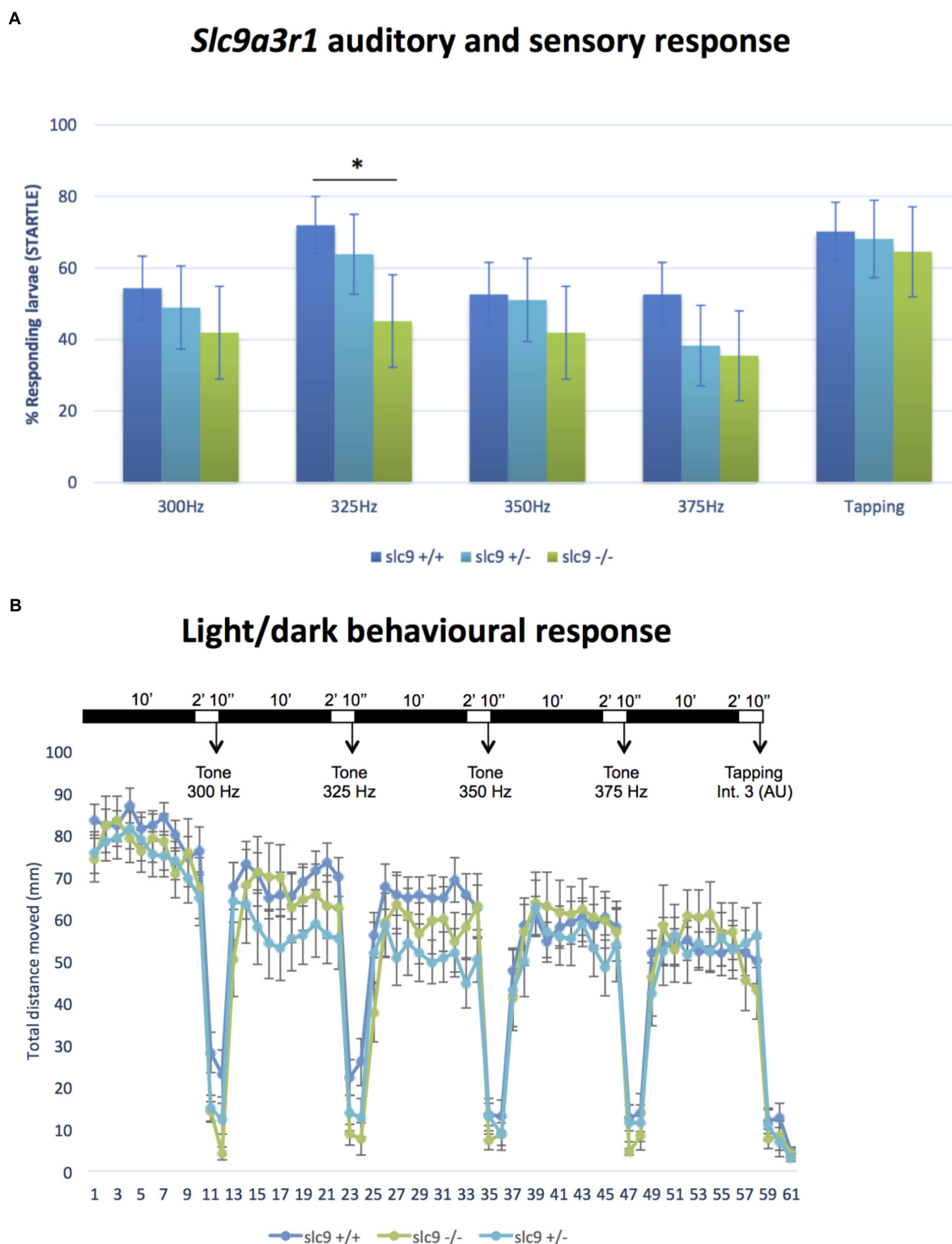


FIGURE 4 | Auditory/sensory response and light/dark behavioral response in Zebrafish larvae. **(A)** Response of *slc9a3r1* KI animals compared to the wild-type ones to all the different sound stimuli tested (number of tested animals: *slc9a3r1*^{+/+} *N* = 57, *slc9a3r1*^{+/-} *N* = 47, *slc9a3r1*^{R180Q/R180Q} *N* = 31). Statistically significant difference indicated with "*" (**p*-value < 0.05). **(B)** Behavior and response activity to visual stimuli of *slc9a3r1*^{R180Q/R180Q} compared to *slc9a3r1*^{+/+} and *slc9a3r1*^{+/-}. Tapping intensity is indicated as arbitrary unit (AU).

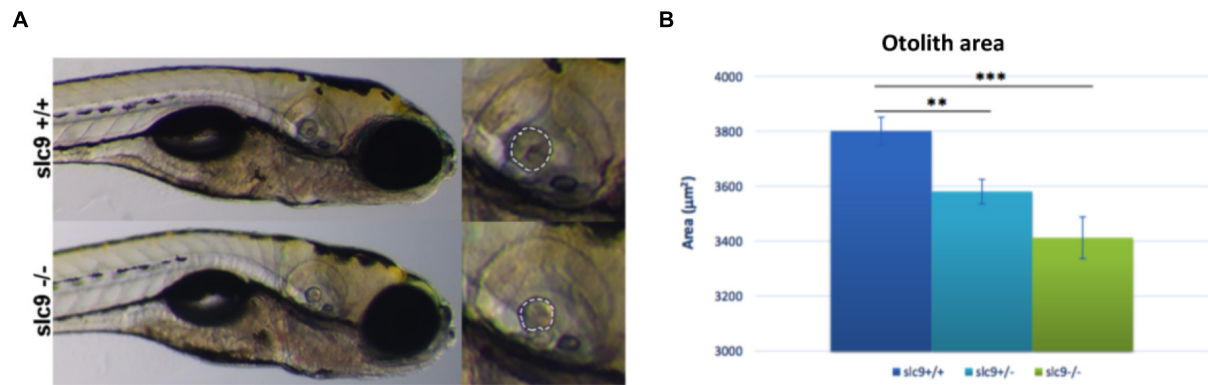


FIGURE 5 | Measurement of the otolith size. **(A)** Measurement of the posterior otolith in *slc9a3r1*^{+/+} and *slc9a3r1*^{R180Q/R180Q} larvae. **(B)** Measurement of posterior otolith area of *slc9a3r1* larvae (number of tested animals: *slc9a3r1*^{+/+} *N* = 57, *slc9a3r1*^{+/-} *N* = 70, *slc9a3r1*^{R180Q/R180Q} *N* = 32). ***p*-value < 0.005 and ****p*-value < 0.0005.

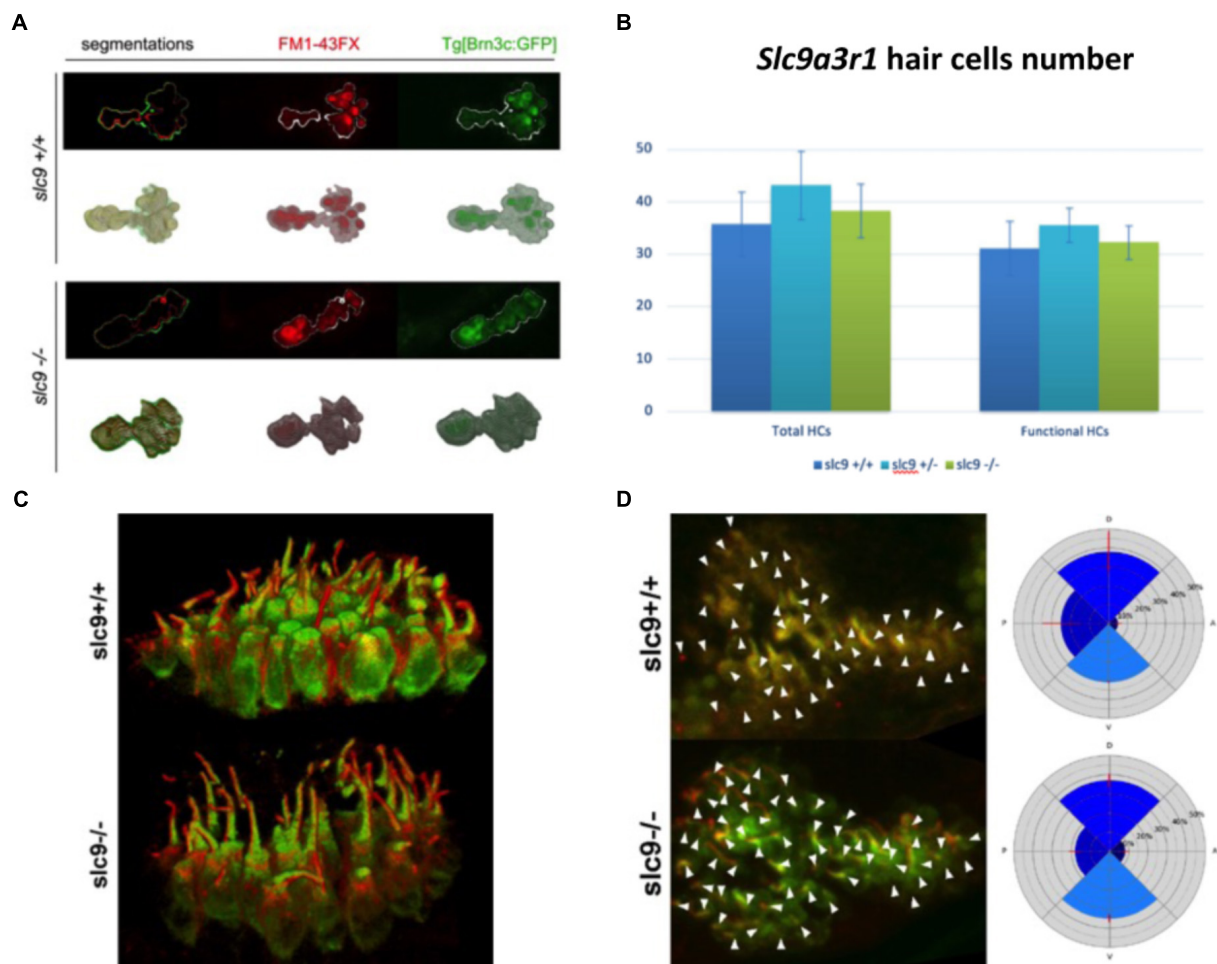
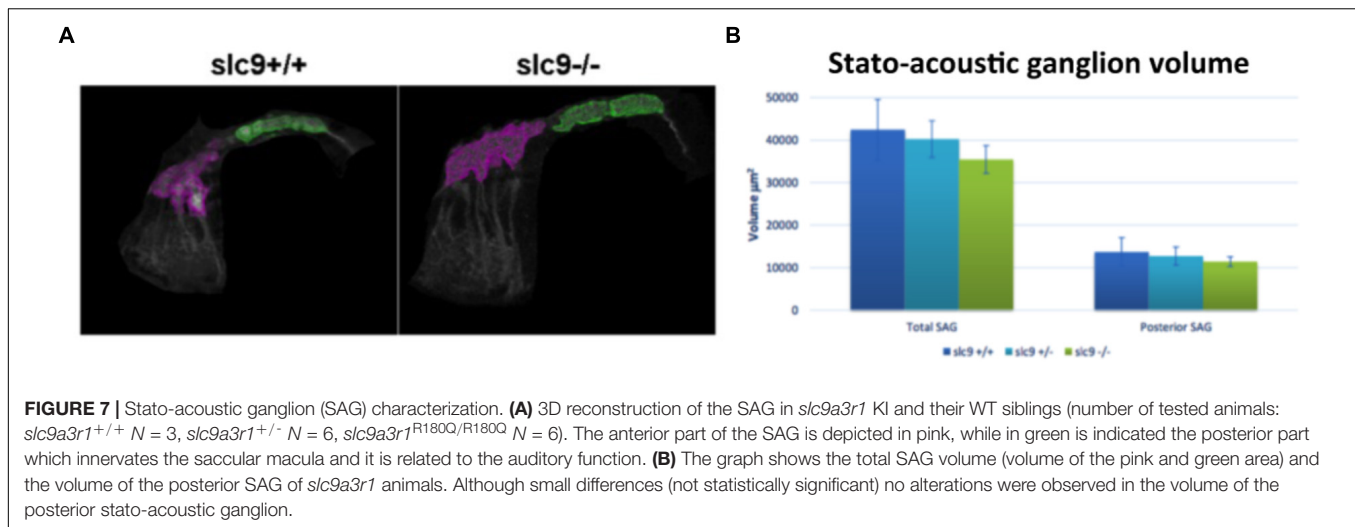


FIGURE 6 | Hair cells number, morphology and orientation in Zebrafish larvae. **(A)** 3D reconstruction of the saccular macula. Hair cells are marked with GFP. FM1-43FX dye labels only mature and fully functional hair cells. **(B)** Saccular hair cells count in *slc9a3r1* larvae (number of tested animals: *slc9a3r1*^{+/+} *N* = 9, *slc9a3r1*^{+/-} *N* = 10, *slc9a3r1*^{R180Q/R180Q} *N* = 6). **(C)** 3D reconstruction of the saccular hair cell layer. *brn3c:mGFP* signal labels the membrane of the hair cells while red signal labels the kinocilia. **(D)** Hair cell polarization within the saccular macula of *slc9a3r1* KI and their wt siblings.



function (Baxendale and Whitfield, 2016; Leventea et al., 2016). Furthermore, a number of genes required for hair-cell function in the zebrafish have been already associated with auditory defects in mice and humans, thus revealing their conservation across species (Coimbra et al., 2002).

Using a combination of genetics/genomics studies followed by functional studies we deeply investigated the role of *SLC9A3R1* in ARHL. In particular, the identification of an ultra-rare human mutation [c.539G > A, p.(R180Q)] in two ARHL patients from an isolated community of the Italian Alps led to a series of functional studies including the generation of the first KI zebrafish model for hearing phenotype developed so far demonstrating the pathogenic effect of this variant and thus a possible relevant role of *SLC9A3R1* in ARHL. *SLC9A3R1* encodes the Na⁺/H⁺ exchange regulatory factor 1 (NHERF1) protein (also called Ezrin-radixin-moesin-binding protein of 50 kDa, Ebp50), which belongs to NHERF family of scaffolding proteins. These proteins play key roles in mediating a number of transports and signaling phenomena, interacting with different molecular partners. Interestingly, several protein–protein interactions involving NHERFs take place in the cochlea, and accordingly the *Nherf1* KO mouse shows hearing defects, due to hair cells anomalies (Kamiya et al., 2014). NHERF1 interacts with many different molecular partners through its PDZ domains. The missense variant here identified affects the second PDZ domain of the protein (PDZ2) and involves the R180 residue, which is essential for protein–protein interaction, as demonstrated by Mamonova et al. (2012). Accordingly, our protein modeling confirmed this data, suggesting that the p.(R180Q) mutation induces changes in the network of interactions by altering the total number of hydrogen bonds and thus compromising the binding activity of the protein itself. Moreover, the KI zebrafish model further support the pathogenic effect of the p.(R180Q) variant, showing that it causes a reduction in size of the saccular otolith, a structure corresponding to human otoconia whose abnormalities might cause vertigo, dizziness and isolated hearing loss. Zebrafish otoliths are biomineralised structures required for balance and hearing. While the utricular

otolith is mainly involved in balancing, the saccular one is implicated in the sense of hearing (Stooke-Vaughan et al., 2015). Interestingly, genetic screenings analyzing several zebrafish mutants with hearing phenotype showed abnormalities in the otolith volume and formation, and revealed that otolith defects are not necessarily associated with other morphological changes of the inner ear (Schibler and Malicki, 2007) as it happens to our *slc9a3r1* KI model. This finding clearly justifies the absence of any balance problem in our KI model and this is in agreement with the ARHL phenotype of the two ARHL patients here described (i.e., high-frequencies hearing loss without any vestibular symptoms). Furthermore, data obtained for *SLC9A3R1* gene are also in agreement with those reported for other genes involved in the otoconia development such as Claudin and Tectorin, in which vestibular signs are frequently absent suggesting a functional compensation by other proteins (Lundberg et al., 2015).

Molecular mechanism underlying KI morphological changes still need to be understood, however, it has been shown that different Na⁺/K⁺-ATPase genes are essential for otolith formation in zebrafish (Blasiolo et al., 2006). In this light, a relevant role might be played by the Na⁺/K⁺-ATPase α1 subunit that binds the PDZ2 domain of NHERF1 in opossum kidney cells (Mahon et al., 2003). Since the strong homology of this subunit between zebrafish and opossum sequences (~71%) it is reasonable to assume that this binding could also happen in zebrafish. Taking into account that our KI model carries a mutation in the PDZ2 domain in a residue critical for protein–protein interaction, it is highly probable that this function could be severely compromised (Voltz et al., 2001; Blasiolo et al., 2006) thus leading to an altered otolith formation and a subsequent hearing loss. Furthermore, this mechanism of action could also explain the phenotype's difference between the KO mouse (that shows anomalies in the shape of the hair cells bundles) and our KI zebrafish model, which specifically reproduces the variant identified in our patients.

Despite carrying the same variant, the KI zebrafish model displays early-onset hearing loss, while the two patients here

described are affected by ARHL. This difference might be explained by several hypotheses.

First of all, even though animal models are extremely useful for recapitulating many human diseases, it is important to keep in mind that in some case the phenotypes' spectra may be different between the two species. Having said that, the *slc9a3r1*^{+/R180Q} model does not show a statistically significant reduction of the hearing thresholds, despite presenting an altered otolith's volume compared to the WT. In this light, considering that the human patients carry the *SLC9A3R1* variant at the heterozygous state, it is reasonable to think that also in this case, despite the presence of an anatomical defect, the hearing phenotype does not occur at an early age.

CONCLUSION

Our results confirmed the role of *SLC9A3R1* in the hearing system function and development, strengthening the previously reported data of the mouse KO model (Kamiya et al., 2014). Moreover, since the KI zebrafish model generated in the present study carries the specific variant identified in two ARHL human patients, this work provides evidence that *SLC9A3R1* gene might play a role in human ARHL. Overall these findings suggest that *SLC9A3R1* could act as a likely Mendelian gene, with a large effect size, in the etiopathogenesis of the late-onset hearing loss form detected in our patients. In this light, larger patients cohorts should be screened at molecular level for the presence of *SLC9A3R1* gene variants/mutations to understand the overall contribution of this gene to the disease itself. Moreover, the identification of other ARHL patients carrying *SLC9A3R1* variants would be also essential for the development of new plans for disease's prevention and treatment.

DATA AVAILABILITY

Sequencing data are available at the European Genome-phenome Archive (EGA) at the following link <https://ega-archive.org/studies/EGAS00001003072>.

AUTHOR CONTRIBUTIONS

GG: study design assessment, data production and analysis, and writing the manuscript. AM: data production and analysis and

writing the manuscript. NK: protein modeling and molecular dynamics simulation. MC: raw data analysis and quality control. BM: raw data analysis and quality control. SB: *in vitro* studies. MLB: data production. MDS: support of functional studies and conceiving the experiments. PG: conceiving the experiments and writing the manuscript.

FUNDING

This research was supported by RBSI14AG8P-SIR2014 to GG (<http://sir.miur.it/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge ZeClinics company for technical contribution in Zebrafish models.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00142/full#supplementary-material>

FIGURE S1 | Clinical features and DNA sequence chromatograms of the two ARHL patients. **(A)** Audiograms of patients 582130 and 5934856; the downward slope indicates that high frequencies are severely affected. **(B)** The figure displays DNA sequence chromatograms showing the nucleotide variant identified in the two ARHL patients and all the details of the mutation. rs ID, Reference SNP ID from dbSNP database. Frequency, frequency of gnomAD Exome all is reported. Predictor tools, GERP++ (higher number is more conserved, >0 is generally conserved), PhyloP (Pathogenicity score: conserved > 0.95, not conserved < 0.95), Polyphen-2 (Pathogenicity score: probably damaging: D, possibly damaging: P, benign: B), SIFT (Pathogenicity score: D: disease causing, N: polymorphism, P: polymorphism automatic), MutationTaster (Pathogenicity score: D: disease causing, N: polymorphism, P: polymorphism automatic), CADD Phred (Pathogenicity score: >10 predicted to be deleterious).

FIGURE S2 | *In vitro* experiments: mRNAs quantification and proteins analyses. After cells transfection with expression vectors containing either the WT or the mutant *SLC9A3R1* cDNA, mRNA and protein analyses were performed. **(A)** qRT-PCR on mRNA showed no difference in the expression levels of the mutant (*) compared to the WT. **(B)** Western blot analysis demonstrated that WT and mutated protein are equally expressed. **(C)** Immunolocalization showed that both the WT and the mutated proteins have a cytoplasmic expression.

REFERENCES

- Abbas, L., and Whitfield, T. T. (2009). Nkcc1 (Slc12a2) is required for the regulation of endolymph volume in the otic vesicle and swim bladder volume in the zebrafish larva. *Development* 136, 2837–2848. doi: 10.1242/dev.034215
- Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* 76, 7.20.1–7.20.41. doi: 10.1002/0471142905.hg0720s76
- Baxendale, S., and Whitfield, T. T. (2016). Methods to study the development, anatomy, and function of the zebrafish inner ear across the life course. *Methods Cell Biol.* 134, 165–209. doi: 10.1016/bs.mcb.2016.02.007
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi: 10.1063/1.448118
- Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., and Hermans, J. (1981). Interaction models for water in relation to protein hydration. *Internol. Forces* 14, 331–342. doi: 10.1007/978-94-015-7658-1_21
- Bhandiwad, A. A., Zeddies, D. G., Raible, D. W., Rubel, E. W., and Sisneros, J. A. (2013). Auditory sensitivity of larval zebrafish (*Danio rerio*) measured using a behavioral prepulse inhibition assay. *J. Exp. Biol.* 216, 3504–3513. doi: 10.1242/jeb.087635
- Bhattacharya, S., Dai, Z., Li, J., Baxter, S., Callaway, D. J. E., Cowburn, D., et al. (2010). A conformational switch in the scaffolding protein NHERF1 controls

- autoinhibition and complex formation. *J. Biol. Chem.* 285, 9981–9994. doi: 10.1074/jbc.M109.074005
- Blasiolo, B., Canfield, V. A., Vollrath, M. A., Huss, D., Mohideen, M.-A. P. K., Dickman, J. D., et al. (2006). Separate Na,K-ATPase genes are required for otolith formation and semicircular canal development in zebrafish. *Dev. Biol.* 294, 148–160. doi: 10.1016/j.ydbio.2006.02.034
- Bovo, R., Ciorba, A., and Martini, A. (2011). Environmental and genetic factors in age-related hearing impairment. *Aging Clin. Exp. Res.* 23, 3–10. doi: 10.1007/BF03324947
- Bowl, M. R., and Dawson, S. J. (2015). The mouse as a model for age-related hearing loss - a mini-review. *Gerontology* 61, 149–157. doi: 10.1159/000368399
- Chun, S., and Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561. doi: 10.1101/gr.092619.109
- Coimbra, R. S., Weil, D., Brottier, P., Blanchard, S., Levi, M., Hardelin, J.-P., et al. (2002). A subtracted cDNA library from the zebrafish (*Danio rerio*) embryonic inner ear. *Genome Res.* 12, 1007–1011. doi: 10.1101/gr.227502
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- de Esch, C., van der Linde, H., Sliker, R., Willemsen, R., Wolterbeek, A., Woutersen, R., et al. (2012). Locomotor activity assay in zebrafish larvae: influence of age, strain and ethanol. *Neurotoxicol. Teratol.* 34, 425–433. doi: 10.1016/j.ntt.2012.03.002
- Di Donato, V., Auer, T. O., Durore, K., and Del Bene, F. (2013). Characterization of the calcium binding protein family in zebrafish. *PLoS One* 8:e53299. doi: 10.1371/journal.pone.0053299
- Ernest, S., Rauch, G. J., Haffter, P., Geisler, R., Petit, C., and Nicolson, T. (2000). Mariner is defective in myosin VIIA: a zebrafish model for human hereditary deafness. *Hum. Mol. Genet.* 9, 2189–2196. doi: 10.1093/hmg/9.14.2189
- Friedman, R. A., Van Laer, L., Huentelman, M. J., Sheth, S. S., Van Eyken, E., Corneveaux, J. J., et al. (2009). GRM7 variants confer susceptibility to age-related hearing impairment. *Hum. Mol. Genet.* 18, 785–796. doi: 10.1093/hmg/ddn402
- Giroto, G., Pirastu, N., Sorice, R., Biino, G., Campbell, H., d'Adamo, A. P., et al. (2011). Hearing function and thresholds: a genome-wide association study in European isolated populations identifies new loci and pathways. *J. Med. Genet.* 48, 369–374. doi: 10.1136/jmg.2010.088310
- Giroto, G., Vuckovic, D., Bunello, A., Lorente-Cánovas, B., Lewis, M., Gasparini, P., et al. (2014). Expression and replication studies to identify new candidate genes involved in normal hearing function. *PLoS One* 9:e85352. doi: 10.1371/journal.pone.0085352
- Haddon, C., and Lewis, J. (1996). Early ear development in the embryo of the zebrafish, *Danio rerio*. *J. Comp. Neurol.* 365, 113–128. doi: 10.1002/(SICI)1096-9861(19960129)365:1<113::AID-CNE9>3.0.CO;2-6
- Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). 3 LINC: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H
- Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4, 435–447. doi: 10.1021/ct700301q
- Howe, K., Clark, M. D., Torroja, C. F., Torrance, J., Berthelot, C., Muffato, M., et al. (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* 496, 498–503. doi: 10.1038/nature12111
- Huang, Q., and Tang, J. (2010). Age-related hearing loss or presbycusis. *Eur. Arch. Otorhinolaryngol.* 267, 1179–1191. doi: 10.1007/s00405-010-1270-7
- Inoue, M., Tanimoto, M., and Oda, Y. (2013). The role of ear stone size in hair cell acoustic sensory transduction. *Sci. Rep.* 3:2114. doi: 10.1038/srep02114
- Kamiya, K., Michel, V., Giraudet, F., Riederer, B., Foucher, I., Papal, S., et al. (2014). An unusually powerful mode of low-frequency sound interference due to defective hair bundles of the auditory outer hair cells. *Proc. Natl. Acad. Sci. U.S.A.* 111, 9307–9312. doi: 10.1073/pnas.1405322111
- Kidd Iii, A. R., and Bao, J. (2012). Recent advances in the study of age-related hearing loss: a mini-review. *Gerontology* 58, 490–496. doi: 10.1159/000338588
- Kingston, R. E., Chen, C. A., Rose, J. K., Kingston, R. E., Chen, C. A., and Rose, J. K. (2003). Calcium phosphate transfection. *Curr. Protoc. Mol. Biol.* 63, 9.1.1–9.1.11. doi: 10.1002/0471142727.mb0901s63
- Krishnamoorthy, N., Yacoub, M. H., and Yaliraki, S. N. (2011). A computational modeling approach for enhancing self-assembly and biofunctionalisation of collagen biomimetic peptides. *Biomaterials* 32, 7275–7285. doi: 10.1016/j.biomaterials.2011.06.074
- Leventea, E., Hazime, K., Zhao, C., and Malicki, J. (2016). Analysis of cilia structure and function in zebrafish. *Methods Cell Biol.* 133, 179–227. doi: 10.1016/bs.mcb.2016.04.016
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lundberg, Y. W., Xu, Y., Thiessen, K. D., and Kramer, K. L. (2015). Mechanisms of otoconia and otolith development. *Dev. Dyn.* 244, 239–253. doi: 10.1002/dvdy.24195
- Mahon, M. J., Cole, J. A., Lederer, E. D., and Segre, G. V. (2003). Na⁺ /H⁺ + exchanger-regulatory factor 1 mediates inhibition of phosphate transport by parathyroid hormone and second messengers by acting at multiple sites in opossum kidney cells. *Mol. Endocrinol.* 17, 2355–2364. doi: 10.1210/me.2003-0043
- Mamonova, T., Kurnikova, M., and Friedman, P. A. (2012). Structural basis for NHERF1 PDZ domain binding. *Biochemistry* 51, 3110–3120. doi: 10.1021/bi201213w
- Miyamoto, S., and Kollman, P. A. (1992). Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. doi: 10.1002/jcc.540130805
- Morgan, A., Vuckovic, D., Krishnamoorthy, N., Rubinato, E., Ambrosetti, U., Castorina, P., et al. (2018). Next-generation sequencing identified SPATC1L as a possible candidate gene for both early-onset and age-related hearing loss. *Eur. J. Hum. Genet.* 27, 70–79. doi: 10.1038/s41431-018-0229-9
- National Health and Nutrition Examination Survey [NHANES] (2015). National Health and Nutrition Examination Survey. Available at: <http://www.cdc.gov/nchs/nhanes.htm> [accessed September 6, 2015].
- Ng, P. C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509
- Nicolson, T. (2005). The genetics of hearing and balance in zebrafish. *Annu. Rev. Genet.* 39, 9–22. doi: 10.1146/annurev.genet.39.073003.105049
- Ohgami, N., Iida, M., Yajima, I., Tamura, H., Ohgami, K., and Kato, M. (2013). Hearing impairments caused by genetic and environmental factors. *Environ. Health Prev. Med.* 18, 10–15. doi: 10.1007/s12199-012-0300-z
- Pittman, A. J., Law, M.-Y., and Chien, C.-B. (2008). Pathfinding in a large vertebrate axon tract: isotopic interactions guide retinotectal axons at multiple choice points. *Development* 135, 2865–2871. doi: 10.1242/dev.025049
- Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* 20, 110–121. doi: 10.1101/gr.097857.109
- Riley, B. B., and Phillips, B. T. (2003). Ringing in the new ear: resolution of cell interactions in otic development. *Dev. Biol.* 261, 289–312. doi: 10.1016/S0012-1606(03)00245-8
- Schibler, A., and Malicki, J. (2007). A screen for genetic defects of the zebrafish ear. *Mech. Dev.* 124, 592–604. doi: 10.1016/j.mod.2007.04.005
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019
- Schwarz, J. M., Rödelberger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. doi: 10.1038/nmeth0810-575
- Stooke-Vaughan, G. A., Obholzer, N. D., Baxendale, S., Megason, S. G., and Whitfield, T. T. (2015). Otolith tethering in the zebrafish otic vesicle requires Otogelin and α -Tectorin. *Development* 142, 1137–1145. doi: 10.1242/dev.116632
- Takayama, K., Igai, K., Hagihara, Y., Hashimoto, R., Hanawa, M., Sakuma, T., et al. (2017). Highly efficient biallelic genome editing of human ES/iPS cells using a CRISPR/Cas9 or TALEN system. *Nucleic Acids Res.* 45, 5198–5207. doi: 10.1093/nar/gkx130
- Thisse, B., Heyer, V., Lux, A., Alunni, V., Degraeve, A., Seiliez, I., et al. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. *Methods Cell Biol.* 77, 505–519. doi: 10.1016/S0091-679X(04)77027-2

- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi: 10.1002/jcc.20291
- van Gunsteren, W. F. (1996). *Biomolecular Simulation: The GROMOS 96 Manual Und User Guide*. Zürich: VDF Hochschulverlag AG an der ETH Zürich.
- Van Laer, L., Huyghe, J. R., Hannula, S., Van Eyken, E., Stephan, D. A., Mäki-Torkko, E., et al. (2010). A genome-wide association study for age-related hearing impairment in the Saami. *Eur. J. Hum. Genet.* 18, 685–693. doi: 10.1038/ejhg.2009.234
- Voltz, J. W., Weinman, E. J., and Shenolikar, S. (2001). Expanding the role of NHERF, a PDZ-domain containing protein adapter, to growth regulation. *Oncogene* 20, 6309–6314. doi: 10.1038/sj.onc.1204774
- Vuckovic, D., Biino, G., Panu, F., Pirastu, M., Gasparini, P., and Giroto, G. (2013). Lifestyle and normal hearing function in Italy and Central Asia: the potential role of coffee. *Hear. Balanc. Commun.* 11, 218–223. doi: 10.3109/21695717.2013.817134
- Vuckovic, D., Biino, G., Panu, F., Pirastu, M., Gasparini, P., and Giroto, G. (2014). Age related hearing loss and level of education: an epidemiological study on a large cohort of isolated populations. *Hear. Balanc. Commun.* 12, 94–98. doi: 10.3109/21695717.2014.911472
- Vuckovic, D., Dawson, S., Scheffer, D. I., Rantanen, T., Morgan, A., Di Stazio, M., et al. (2015). Genome-wide association analysis on normal hearing function identifies PCDH20 and SLC28A3 as candidates for hearing function and loss. *Hum. Mol. Genet.* 24, 5655–5664. doi: 10.1093/hmg/ddv279
- Whitfield, T. T., Riley, B. B., Chiang, M.-Y., and Phillips, B. (2002). Development of the zebrafish inner ear. *Dev. Dyn.* 223, 427–458. doi: 10.1002/dvdy.10073
- Wolber, L. E., Giroto, G., Buniello, A., Vuckovic, D., Pirastu, N., Lorente-Cánovas, B., et al. (2014). Salt-inducible kinase 3, SIK3, is a new gene associated with hearing. *Hum. Mol. Genet.* 23, 6407–6418. doi: 10.1093/hmg/ddu346
- Xiao, T., Roeser, T., Staub, W., and Baier, H. (2005). A GFP-based genetic screen reveals mutations that disrupt the architecture of the zebrafish retinotectal projection. *Development* 132, 2955–2967. doi: 10.1242/dev.01861

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Giroto, Morgan, Krishnamoorthy, Cocca, Brumat, Bassani, La Bianca, Di Stazio and Gasparini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Biochemical, Molecular, and Clinical Characterization of Patients With Primary Carnitine Deficiency via Large-Scale Newborn Screening in Xuzhou Area

Wei Zhou*, Huizhong Li, Ting Huang, Yan Zhang, Chuanxia Wang and Maosheng Gu*

Xuzhou Maternity and Child Health Care Hospital, Xuzhou, China

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Fan Jin,
Zhejiang University, China
Ruili Huang,
National Center for Advancing
Translational Sciences (NCATS),
United States

*Correspondence:

Wei Zhou
wei0743916@163.com
Maosheng Gu
gumaosheng2007@126.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Pediatrics

Received: 07 October 2018

Accepted: 06 February 2019

Published: 26 February 2019

Citation:

Zhou W, Li H, Huang T, Zhang Y,
Wang C and Gu M (2019)
Biochemical, Molecular, and Clinical
Characterization of Patients With
Primary Carnitine Deficiency via
Large-Scale Newborn Screening in
Xuzhou Area. *Front. Pediatr.* 7:50.
doi: 10.3389/fped.2019.00050

Background: Primary carnitine deficiency (PCD) is attributed to a variation in the *SLC22A5* (OCTN2) gene which encodes the key protein of the carnitine cycle, the OCTN2 carnitine transporter. PCD is typically identified in childhood by either hypoketotic hypoglycemia, or skeletal and cardiac myopathy. The aim of this study was to the clinical, biochemical, and molecular characteristics of PCD patients via newborn screening with tandem mass spectrometry (MS/MS).

Methods: MS/MS was performed to screen newborns for inherited metabolic diseases. *SLC22A5* gene mutations were detected in the individual and/or their family member by DNA mass array and next-generation sequencing (NGS).

Results: Among the 236,368 newborns tested, ten exhibited PCD, and six others were diagnosed with low carnitine levels caused by their mothers, who had asymptomatic PCD. The incidence of PCD in the Xuzhou area is $\sim 1:23,637$. The mean initial free carnitine (C_0) concentration of patients was $6.41 \pm 2.01 \mu\text{mol/L}$, and the follow-up screening concentration was $5.80 \pm 1.29 \mu\text{mol/L}$. After treatment, the concentration increased to $22.8 \pm 4.13 \mu\text{mol/L}$.

Conclusion: This study demonstrates the important clinical value of combining MS/MS and NGS for the diagnosis of PCD and provides new insight into the diagnosis of PCD and maternal patients with PCD using C_0 concentration and *SLC22A5* mutations.

Keywords: primary carnitine deficiency (PCD), *SLC22A5* gene, newborn screening, tandem mass spectrometry (MS/MS), maternal PCD

INTRODUCTION

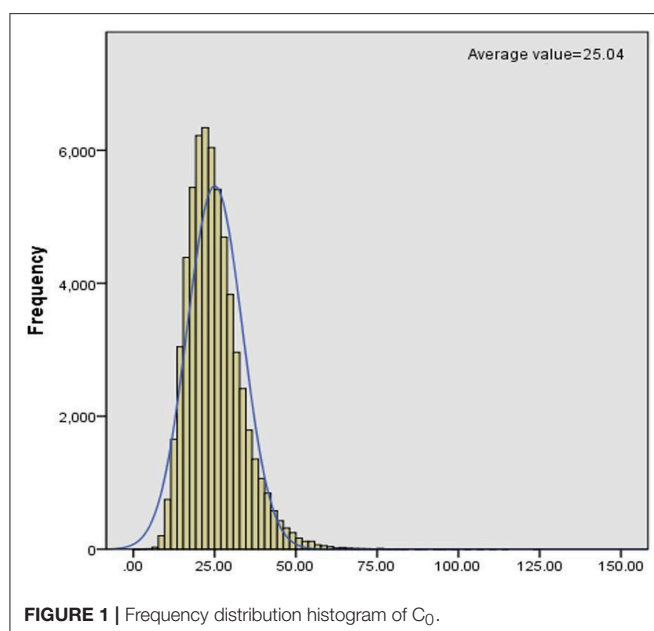
Primary carnitine deficiency (PCD), a of carnitine cycle disorder, represents an autosomal recessive defect that occurs on the *SLC22A5* (OCTN2) gene (1, 2). The organic cation transporter (OCTN) family is essential for transporting organic cation compounds. One member of this family, OCTN2, which is encoded by the *SLC22A5* gene, can transfer carnitine across the cell membrane (3). Without the ability to transport carnitine into the cell, long-chain fatty acids, which are participated in fatty acid β -oxidation, cannot enter the mitochondrial matrix and provide energy for the body (4). As one of the most common fatty acid oxidation and metabolic diseases, PCD was first detected

by measuring plasma free carnitine (C_0) levels in 1988 (5, 6). However, the connection between PCD and mutations located on the *SLC22A5* gene was not demonstrated until 10 years later (1, 2).

The incidence rate of PCD is approximately 1:40,000 in Japan, but the incidence is 1:297 in the Faroe Islands, which is extremely high (7, 8). In Taiwan, the incidence of PCD is 1:18,543, which is similar to that in the Xuzhou area, which has a current prevalence of 1:23,637 (9).

The most typical form of PCD is characterized by progressive infantile-onset cardiomyopathy, with weakness, peripheral neuropathy, and recurrent hypoglycemic hypoketotic encephalopathy. Cardiomyopathy, skeletal muscle weakness, and mildly elevated creatine kinase levels occasionally occur, which have a considerable impact on metabolic decompensation (10, 11). The occurrence of death due to cardiac failure before diagnosis suggests that PCD can be fatal without treatment. Some infants of asymptomatic maternal patients with PCD have been reported to have low C_0 levels, which emphasize the necessity of universal newborn screening, including the screening of asymptomatic newborns and their mothers (12). Infants with low levels of C_0 were screened for PCD via newborn screening programs through tandem mass spectrometry (MS/MS) (13, 14). With the application of MS/MS in neonatal screening, increasing numbers of children in our area are examined, diagnosed and treated early, and the prognosis has been favorable.

Here, ten patients with PCD were screened by MS/MS and diagnosed by DNA sequence analysis, while six infants whose mothers were diagnosed with PCD exhibited low C_0 levels at the first neonatal screening. The clinical, biochemical, and molecular characteristics of the six infants were subsequently analyzed. In addition, two other families were conducted PCD, and the younger family members presented low C_0 levels.



PATIENTS AND METHODS

Study Population

From November 2015 to December 2017, 236,368 newborns were recruited for PCD screening at the Genetic Medicine Center in Xuzhou Maternity and Child Health Care Hospital. C_0 levels were measured by MS/MS, with a dried spot of 2 mg/dL (134 μ mol/L) whole blood that was collected from the infants' plantar surface 48–72 h after birth. Subsequently, the NeoBase Non-derivatized MSMS Kit (PerkinElmer, Finland) was used for preparation of samples at room temperature (21–24°C) and an appropriate humidity (50–70%). Several instrumental parameters of the MS/MS are as follows: ion mode-electrospray⁺/MS/MS; source temperature-120°C/desolvation temperature-350°C. Informed consent was issued by the guardians of all the patients before clinical testing.

MS/MS Analysis and Diagnostic Criteria

The plasma C_0 concentration was quantified by MS/MS, with a cut-off value of 2 mg/dL (134 μ mol/L) in the whole blood. The cut-off value of C_0 was 9.63–54 μ mol/L.

Patients were diagnosed with PCD based on the following criteria (4):

- (1) Newborns with a C_0 value <9.63 μ mol/L were rescreened; if the C_0 value was <5.0 μ mol/L in the second screening, the newborn was diagnosed with PCD, and the mothers' C_0 levels were tested to exclude maternal carnitine deficiency.
- (2) Newborns with persistent low C_0 levels (normal >9.63 μ mol/L) were readmitted for further diagnosis.
- (3) Two pathogenic mutations were detected in the *SLC22A5* gene of the OCTN; one mutation was derived from the father and the other was derived from the mother (4).

DNA Sequence Analysis, Treatment, and Follow-Up

Infants and/or their family members who had abnormal results from the PCD screening provided blood samples, which were sent to Bioscan Genomics, Hangzhou. Subsequently, the targeted DNA sequence was mapped and analyzed by comparing the DNA sequence with a genetic diagnosis panel of hereditary metabolic diseases covering 51 diseases and 98 genes. One of the panels included 16 fatty acid metabolism diseases and 21 genes, including *SLC22A5*. Peripheral blood samples from the patients were used to extract genomic DNA with the omega Genomic DNA Extraction Kit (omega Biotech, USA), and genealogies were determined via Sanger sequencing.

All patients, including those with maternal PCD, were asymptomatic when diagnosed. After diagnosis, 100–300 mg/kg/day of L-carnitine was granted orally 3–4 times per day. The patients were monitored in the clinic once every 2–3 weeks during the initial treatment and then once every 3 months thereafter or a little longer after serum levels of C_0 stabilized and returned to normal. A normal serum level of C_0 is approximately 20 μ mol/L. C_0 is considered a reliable marker during treatment, and after treatment, the levels of a series of other acylcarnitines were found to be normal (15). The youngest infant with PCD was monitored for 2.3 years after birth.

Bioinformatics and Statistical Analysis

Fifty amino acid sequences of *SLC22A5* were downloaded from the NCBI database. Then, sequence logo of *SLC22A5* was built by with WebLogo (16). The point mutation sites in this study were well-conserved in the wild-type (*SLC22A5*). Electrostatic interactions played a crucial role in the process of protein function. To investigate changes in electrostatic properties caused by mutations, adaptive Poisson-Boltzmann solver (APBS) and PDB2PQR were applied to each mutant and wild-type (17). The pqr file of each structure was generated using the PDB2PQR program. The dx file of each structure was generated by utilizing APBS. The pqr file and dx file were then uploaded in VMD to show the molecular surface electrostatic potential map (18). High-quality 3-D images of the protein were drawn by PyMOL (19).

SPSS 16.0 statistical software package was used (SPSS Inc., Chicago, IL, USA). Logistic regression was performed with the 200,000 healthy newborns in Xuzhou. A confidence interval of 0.5–99.5% was selected as standard clinic reference interval.

RESULTS

Clinical and Biochemical Description

C_0 levels in dry blood spots were detected through MS/MS. We collected clinical data of PCD patients from November 2015 to December 2017 and then analyzed the levels of C_0 , which was described as a normal distribution (Figure 1). Statistically, the reference interval was in line with the percentiles of the 200,000 health screening samples in Xuzhou. We chose a confidence

interval of 0.5–99.5%, [CI]: 9.65–54.59. As a consequence, the reference range of C_0 in clinic was 9.65–54.59 $\mu\text{mol/L}$ in Xuzhou.

Among the 236,368 newborns screened by MS/MS, 186 infants initially exhibited C_0 levels that were less than the normal range. In addition, the PCD screening protocol has been modified in daily practice, as showed in Figure 2. Out of 186 infants that were suspected to have PCD from the neonatal screening, 16 of the 186 cases were confirmed to have *SLC22A5* mutations. Ten newborns were diagnosed with PCD, and another six individuals were diagnosed with maternal PCD (theoretically 1 in 23,637 in our region). However, data on PCD incidence are limited and insufficient, which is partially attributed to asymptomatic individuals. To evaluate the incidence of PCD, further work should be performed. Ten PCD patients included eight males and two females; the PCD patients were full-term children with normal birth weight, with the exception of Case 7 and had no clinical symptoms. The clinical and biochemical characteristics of the ten patients with PCD are showed in Table 1. The infant in Case 9 had a 12-year-old sister, and the infant in Case 10 had a twin-brother. Neither patient was deemed to have any significant medical history. The Case 9 infant's sister had the same genetic mutations as the Case 9 infant; however, the Case 10 infant's brother had a normal serum carnitine level.

Clinical Characteristics of Maternal PCD

In six families, the infants were identified to be free of PCD; conversely, their mothers were diagnosed. The six families came from different counties in Xuzhou; Families 1 and 3 were from Peixian, Family 2 was from Suining, Family 4 was from Xinyi and

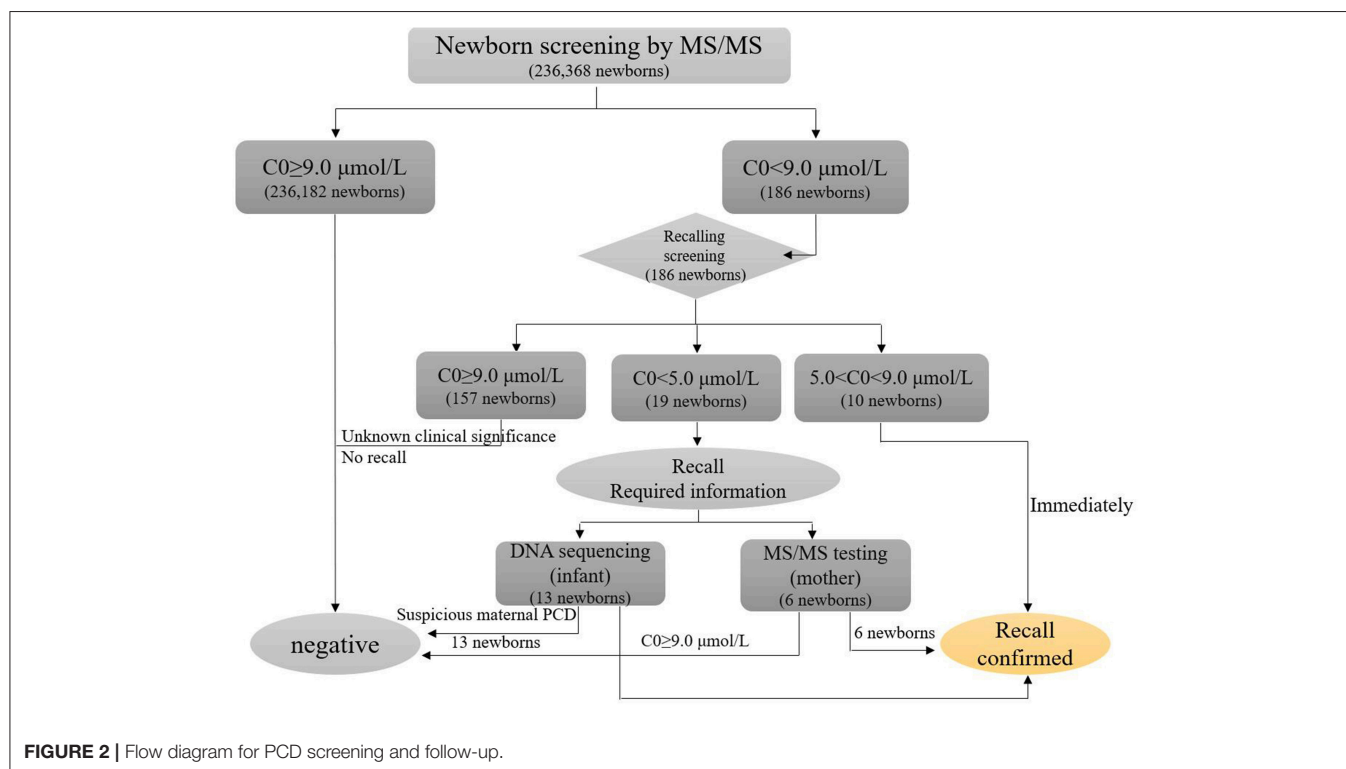


TABLE 1 | Clinical, biochemical, and molecular characteristics of the ten cases of children with PCD.

Case	Current age	Sex	Birth weight (kg)	Gestational weeks	C ₀ level, $\mu\text{mol/L}$		SLC22A5 gene mutation	
					Primary screening	After treatment(mean)	Allele 1 (maternal)	Allele 2 (paternal)
1	2y4m	M	3.00	40	6.74	30.23	c.1400C > G, p. S467C	c.1400C > G, p. S467C
2	1y5m	M	3.10	38	4.02	56.60	c.1400C > z G, p. S467C	c.1400C > G, p. S467C
3	1y3m	F	3.25	39 + 4	6.36	42.81	c.1093A > C, p. T365P	c.1400C > G, p. S467C
4	1y2m	M	3.40	39 + 1	7.05	3.71	c.428C > T, p. P143L	c.1400C > G, p. S467C
5	9m	M	3.50	39 + 4	7.36	36.94	c.1400C > G, p. S467C	c.1400C > G, p. S467C
6	7m	F	3.00	39 + 5	7.18	32.75	c.1400C > G, p. S467C	c.1400C > G, p. S467C
7	7m	M	2.50	33 + 2	7.82	-	c.761G > A, p. R254Q	c.865C > T, p. R289*
8								
Infant	1y7m	M	3.20	41 + 1	7.80	26.78	c.1195C > T, p. R399W	c.1400C > G, p. S467C
Mother					18.56	-	c.1195C > T, p. R399W	-
Father							-	c.1400C > G, p. S467C
9								
Infant	1y3m	M	3.30	39 + 2	6.89	15.01	c.1400C > G, p. S467C	c.92C > T, p. P31L
Mother					21.09	-	c.1400C > G, p. S467C	-
Father							NT	
Sister	12y7m	F	3.00	39	3.25	11.64	c.1400C > G, p. S467C	c.92C > T, p. P31L
10								
Infant	7m	M	3.60	41 + 2	7.65	14.56	c.1400C > G, p. S467C	c.1195C > T, p. R399W
Mother					15.45	-	c.1400C > G, p. S467C	-
Father							NT	
Brother		M	3.50	41 + 2	11.46	-	NT	

F, Female; M, Male; y, year; m, month; NT, not tested. *, nonsense mutation.

the last two families were from Fengxian. All six newborns were normal-birth-weight infants who had unremarkable neonatal physical examinations. All births were by eutocia, and none of the mothers had complications or discomfort. The newborn screening of each infant was performed within 7 days after birth, and if the infant exhibited a low C₀ value, the value was verified by a subsequent plasma carnitine analysis (Table 2). Subsequently, all individuals received carnitine supplementation at a dose of 100–300 mg/kg/day. After treatment with L-carnitine, follow-up carnitine profiles showed normal or elevated carnitine levels in all infants. The recent follow-up monitoring of C₀ levels in the six infants took place at the ages of 1, 3, 3, 2, 2, and 3 months. All of the infants were asymptomatic and were developing age appropriately.

Table 2 shows a clear trend of decreasing of plasma carnitine levels in the six mothers. The mothers of Families 1 and 5 were 24 years old, while both the mothers of Families 2 and 4 were 28 years old. The mothers of Families 3 and 6 were 26 and 27 years old, respectively. None of the mothers were reported to have significant medical history. Each infant was their first child. The mothers were given 1–2g of L-carnitine daily. After treatment, C₀ values of the mothers of Families 2, 4, and 6 increased to within the normal range. However, the mother of Family 3 who did not receive carnitine supplementation sustained a low carnitine level (Table 2). Unfortunately, the mothers of Families 1 and 5 did not follow-up after the initial detection for several reasons.

In summary, of the 236,368 newborns screened by MS/MS, 29 (0.012%) had C₀ levels that were below the cut-off values and were recorded as positive for PCD first the screening. According to the follow-up testing, 10 cases of PCD and 6 cases of maternal PCD, were confirmed and treated in our center. The other 13 cases with no mutations were regarded as negative for PCD in the clinic after a period of monitoring. As showed in Figure 3, C₀ was primarily low in PCD patients, while the level of C₀ was not different between healthy and PCD newborns. The serum levels of C₀ in newborns with maternal PCD were far lower than the levels of other newborns at the follow-up screening, (as showed in Figure 3), but the distribution of the population without mutations was near the normal level. To avoid misdiagnose, further data collection is required due to other rare forms of PCD that can affect the diagnosis. Almost no individuals in the mutation group were followed up, but the results were normal at present.

SLC22A5 Gene Sequencing Results

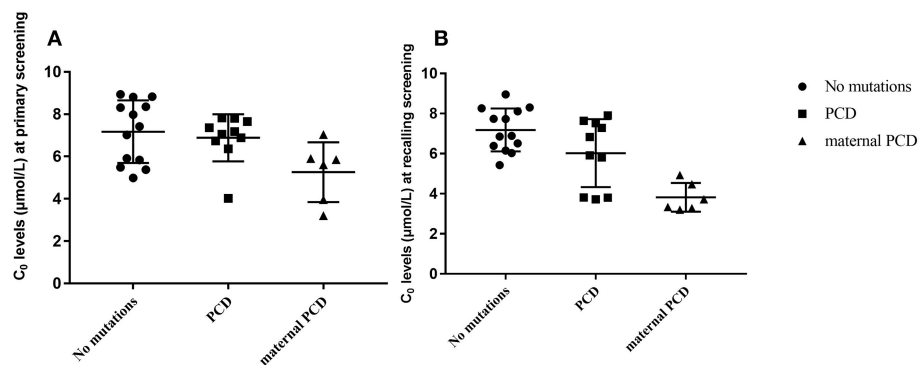
DNA sequencing was further performed in 16 newborns with low C₀ levels from 16 separate families. The SLC22A5 gene mutations are described in Tables 1, 2.

Infants with PCD were shown to have two mutations, one from each of their parents, that is so-called paternal and maternal alleles. The patients in Cases 3, 4, and 9 are being detected with compound heterozygous for one novel missense variant,

TABLE 2 | Plasma carnitine and *SLC22A5* gene sequencing results for reported maternal PCD.

Newborn screening free carnitine, $\mu\text{mol/L}$ 9.63-54		Plasma carnitine, $\mu\text{mol/L}$		SLC22A5 gene sequencing	
		Timing relative to carnitine supplementation	Free carnitine(C ₀) 9.63-54	Allele 1	Allele 2
Family 1					
Infant	7.04	Before	7.04	c.1400C > G, p. S467C	–
		6 days after	8.47		
Mother		Before	1.98	c.1400C > G, p. S467C	c.1400C > G, p. S467C
		After	-		
Family 2					
Infant	5.61	Before	13.28	NT	
		2 weeks after	16.84		
Mother		Before	5.66	c.797C > T, p. P266L	c.1400C > G, p. S467C
		6 weeks after	20.37		
Family 3					
Infant	5.85	Before	3.72	c.95A > G, p. N32S	-
		2 months after	48.23		
Mother		Before	1.73	c.95A > G, p. N32S	c.1400C > G, p. S467C
		2 months after	4.10		
Family 4					
Infant	3.20	Before	12.49	c.1400C > G, p. S467C	-
		2 months after	20.32		
Mother		Before	3.62	c.1400C > G, p. S467C	c.1400C > G, p. S467C
		9 weeks after	16.12		
Family 5					
Infant	5.90	Before	4.93	c.1400C > G, p. S467C	-
		5 weeks after	74.64		
Mother		Before	2.05	c.1400C > G, p. S467C	c.1400C > G, p. S467C
		After	-		
Family 6					
Infant	3.96	Before	3.34	c.1400C > G, p. S467C	-
		3 months after	87.03		
Mother		Before	2.73	c.1462C > T, p. R488C	c.1400C > G, p. S467C
		3 months after	16.80		

NT, not tested.

**FIGURE 3 |** Metabolite measurement in PCD compared to no mutation, PCD and maternal PCD patients. **(A)** Primary screening result of C_0 . **(B)** Recalling screening results of C_0 .

which was a paternal or maternal allele. The most common mutation of the *SLC22A5* gene in our area is c.1400C > G (p. S467C) mutation. **Figure 4** shows the genetic family map of Cases 8, 9, and 10. The infant in Case 8 was compound heterozygous for c.1195C > T (p. R399W) and c.1400C > G (p. S467C) because his mother and father were heterozygous for the missense mutations c.1195C > T (p. R399W) and c.1400C > G (p. S467C), respectively. However, the new mutation type c.92C > T (p. P31L) may have been from a father who was not tested. The Case 10 infant was identified as compound heterozygous for two missense variants c.1195C > T (p. R399W) and c.1400C > G (p. S467C), whereas his mother was a carrier for only the c.1400C > G (p. S467C) variant. Neither the infant's twin-brother nor their father were tested.

In the maternal PCD families, the mothers of Families 1, 4 and 5 were found to be homozygous for a classic missense mutation, c.1400C > G (p. S467C). The infants were carriers of the c.1400C > G (p. S467C) variant. The mother of Family 3 was compound heterozygous for two missense mutations, c. 95A > G (p. N32S) and c.1400C > G (p. S467C). Additionally, the mother of Family 6 was also compound heterozygous for two common missense mutations, c.1462C > T (p. R488C) and c.1400C > G (p. S467C). In these two families, carriers were verified with c. 95A > G (p. N32S) or c.1400C > G (p. S467C) mutations. The mother of Family 2 was shown to be compound heterozygous, carrying a novel missense mutation, c.797C > T (p. P266L), and a classified missense mutation, c.1400C > G (p. S467C). The infant in Family 2 was assessed clinically and was found to be free of carnitine deficiency; DNA analysis was not performed, but she was presumed to be an obligate carrier given the diagnosis of her mother.

The Frequencies and Locations of *SLC22A5* Gene Mutations

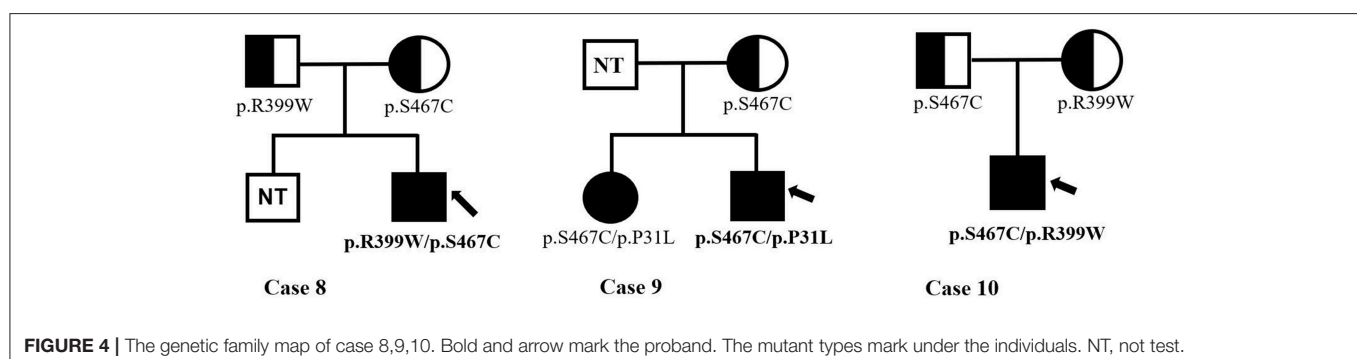
It can be observed in the data that a total of 43 mutant alleles with 10 different mutations/unclassified missense variants (**Table 3** and **Figure 5**) were verified in 26 cases (**Tables 1, 2**). Except for five classical mutations, the additional mutations were not classified and might be considered unclassified types. Five novel missense variants were considered unclassified mutations. Indeed, previous studies have suffered from methodological limitations and rare cases, which have found it difficult to determine the significance of these mutations. Some statistical

analyses cannot be conducted to explain the correlation between the apparent “novel” mutations and their clinical relevance. However, based on the landscapes of the electrostatic potential for the five *SLC22A5* proteins, the point mutations in the wild-type *SLC22A5* protein could affect its surface electrostatic potential which plays an important role in protein-protein interaction. This result suggests that these genetic substitutions probably result in various *SLC22A5* functions as showed in **Figure S1**. Moreover, further work will be performed in the future. Among the 10 different mutations that were identified, the most frequently occurring mutation was the c.1400C > G (p. S467C) with a frequency of ~72% (31/43). Serine at position 467 is revolutionaries conserved, as the amino acid change in p. S467C can affect the transmembrane domain 11 (TM11) of the OCTN2 protein, resulting in PCD. Seven cases were homozygous for the constant mutations of c.1400C>G. Seventeen cases, including the 7 homozygous cases, had 2 mutations; 6 cases carried 1 unclassified variant and 1classified mutation; and other cases harbored two classified mutations. Hence, at least one mutation was detected in each person analyzed. Not only were classical mutations identified but unclassified missense variants in all exons were also identified in our area. However, **Table 3** is shown that more than 72% of mutant alleles (31/43) were located

TABLE 3 | Detected mutations and unclassified variants, and their frequencies and locations.

Mutations/unclassified missense variants	Freq	Exons	Domains
c.92C > T (p. P31L)	2	Exon 1	TM 1
c.95A > G (p. N32S)	2		
c.428C > T (p. P143L)	1	Exon 2	TM 2
c.761G > A (p. R254Q)	1	Exon 4	L 5
c.797C > T (p. P266L)	1		TM 6
c.865C > T (p. R289*)	1	Exon 5	L 6
c.1093A > C (p. T365P)	1	Exon 7	TM 7
c.1195C (p. R399W)	3		L 8
c.1400C > G (p. S467C)	22	Exon 8	TM 11
c.1462C > T (p. R488C)	1	Exon 9	L 11
Total	35		

*Bold novel changes in this study; Gray, unclassified missense variant. TM, transmembrane domain. L, inter-transmembrane loop. *, nonsense mutation.*



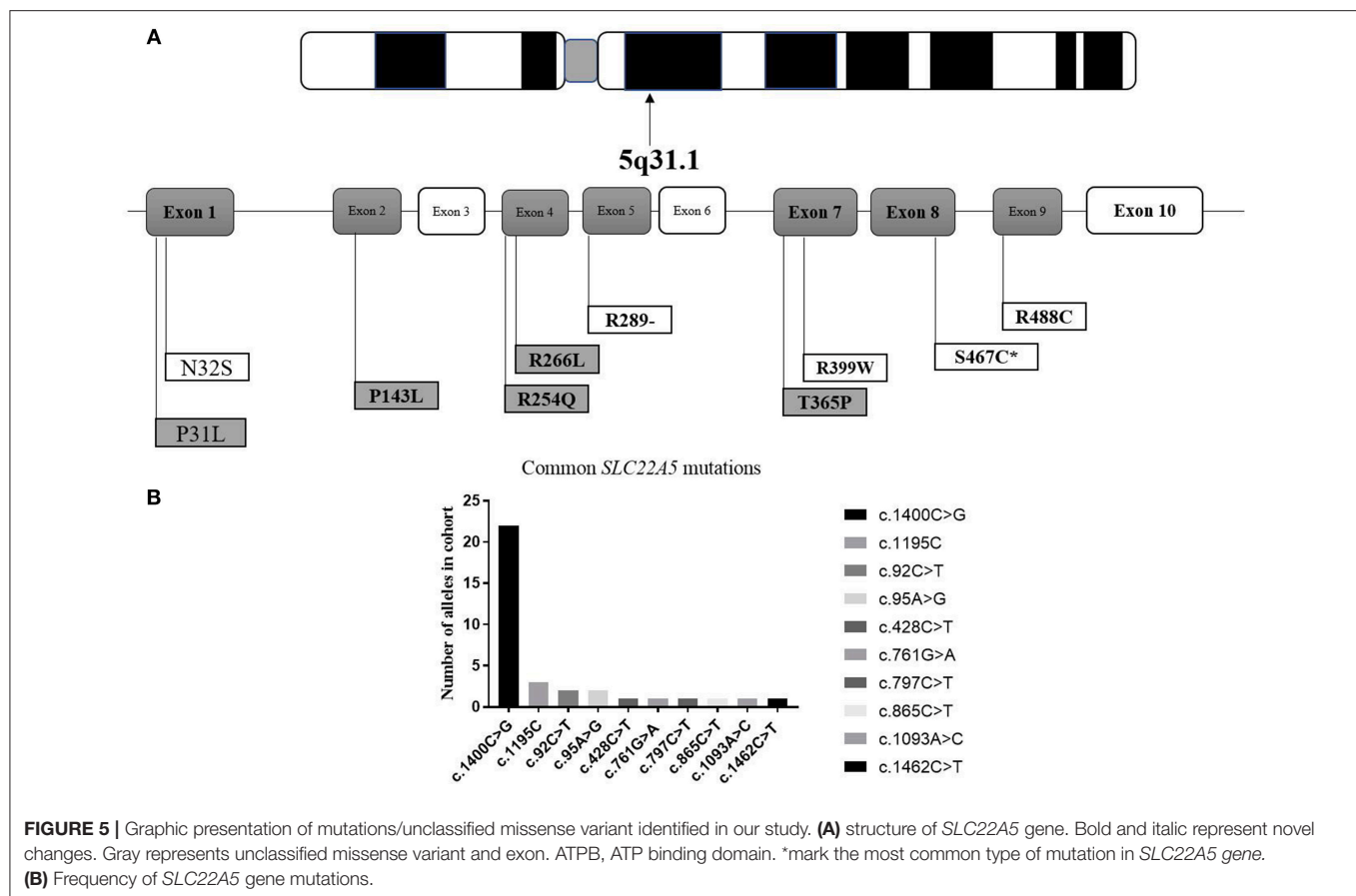


FIGURE 5 | Graphic presentation of mutations/unclassified missense variant identified in our study. **(A)** structure of *SLC22A5* gene. Bold and italic represent novel changes. Gray represents unclassified missense variant and exon. ATPB, ATP binding domain. *mark the most common type of mutation in *SLC22A5* gene. **(B)** Frequency of *SLC22A5* gene mutations.

in exon 8 while the most frequently affected domain was TM11 of the OCTN2 protein.

DISCUSSION

As indicated in the literature review, PCD is a disease caused by a carnitine absorption barrier. Unfortunately, some significant clinical consequences, even death, can occur if PCD is not treated on time. However, PCD can be dramatically improved with carnitine supplementation. Previous investigations have not presented any exact data about the nationwide prevalence of PCD in China. It has been reported that the incidence of PCD was approximately 1:45,000 in Shanghai (20, 21), approximately 1:22,384 in Zhejiang Province (22) and 1:8,938 in Nanjing, Jiangsu Province (4). In this study, 236,368 newborns were screened and ten PCD cases in infants and six maternal PCD cases were identified by MS/MS from 2015 to 2017 in Xuzhou, China. Thus, the incidence of PCD in the Xuzhou area was 1:23,637, which was similar to that of Zhejiang Province.

The *SLC22A5* gene contains 10 exons and 3 introns located on chromosome 5q31.1. To date, over 110 mutations in the *SLC22A5* gene have been associated with PCD, while c.1400C>G (p. S467C) is considered to be the most common mutation among Chinese patients. Subsequently, 7 pathogenic homozygous mutations were detected in 16 patients and other heterozygous mutations were also noted. This result is in accordance with

the findings of earlier reports indicating that the frequency of the c.1400C > G (p. S467C) mutation was as high as 72% (31/43). Unfortunately, what is not yet clear is the prevalence of PCD false-negative assessments in the primary screening in our area. To account for the technical challenge, a rapid novel 2nd-tier test application might reduce the false-negative rate in the clinical setting.

Traditionally, serum C_0 levels have been assessed (23, 24) based on quantifiable biochemical manifestations and DNA analysis. Sixteen patients showed a lower C_0 level (<9.63 $\mu\text{mol/L}$) at the primary screening; with a mean value of 6.48 $\mu\text{mol/L}$, emphasizing the utility of MS/MS neonatal screening for PCD diagnosis. Extremely low levels of C_0 were observed in these individuals at the second screening, which seems to be consistent with previous findings that showed that PCD patients often exhibit low C_0 levels at the follow-up screening. Asymptomatic carriers are individuals who are identified with heterozygous mutations without clinical consequences due to half-normal carnitine transport in their fibroblasts. Despite slightly lower C_0 levels, these infants had no impact to the degree of the PCD patients. Collectively, plasma carnitine levels alone are not a reliable indicator of a PCD diagnosis, and DNA analysis should also be taken into account.

C_0 levels of infants reflect their mothers' carnitine levels shortly after birth, because carnitine is provided by the placental tissue matrix to the fetus during intrauterine life (25). Therefore,

one reason infant present with low C_0 levels at the initial screening is that their mothers have PCD. These results reinforced several notions that are in charge of identifying some maternal inborn errors of metabolism (26, 27). To decrease false-positive or false-negative diagnoses, we should also pay more attention to maternal evaluation after the identification of a low C_0 level at the primary screening. There is some evidence that PCD patients might be asymptomatic, such as the mothers of Families 1–6. Lacking of sufficient asymptomatic cases and follow-up, it is not yet clear whether potential health risks exist (12). A possible explanation for these asymptomatic cases might be that asymptomatic adult individuals may be unaware of their own defects. Fatty acid oxidation defects, such as medium chain acyl-CoA dehydrogenase deficiency, might not result in sudden death or other acute conditions unless the individual with the condition is under severe stress (28–30). Hence, it is crucial that asymptomatic PCD patients receive preventive treatments with carnitine supplementation for any potential decompensation attributed to intercurrent illness or stress.

In conclusion, the aim of the research was to determine if the application of MS/MS for neonatal screening combined with DNA analysis can diagnose PCD. Overall, newborn screening can result in the diagnosis of maternal PCD, and it is vital that low maternal C_0 levels are evaluated at the secondary screening. PCD can be present at any age, with a wide phenotypic spectrum ranging from metabolic decompensation in infancy to an asymptomatic presentation in adulthood, or PCD may be asymptomatic but manifest or be exacerbated during pregnancy. To our knowledge, early universal L-carnitine treatment played a significant role in the favorable prognosis of patients with PCD. In addition, the current data highlights the importance of the diagnosis of PCD in neonatal screening. Because of the limited time, this study lacks more clinical cases, including PCD cases with exon skipping and deletions. Despite its exploratory nature, this study offers some insight into maternal PCD in newborn screening. Therefore, there is a definite need for additional research in the future.

REFERENCES

- Kilic M, Ozgul RK, Coskun T, Yucel D, Karaca M, Sivri HS, Tokatli A, et al. Identification of mutations and evaluation of cardiomyopathy in Turkish patients with primary carnitine deficiency. *JIMD Rep.* (2012) 3:17–23. doi: 10.1007/8904_2011_36
- Nezu J, Tamai I, Oku A, Ohashi R, Yabuuchi H, Hashimoto N, et al. Primary systemic carnitine deficiency is caused by mutations in a gene encoding sodium ion-dependent carnitine transporter. *Nat Genet.* (1999) 21:91–4. doi: 10.1038/5030
- Li FY, El-Hattab AW, Bawle EV, Boles RG, Schmitt ES, Scaglia F, et al. Molecular spectrum of SLC22A5 (OCTN2) gene mutations detected in 143 subjects evaluated for systemic carnitine deficiency. *Hum Mutat.* (2010) 31:E1632–51. doi: 10.1002/humu.21311
- Sun Y, Wang YY, Jiang T. Clinical features and genotyping of patients with primary carnitine deficiency identified by newborn screening. *J Pediatr Endocrinol Metabol.* (2017) 30:879–83. doi: 10.1515/jpem-2017-0002
- Amat di San Filippo C, Pasquali M, Longo N. Pharmacological rescue of carnitine transport in primary carnitine deficiency. *Hum Mutat.* (2006) 27:513–23. doi: 10.1002/humu.20314

ETHICS STATEMENT

This study was approved by the Ethics Committee of Xuzhou Maternity and Child Health Care Hospital, and the individual written informed consents were obtained from the parents of the infants involved in this study.

AUTHOR CONTRIBUTIONS

WZ and MG: conception and design of study; WZ, HL, and YZ: acquisition of data; WZ, HL, TH, and CW: analysis and/or interpretation of data; WZ, HL, MG, and CW: drafting the manuscript; WZ, TH, and CW: revising the manuscript critically for important intellectual content.

ACKNOWLEDGMENTS

We thank all those patients for their participation in this study. In addition, special thanks to the personnel from Xuzhou Maternity and Child Health Care Hospital and Zhejiang Biosan Biochemical Technologies Co., Ltd, who provided technical assistance in this study. Simultaneously, for help in bioinformatics analysis, we would like to express our heartfelt gratitude to Dr. Chuanjun Shu, who works for Department of Bioinformatics, School of Biomedical Engineering and Informatics, Nanjing Medical University. Funding for this program is provided by the Artificial Intelligence-Aided Diagnosis Platform for Genetic Metabolic Disease (Grant number: 2017YFC1001703). Ultimately, the reviewers have also contributed considerably to the publication of this paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fped.2019.00050/full#supplementary-material>

- Longo N. Primary carnitine deficiency and newborn screening for disorders of the carnitine cycle. *Ann Nutr Metabol.* (2016) 68 (Suppl 3):5–9. doi: 10.1159/000448321
- Rasmussen J, Nielsen OW, Janzen N, Duno M, Gislason H, Kober L, et al. Carnitine levels in 26,462 individuals from the nationwide screening program for primary carnitine deficiency in the Faroe Islands. *J Inherit Metab Dis.* (2014) 37:215–22. doi: 10.1007/s10545-013-9606-2
- Rasmussen J, Nielsen OW, Lund AM, Kober L, Djurhuus H. Primary carnitine deficiency and pivalic acid exposure causing encephalopathy and fatal cardiac events. *J Inherit Metab Dis.* (2013) 36:35–41. doi: 10.1007/s10545-012-9488-8
- Niu DM, Chien YH, Chiang CC, Ho HC, Hwu WL, Kao SM, et al. Nationwide survey of extended newborn screening by tandem mass spectrometry in Taiwan. *J Inherit Metab Dis.* (2010) 33:S295–305. doi: 10.1007/s10545-010-9129-z
- Kepka A, Szajda SD, Waszkiewicz N, Pludowski P, Chojnowska S, Rudy M, et al. Carnitine: function, metabolism and value in hepatic failure during chronic alcohol intoxication. *Postępy Higieny i Medycyny Doswiadczalne* (2011) 65:645–53. doi: 10.5604/17322693.962226

11. Saudubray JM, and Garcia-Cazorla A. Inborn errors of metabolism overview: pathophysiology, manifestations, evaluation, and management. *Pediatr Clin North Am.* (2018) 65:179–208. doi: 10.1016/j.pcl.2017.11.002
12. El-Hattab AW, Li YF, Shen J, Powell BR, Bawle EV, Adams DJ, et al. Maternal systemic primary carnitine deficiency uncovered by newborn screening: clinical, biochemical, and molecular aspects. *Genet Med.* (2010) 12:19–24. doi: 10.1097/GIM.0b013e3181c5e6f7
13. Guo K, Zhou X, Chen X, Wu Y, Liu C, Kong Q. Expanded newborn screening for inborn errors of metabolism and genetic characteristics in a Chinese population. *Front Genet.* (2018) 9:1–7. doi: 10.3389/fgene.2018.00122
14. Kim JU, Kannan K. Method for the determination of iodide in dried blood spots from newborns by high performance liquid chromatography tandem mass spectrometry. *Anal Chem.* (2018) 90:3291–8. doi: 10.1021/acs.analchem.7b04827
15. Rasmussen J, Thomsen JA, Olesen JH, Lund TM, Mohr M, Clementsen J, et al. Carnitine levels in skeletal muscle, blood, and urine in patients with primary carnitine deficiency during intermission of l-carnitine supplementation. *J Int Med Dentistr Rep.* (2014) 20:103–111. doi: 10.1007/8904_2014_398
16. Crooks GE, Hon G, Chandonia MJ, Brenner SE. WebLogo: a sequence logo generator. *Genome Res.* (2004) 14:1188–90. doi: 10.1101/gr.849004
17. Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW, Nielsen JE, Baker NA. Web servers and services for electrostatics calculations with APBS and PDB2PQR. *J Comput Chem.* (2011) 32:1488–91. doi: 10.1002/jcc.21720
18. Eargle J, Wright D, Luthey-Schulten Z. Multiple alignment of protein structures and sequences for VMD. *Bioinformatics* (2006) 22:504–6. doi: 10.1093/bioinformatics/bti825
19. McKay T, Hart K, Bernstein H, Tedla-Boyd W, Craig P. Protein function prediction using ProMOL and PyMOL. *Ann Funct Anal.* (2015) 6:155–165.
20. Han L, Wang F, Wang Y, Ye J, Qiu W, Zhang H, et al. Analysis of genetic mutations in Chinese patients with systemic primary carnitine deficiency. *Eur J Med Genet.* (2014) 57:571–5. doi: 10.1016/j.ejmg.2014.08.001
21. Han LS, Ye J, Qiu WJ, Zhang HW, Wang Y, Ji WJ, et al. Primary carnitine deficiency in 17 patients: diagnosis, treatment and follow up. *Zhonghua Er Ke Za Zhi* (2012) 50:405–9.
22. Pingyue M. *Analysis the Gene Types and Clinical Feature of 50 Newborns With Primary Carnitine Deficiency.* Pediatrics Zhejiang University (2015).
23. Gallant NM, Leydiker K, Wilnai Y, Lee C, Lorey F, Feuchtbaum L, et al. Biochemical characteristics of newborns with carnitine transporter defect identified by newborn screening in California. *Mol Genet Metabol.* (2017) 122:76–84. doi: 10.1016/j.ymgme.2017.06.015
24. Rasmussen J, Hougaard DM, Sandhu N, Fjaellegaard K, Petersen PR, Steuerwald U, Lund AM. Primary carnitine deficiency: is foetal development affected and can newborn screening be improved? *J Int Med Dentistr Rep.* (2017) 36:35–40. doi: 10.1007/8904_2016_30
25. Magoulas PL, El-Hattab AW. Systemic primary carnitine deficiency: an overview of clinical manifestations, diagnosis, and management. *Orphanet J Rare Dis.* (2012) 7. 68. doi: 10.1186/1750-1172-7-68.
26. Pierpont ME, Breningstall GN, Stanley CA, Singh A. Familial carnitine transporter defect: a treatable cause of cardiomyopathy in children. *Am Heart J.* (2000) 139:S96–106. doi: 10.1067/mhj.2000.103921
27. Vijay S, Patterson A, Olpin S, Henderson MJ, Clark S, Day C, et al. Carnitine transporter defect: diagnosis in asymptomatic adult women following analysis of acylcarnitines in their newborn infants. *J Inherit Metabol Dis.* (2006) 29:627–30. doi: 10.1007/s10545-006-0376-y
28. Dobrowolski SF, Ghaloul-Gonzalez L, Vockley J. Medium chain acyl-CoA dehydrogenase deficiency in a premature infant. *Pediatric Rep.* (2017) 9:7045. doi: 10.4081/pr.2017.7045
29. Lim SC, Tajika M, Shimura M, Carey KT, Stroud DA, Murayama K, et al. Loss of the mitochondrial fatty acid beta-oxidation protein medium-chain acyl-coenzyme a dehydrogenase disrupts oxidative phosphorylation protein complex stability and function. *Sci Rep.* (2018) 8:153. doi: 10.1038/s41598-017-18530-4
30. Smon A, Groselj U, Debeljak M, Zerjav Tansek M, Bertok S, Avbelj Stefaniya M, et al. Medium-chain acyl-CoA dehydrogenase deficiency: Two novel ACADM mutations identified in a retrospective screening. *J Int Med Res.* (2018) 46:1339–48. doi: 10.1177/0300060517734123

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhou, Li, Huang, Zhang, Wang and Gu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Novel Homozygous Frameshift Variant in *XYLT2* Causes Spondyloocular Syndrome in a Consanguineous Pakistani Family

Mehran Kausar^{1,2,3}, Elaine Guo Yan Chew^{4,5}, Hazrat Ullah⁶, Mariam Anees¹, Chiea Chuen Khor⁵, Jia Nee Foo^{4,5}, Outi Makitie^{3,7*} and Saima Siddiqi^{2*}

¹ Department of Biochemistry, Quaid-i-Azam University, Islamabad, Pakistan, ² Institute of Biomedical and Genetic Engineering (IBGE), Islamabad, Pakistan, ³ Folkhälsan Institute of Genetics, Helsinki, Finland, ⁴ Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore, ⁵ Human Genetics, Genome Institute of Singapore, A*STAR, Singapore, Singapore, ⁶ National Institute of Rehabilitation Medicine (NIRM), Islamabad, Pakistan, ⁷ Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Muhammad Tariq,
University of Tabuk, Saudi Arabia
Muhammad Umair,
Ministry of National Guard Health
Affairs (MNGHA), Saudi Arabia
Koichiro Wasano,
Tokyo Medical Center (NHO), Japan

*Correspondence:

Outi Makitie
outi.makitie@helsinki.fi
Saima Siddiqi
saimasiddiqi2@gmail.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 08 November 2018

Accepted: 12 February 2019

Published: 05 March 2019

Citation:

Kausar M, Chew EGY, Ullah H,
Anees M, Khor CC, Foo JN, Makitie O
and Siddiqi S (2019) A Novel
Homozygous Frameshift Variant
in *XYLT2* Causes Spondyloocular
Syndrome in a Consanguineous
Pakistani Family.
Front. Genet. 10:144.
doi: 10.3389/fgene.2019.00144

We report on three new patients with spondyloocular syndrome (SOS) in a consanguineous Pakistani family. All three patients present progressive generalized osteoporosis, short stature, recurrent fractures, hearing loss and visual impairments. WES revealed a novel homozygous frameshift variant in exon 11 of *XYLT2* (NG 012175.1, NP_071450.2) resulting in loss of evolutionary conserved amino acid sequences (840 – 865/865) at C-terminus p.R840fs*115. Sanger Sequencing confirmed the presence of the novel homozygous mutation in all three patients while the parents were heterozygous carriers of the mutation, in accordance with an autosomal recessive inheritance pattern. Only nine variants worldwide have previously been reported in *XYLT2* in patients with SOS phenotype. These three patients with novel homozygous variant extend the genotypic and phenotypic spectrum of SOS.

Keywords: spondyloocular syndrome (SOS), whole-exome-sequencing (WES), osteoporosis, xylosyltransferase II (*XYLT2*), cataract

INTRODUCTION

Spondyloocular syndrome (OMIM 605822), a very rare form of autosomal recessive genetic skeletal dysplasia, was first defined and reported by Schmidt et al. (2001), Taylan et al. (2017). In recent years, spondyloocular syndrome (SOS) has been linked to variant in the xylosyltransferase II encoded by *XYLT2* (MIM 608125) (Munns et al., 2015; Taylan et al., 2016). *XYLT2* is involved in the biosynthesis of proteoglycans (PGs).

PGs, surface-associated and extracellular matrix proteins, play a vital role in sustaining the homeostasis in various tissues including skin, bone and cartilage (Couchman and Pataki, 2012). PGs are composed of core protein on which glycosaminoglycan (GAGS) chains, made up of tetrasaccharide linkers, are assembled. Mainly two types of sulfated PGs, heparin sulfate PGs and chondroitin sulfate/dermatan sulfate PGs are involved (Couchman and Pataki, 2012). Synthesis of tetrasaccharide linker molecules is catalyzed by four consecutive enzymatic reactions, first reaction requiring the transfer of xylose from uridine diphosphate-xylose (UDP-Xyl) to particular serine residue of the core protein by xylosyltransferase I (MIM 608124) and II. Next step involves the addition of two galactose residues by galactosyltransferase I and II, encoded by *B4GALT7* (MIM 604327) and *B3GALT6* (MIM 615291), respectively. Final reaction involves the transfer

of glucuronic acid by glucuronosyltransferase I, encoded by *B3GAT3* (MIM 606374). Variants in genes encoding these five enzymes have been reported in various genetic skeletal diseases, which are collectively known as linkeropathies (Mizumoto et al., 2015; Taylan et al., 2016).

To date, only nine variants in *XYLT2* have been identified in the patients with SOS phenotype worldwide (Taylan et al., 2017; Umair et al., 2017). Here we present a consanguineous Pakistani family with three children with moderate to severe manifestations of SOS due to a novel homozygous frameshift variant in *XYLT2*.

CASE PRESENTATION AND METHOD

We identified a consanguineous Pakistani family in which three family members, 12 and 9 years old brothers (IV:1, IV:3) and their 10 years old cousin (IV:5), were affected with a rare skeletal dysplasia (**Figure 1A,B**). A written informed consent was obtained from the head of family (father) before the collection of blood samples for genetic analyses from all available family members and publication of the results.

All three patients were under treatment in Children's Hospital Lahore, Pakistan. Growth parameters including height and weight of patients, were measured and compared with reference values (Aziz et al., 2012) among Pakistani population (Aziz et al., 2012). The clinical data, including radiographs, were assessed retrospectively. Genetic studies were performed using peripheral blood genomic DNA. We used 1 µg of genomic DNA for targeted enrichment from Patient 1 (IV:1) using the Nimblegen SeqCap EZ Exome v3 kit and barcoded for sequencing on a single lane of a multiplexed 2x151 bp sequencing run on the Illumina HiSeq 2000 platform. This individual's sample was sequenced to a mean coverage of 37.8 reads per target base, with 97% of the target exome covered by 10 or more reads. Reads were mapped using BWA v0.7.17 (Genomes Project et al., 2012) and variants were called using the GATK v2 Unified Genotyper following the recommended guidelines by GATK 'Best practices for variant calling v3' (Genomes Project et al., 2012). Variants present at $\leq 1\%$ frequency in 1000 Genomes, HapMap and ExAC populations were identified from exome sequencing data. SIFT (Sim et al., 2012) and Polyphen2 (Adzhubei et al., 2010) were used to predict effect of each missense variant. The variant of interest was looked up within in-house exome data of 95 Pakistani individuals (16 normal individuals and 79 individuals with an unrelated disease phenotype) which were processed as described above. We used the following primer sequences for the segregation analysis of detected mutations in *XYLT2* by polymerase chain reaction and Sanger sequencing: *XYLT2*-Forward, GCAAGCTGTGACTCAGAAGTA and *XYLT2*-Reverse, AGCCTCGTGACGAACAATAG.

RESULTS

Clinical Findings

There were three patients in the family, the index Patient 1 (IV:1) is a 12 years old boy born to consanguineous healthy parents from

uncomplicated pregnancy at full term. He presented with delayed milestones and multiple compression fractures at the age of 9 months. His first right femoral fracture occurred at 12 months and generalized osteoporosis was noted. Presently, his height is 124 cm (Z-score -2.4) and weight 36 kg (Z-score -0.2). Physical evaluation revealed a low posterior hairline, short and webbed neck, low set ears, shield chest, long fingers and toes. Sclerae were normal and no dental problems were observed. Radiographs revealed generalized osteoporosis, mild to moderate thoracic kyphosis, platyspondyly and increased intervertebral disk space. Intravenous pamidronate treatment was started at the age of 36 months and resulted in reshaping of some vertebrae. At the age of 12 years the patient is ambulant but has unsteady gait due to muscle weakness.

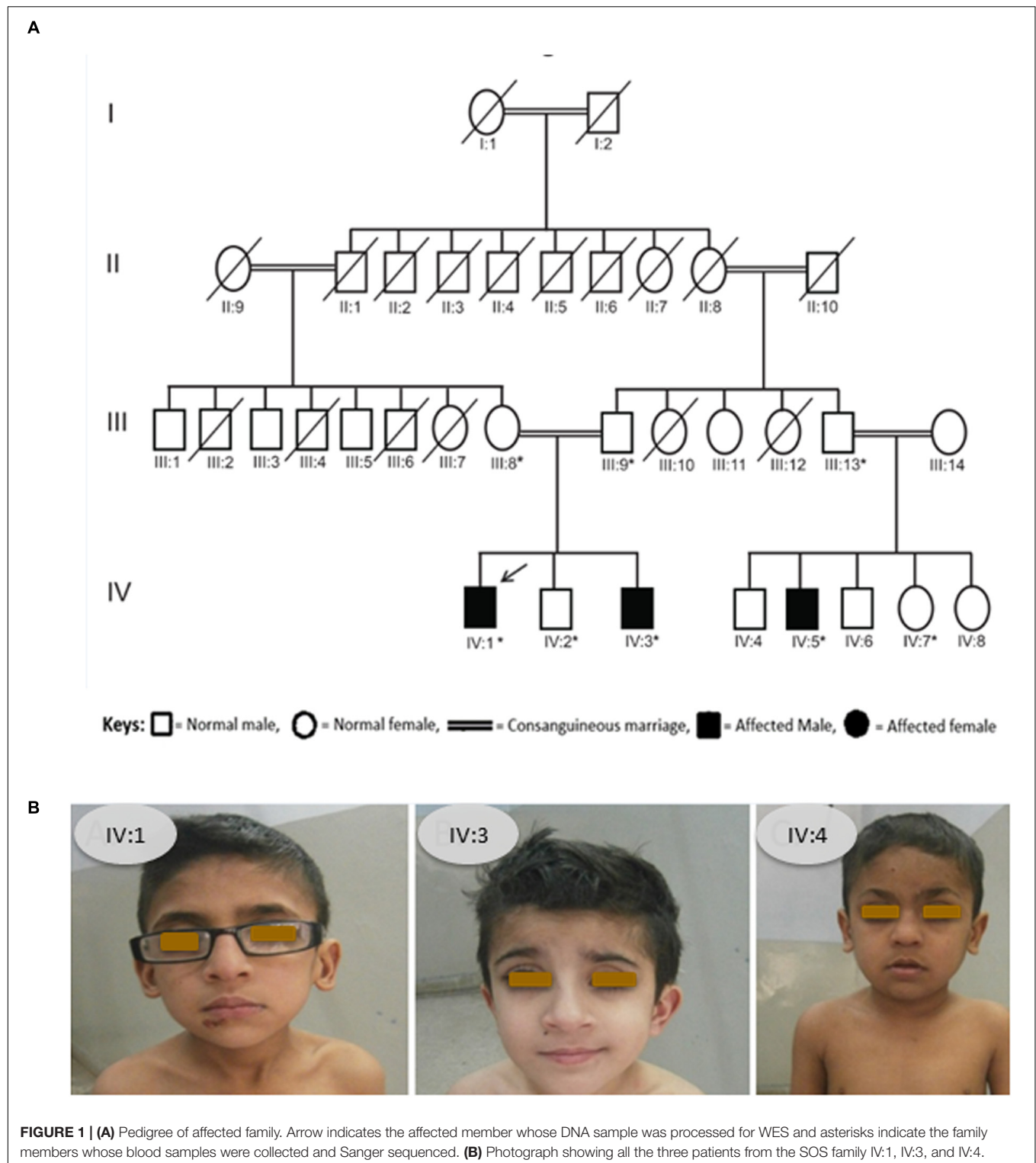
In addition to skeletal problems, the patient was diagnosed with moderate hearing loss at 8 years of age. Eye examination revealed nystagmus and amblyopia, and spontaneous left retinal detachment occurred. Cataracts were noted at 12 years. Learning difficulties were also obvious since early childhood. Biochemical analysis of patient's blood revealed normal levels of calcium, alkaline phosphatase, creatinine and 25-OH-vitamin D.

The 2nd patient (IV:3) is 9 years old, younger brother of patient 1, and had a similar clinical course and manifestations as his elder brother. He was diagnosed with generalized vertebral flattening and multiple compression fractures at the age of 6 months. Currently, his height is 109 cm (Z-score -3.5) and weight 29 kg (Z-score -0.7). He also had pamidronate infusions at the age of 15 months. Reshaping of some vertebrae were noted, but new compression fractures have also occurred. He has sustained recurrent long bone fractures but is ambulant. Mild to moderate hearing loss and vision impairment appeared at the age of 6 years and he also had learning difficulties. Biochemical analysis of patient's blood revealed normal levels of calcium, alkaline phosphatase, creatinine and 25-OH-vitamin D.

The 3rd patient of the family (IV:4) is a 10 years old boy who has had similar clinical course and manifestations as his two older cousins. Presently, his height is 111 cm (Z-score -3.7) and weight 33 kg (Z-score -0.4). Delayed milestones and multiple compression fractures were apparent at the age of 9 months. His first femoral fracture occurred at the age of 18 months and generalized osteopenia was observed; multiple other fractures have occurred thereafter. Physical evaluation revealed a low posterior hairline, short and webbed neck, low set ears, shield chest, long fingers and toes. Sclerae and teeth were normal. Radiographs revealed moderate thoracic kyphosis and platyspondyly. He started intravenous pamidronate treatment at the age of 3 years which improved the compression fractures but did not completely prevent new fractures. Along with skeletal problems he was also diagnosed with hearing loss and visual impairment at the age of 5 years. He was operated on for bilateral cataract. Learning difficulties were observed since early childhood. Blood biochemistry for calcium, alkaline phosphatase, creatinine and 25-OH-vitamin D was normal.

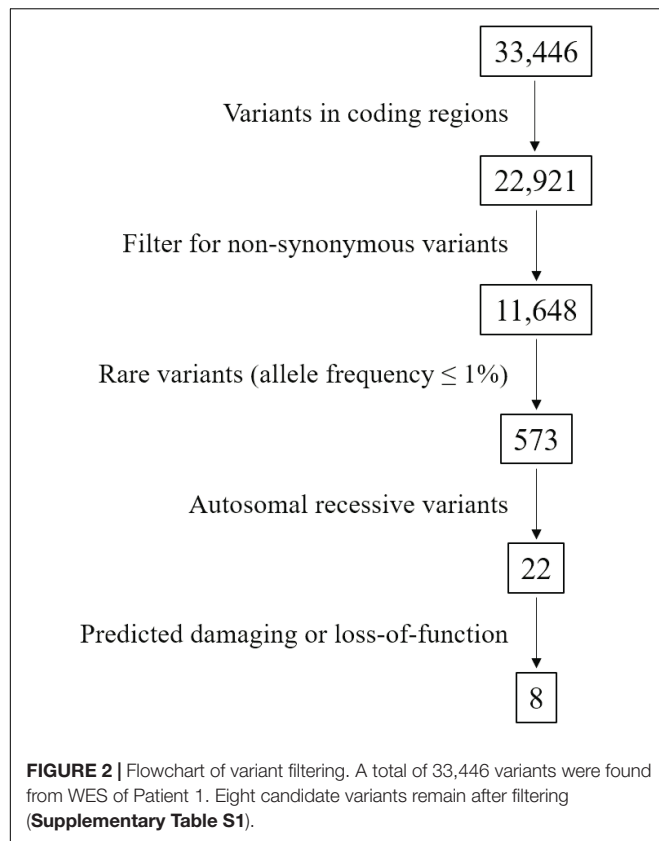
Genetic Findings

We identified a total of 33,446 variants from WES of Patient #1. Eight variants remained after filtering for variants which



localized to coding regions, are non-synonymous, rare (allele frequency $\leq 1\%$), autosomal recessive and predicted to be damaging (Figure 2, Supplementary Table S1). Based on the functions of the genes in which these 8 candidate variants localized and the previously reported gene association with SOS,

we focused on the 17:g.48437571CAG>C deletion variant in *XYLT2* (xylosyltransferase II). The deletion g.48437572delAG (p.R840Tfs*115) which is present at 17q21.33 alters the reading frame, substituting the last 25 amino acids (p.840 – 865) at carboxy terminal end of the xylosyltransferase II. Sanger



sequencing confirmed that all three patients were homozygous for this variant and their parents were heterozygous (Figure 3). Two healthy siblings were negative for the variant. This variant

was not present in exome-sequencing data of 95 Pakistani individuals, and was also not found in the available population databases, including the exome aggregation consortium database (ExAC) and genome aggregation database (gnomAD). This novel variant segregated perfectly with SOS manifestations and was regarded as the cause of the patient's phenotype.

DISCUSSION

We report the phenotypic characteristics and molecular diagnosis of SOS in three children from a consanguineous Pakistani family. All three patients present with generalized osteoporosis, multiple compression fractures, hearing loss and visual impairment, all consistent with the diagnosis of SOS. We report a novel homozygous pathogenic frameshift variant in *XYLT2* in the family under study. Chromosomal location of the *XYLT2* is 17q21.3-q22. *XYLT2* contains 11 exons and 10 introns, and encodes an 865 amino acids long protein. The *XYLT2* is comprised of four domains; N-terminus domain, catalytic domain, a core2/I-branching enzyme domain and a C-terminus domain (Umair et al., 2017). Schmidt et al. (2001) published the first clinical report on SOS in a consanguineous Iraqi family with 6 affected individuals. Alanay et al. (2006) reported in 2006 another patient with SOS in a Turkish family. Munns et al. (2015) identified a pathogenic biallelic frameshift insertion in *XYLT2* as the genetic cause of SOS in a non-consanguineous Australian European family with two affected brothers; another patient had a deleterious homozygous frameshift deletion. Taylan et al. (2016) identified a nonsense pathogenic frameshift variant (p.Arg730*) in a Turkish patient and two different pathogenic missense variants (p.Arg563Gly and p.Leu605Pro)

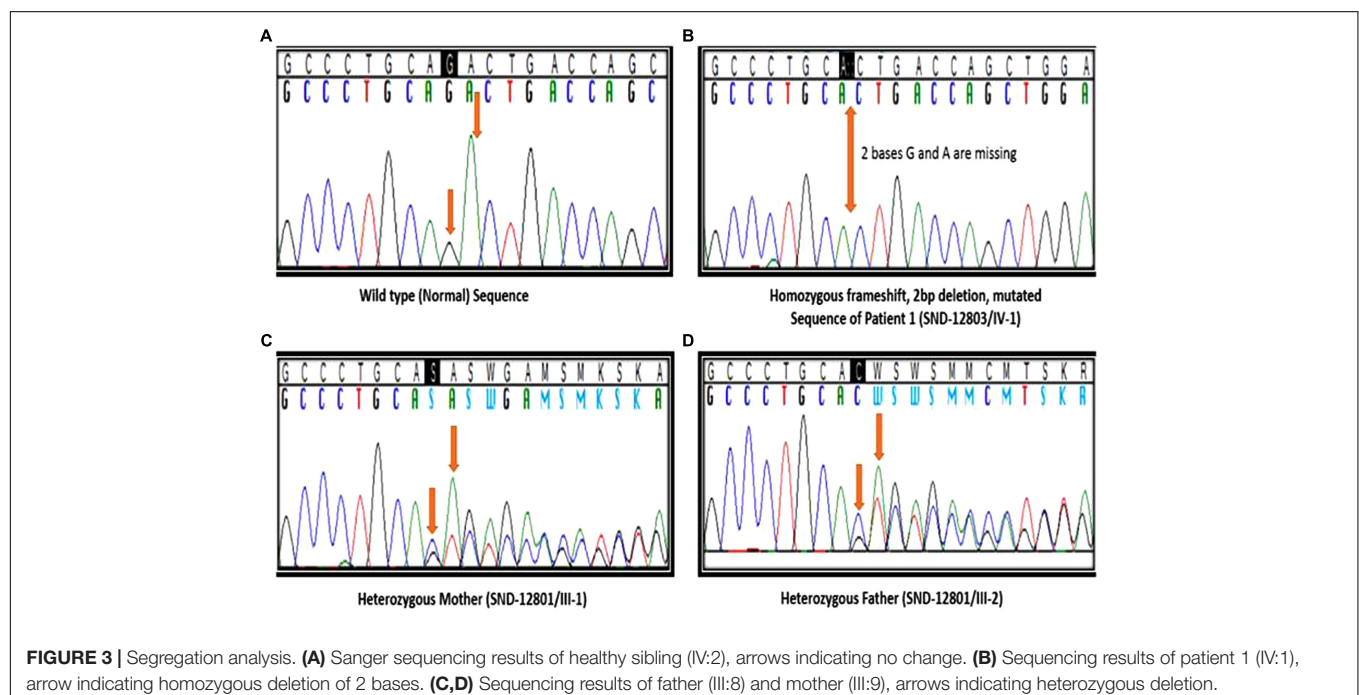


TABLE 1 | All reported cases of SOS with *XYLT2* mutations and phenotypic differences.

	This report						Umair et al., 2017			Taylan et al., 2017			Taylan et al., 2016			Munns et al., 2015		
	Pakistani			Pakistani			Iraqi			Turkish			Turkish			Canadian		
	Patient1	Patient2	Patient3	Patient1	Patient2	Patient3	Family A	Family B		Patient1	Patient2		Patient1	Patient2		Patient1	Patient2	Patient3
Ethnicity	Pakistani	Pakistani	Pakistani	Pakistani	Pakistani	Pakistani	Iraqi	Turkish		Turkish	Turkish	Turkish	Turkish	Canadian	Iraqi	European Australian		
Consanguinity	+	+	+	+	+	+	+	–	–	–	+	+	+	+	+	–	–	+
Mutation	c.2518 – 2519delAG	p.R840fs*115					c.1159G>T p.R387W	c.2548G>C p.D850H	p.M237R	p.E854As*101	c.2188C>T p.R730*	c.1687C>G p.R563G	c.1814T>C p.L605P	c.692dupC p.V232Gfs*54	c.692dupC p.V232Gfs*54	c.520delA p.A174Pfs*35		
Skeletal fragility	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Recurrent fractures	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Decreased spine mobility	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Facial dysmorphism	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Platyspondyly	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Cataract	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ureter dilation	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Hearing loss	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Learning difficulties	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–
Dental problems	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Cardiac issues	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Effect of bisphosphonate therapy	+/-	+/-	+/-	+/-	+/-	+/-	NA	NA	+	+	+	+	+	+	+	+	+	+/-

in consanguineous Canadian and Iraqi families. To date, two homozygous frameshift deletions, one homozygous frameshift insertion, one homozygous nonsense and five homozygous missense variants in *XYLT2* with typical SOS phenotypic characteristics have been reported (Taylan and Makitie, 2016; Taylan et al., 2017; Umair et al., 2017; **Table 1**). Here, we report a novel homozygous nonsense frameshift variant in a consanguineous Pakistani family with three affected individuals with manifestations of SOS including moderate to severe early-childhood onset osteoporosis, multiple compression fractures, gradual hearing loss and visual impairment. Our patients had a frameshift change in exon 11 of *XYLT2*. The p.Ala174Profs*35 and p.Val232Glyfs*54 frameshifts previously reported by Munns et al. (2015) in exon2 and 3 lead to premature termination and truncated mRNA (**Table 1**). None of our patients currently present with severe cardiac problems or congenital heart defects. The described frameshift **p.R840Tfs*115**, is predicted to adversely affect the catalytic subunit of xylosyltransferase II. The described frameshift results in the removal of the last 25 evolutionarily conserved amino acids in the last exon of *XYLT2* (exon11). Defects in exon 11 inhibit the catalysis of sugar transfer to serine residue of PG core proteins thus resulting in improper PGs synthesis. Phenotypic features in our patients are more serious as compared with those reported in subjects with homozygous missense mutations by Taylan et al. (2016). It is possible that the single amino acid replacement described by Taylan et al. (2016) may allow residual protein function while in our patients the removal of 25 amino acids may result in complete loss of *XYLT2* function. All these patients have been treated with pamidronate infusions, which has helped in reducing the fractures but some new compression fractures still occurred. As expected, pamidronate infusion did not improve vision or hearing.

To date, all patients including our patients with SOS manifestation and defect in xylosyltransferase II have showed autosomal recessive pattern of inheritance, and no clinical manifestations have been reported in a heterozygous state. This finding suggests that only biallelic deleterious variants in *XYLT2* lead to such a loss in enzyme function that clinical symptoms appear, while in a heterozygous state the *XYLT2* function is compensated by the normal allele.

CONCLUSION

We identified a novel deleterious nonsense frame shift variant in the *XYLT2* in three children in a consanguineous Pakistani family. This novel variant results in marked skeletal and extra-skeletal features including generalized osteoporosis, platyspondyly, gradual hearing loss, and visual impairment. Differences in

phenotypic presentations, from mild to severe forms, between our and previously reported patients are likely to be dependent on the nature and location of the variant along with the expression pattern of *XYLT2* in different tissues. Pamidronate therapy was beneficial but did not fully restore the bone health.

DATA AVAILABILITY

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

ETHICS STATEMENT

Ethical committee from IBGE provided the ethical approval for our project on the Genetics of Rare Syndromes. Consent was also obtained from the head of the family for the publication of the results.

AUTHOR CONTRIBUTIONS

MK and SS did the family collection. HU and OM did the clinical evaluation. EC, CK, SS, and JF did the NGS data analysis. MK, EC, MA, JF, SS, and OM compiled the data and wrote the manuscript.

FUNDING

This work was funded by Sigrid Jusélius Foundation, Novo Nordisk Foundation (NNF180C0034982), Academy of Finland (318137) and Folkhälsan Research Foundation. IRSIP scholarship by Higher Education Commission (HEC) Pakistan and CIMO's Fellowship (Finland) were awarded to MK. JF is a Singapore National Research Foundation Fellow (NRF-NRFF2016-03).

ACKNOWLEDGMENTS

We are grateful to family members for their invaluable cooperation and participation in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00144/full#supplementary-material>

REFERENCES

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth.0410-248

Alanay, Y., Superti-Furga, A., Karel, F., and Tuncbilek, E. (2006). Spondylo-ocular syndrome: a new entity involving the eye and spine. *Am. J. Med. Genet. A* 140, 652–656. doi: 10.1002/ajmg.a.31119

Aziz, S., Noor-Ul-Ain, W., Majeed, R., Khan, M. A., Qayum, I., Ahmed, I., et al. (2012). Growth centile charts (anthropometric measurement) of Pakistani pediatric population. *J. Pak. Med. Assoc.* 62, 367–377.

- Couchman, J. R., and Pataki, C. A. (2012). An introduction to proteoglycans and their localization. *J. Histochem. Cytochem.* 60, 885–897. doi: 10.1369/0022155412464638
- Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632
- Mizumoto, S., Yamada, S., and Sugahara, K. (2015). Mutations in biosynthetic enzymes for the protein linker region of chondroitin/dermatan/heparan sulfate cause skeletal and skin dysplasias. *Biomed. Res. Int.* 2015:861752. doi: 10.1155/2015/861752
- Munns, C. F., Fahiminiya, S., Poudel, N., Munteanu, M. C., Majewski, J., Sillence, D. O., et al. (2015). Homozygosity for frameshift mutations in XYLT2 result in a spondylo-ocular syndrome with bone fragility, cataracts, and hearing defects. *Am. J. Hum. Genet.* 96, 971–978. doi: 10.1016/j.ajhg.2015.04.017
- Schmidt, H., Rudolph, G., Hergersberg, M., Schneider, K., Moradi, S., and Meitinger, T. (2001). Retinal detachment and cataract, facial dysmorphism, generalized osteoporosis, immobile spine and platyspondyly in a consanguineous kindred—a possible new syndrome. *Clin. Genet.* 59, 99–105. doi: 10.1034/j.1399-0004.2001.590206.x
- Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* 40, W452–W457. doi: 10.1093/nar/gks539
- Taylan, F., Costantini, A., Coles, N., Pekkinen, M., Heon, E., Siklar, Z., et al. (2016). Spondyloocular syndrome: novel mutations in XYLT2 gene and expansion of the phenotypic spectrum. *J. Bone Miner. Res.* 31, 1577–1585. doi: 10.1002/jbmr.2834
- Taylan, F., and Makitie, O. (2016). Abnormal proteoglycan synthesis due to gene defects causes skeletal diseases with overlapping phenotypes. *Horm. Metab. Res.* 48, 745–754. doi: 10.1055/s-0042-118706
- Taylan, F., Yavas Abali, Z., Jantti, N., Gunes, N., Darendeliler, F., Bas, F., et al. (2017). Two novel mutations in XYLT2 cause spondyloocular syndrome. *Am. J. Med. Genet. A* 173, 3195–3200. doi: 10.1002/ajmg.a.38470
- Umair, M., Eckstein, G., Rudolph, G., Strom, T., Graf, E., Hendig, D., et al. (2017). Homozygous XYLT2 variants as a cause of spondyloocular syndrome. *Clin. Genet.* 93, 913–918. doi: 10.1111/cge.13179

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Kausar, Chew, Ullah, Anees, Khor, Foo, Makitie and Siddiqi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Compound Heterozygous *CHAT* Gene Mutations of a Large Deletion and a Missense Variant in a Chinese Patient With Severe Congenital Myasthenic Syndrome With Episodic Apnea

Zhimei Liu^{1†}, Li Zhang^{2,3†}, Danmin Shen¹, Changhong Ding¹, Xinying Yang¹, Weihua Zhang¹, Jiuwei Li¹, Jie Deng¹, Shuai Gong¹, Jun Liu⁴, Suyun Qian⁴ and Fang Fang^{1*}

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University, China

Reviewed by:

Judith Cossins,
University of Oxford, United Kingdom
Xingbin Ai,
Brigham and Women's Hospital,
Harvard Medical School,
United States

*Correspondence:

Fang Fang
13910150389@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 16 November 2018

Accepted: 28 February 2019

Published: 12 March 2019

Citation:

Liu Z, Zhang L, Shen D, Ding C,
Yang X, Zhang W, Li J, Deng J,
Gong S, Liu J, Qian S and Fang F
(2019) Compound Heterozygous
CHAT Gene Mutations of a Large
Deletion and a Missense Variant in a
Chinese Patient With Severe
Congenital Myasthenic Syndrome
With Episodic Apnea.
Front. Pharmacol. 10:259.
doi: 10.3389/fphar.2019.00259

¹ Department of Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China, ² Center for Bioinformatics and Computational Biology, Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, ³ School of Statistics, Faculty of Economics and Management, East China Normal University, Shanghai, China, ⁴ Department of Pediatric Intensive Care Unit, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

Congenital myasthenic syndromes (CMSs) are a group of inherited disorders caused by genetic defects in neuromuscular junctions. Mutations in *CHAT*, encoding choline acetyltransferase, cause congenital myasthenic syndrome with episodic apnea (CMS-EA), a rare autosomal recessive disease characterized by respiratory insufficiency with cyanosis and apnea after infections, fever, vomiting, or excitement. To date, no studies have reported deletions comprised of multiple exons. Here, using next generation sequencing, we identified compound heterozygous mutations, namely a large maternally inherited deletion, including exons 4, 5, and 6, and a paternally inherited missense variant (c.914T>C [p.Ile305Thr]) in *CHAT* in a Chinese patient with a severe phenotype of CMS-EA. Furthermore, the large deletion was also validated by real-time fluorescence quantitative polymerase chain reaction. The patient was a 10-month-old boy, who presented with a weak cry and feeding difficulties soon after birth, ptosis at 4 months old, episodic apnea after fever at 9 months old, and respiratory insufficiency with cyanosis and apnea that required intubation after a respiratory tract infection at 10 months old. Unfortunately, he died in the Pediatric Intensive Care Unit soon after hospitalization. The patient's elder sister had similar clinical manifestations, and she died prior to the age of 2 months old without a diagnosis. Genotype-phenotype correlation analysis revealed that loss-of-function mutations in exons 4–6 of *CHAT* might cause more severe CMS-EA. To our knowledge, this is the first study to show compound heterozygous *CHAT* mutations consisting of a large deletion and missense mutation in a patient with CMS-EA.

Keywords: *CHAT*, congenital myasthenic syndromes, episodic apnea, large deletion, severe

INTRODUCTION

Congenital myasthenic syndromes (CMSs) are a group of inherited disorders caused by genetic defects in neuromuscular junctions. CMSs have been recognized as clinical entities since the 1970s (Conomy et al., 1975; Engel et al., 1977) and are classified into the presynaptic, synaptic, or postsynaptic syndrome type according to the involved mutation sites or genes. With the development of next generation sequencing (NGS) technology, over 30 CMS disease-related genes have now been reported, including *CHAT*, *CHRNE*, *COLQ*, *RAPSN*, and so on; of these, *CHAT* accounts for 4–5% (Abicht et al., 1993–2019; Engel, 2018; Rodriguez et al., 2018).

The *CHAT* gene, located on chromosome 10q11.23, encodes choline acetyltransferase (ChAT), which catalyzes the synthesis of the neurotransmitter acetylcholine from acetyl coenzyme A (AcCoA) and choline. In Ohno et al. (2001), *CHAT* mutations were first reported to cause congenital myasthenic syndrome with episodic apnea (CMS-EA), also named familial infantile myasthenia. Usually, CMS-EA manifests at birth or in early infancy with hypotonia, variable eyelid ptosis, severe bulbar weakness causing dysphagia, and respiratory insufficiency with cyanosis and apnea; the crises recur with infections, fever, excitement, vomiting, or overexertion, and can be prevented or mitigated by anticholinesterase drugs (Ohno et al., 2001). To date, more than 40 *CHAT* mutations have been identified to cause CMS-EA (Human Gene Mutation Database [HGMD®] Professional version 2018.1). Although some genetic heterogeneity regarding catalytic activity and phenotypic heterogeneity regarding onset, severity of crises, and prognosis have been described, no genotype-phenotype correlation has been identified. Here, we present the case of a 10-month-old Chinese boy with compound heterozygous *CHAT* variants, including a large deletion (exons 4, 5, and 6) and a missense variant c.914T>C (p.Ile305Thr), which manifested as severe CMS-EA.

MATERIALS AND METHODS

Ethics Statement

The present study was approved by the Ethics Committee of Beijing Children's Hospital, Capital Medical University, Beijing, China, and was conducted according to the principles expressed in the Declaration of Helsinki. Participants and/or their legal guardians involved in this study gave a written informed consent prior to inclusion in the study. Participants and/or their legal guardians also provided their written informed consent for the material to appear in *Frontiers in Pharmacology* and associated publications without limit on the duration of publication.

Sample Collection and Library Preparation

The present study included DNA samples from three family members, the parents and the proband. Genomic DNA was isolated using a blood DNA extraction kit according to the manufacturer's recommendations (Beijing ComWin Biotech

Co., Ltd., Beijing, China). A minimum of 3 µg DNA was used to make the indexed Illumina libraries according to the manufacturer's protocol. The 300–400 bp library size including adapter sequences was finally selected.

Targeted NGS

Targeted sequencing was performed on the whole mitochondrial genome and 1,033 nuclear genes (**Supplementary Table S1**), that affect mitochondrial structure and function, or cause some disease difficult to differentiate from mitochondrial disease, such as Krabbe disease, succinic semialdehyde dehydrogenase deficiency, CMS-EA, and so on (Fang et al., 2017).

Sanger Sequencing

The variant prioritized through NGS was verified by Sanger sequencing in the patient and his parents. The primer sequences used were as follows: F: 5'-GCCGAGAG AAGATCAGCATAAGCA-3', and R: 5'-GTACAGGTGGAGGT CTCGATCA-3'.

Reads Mapping and Variant Calling

Paired-end reads of 200 bp (100 bp at each end) from the targeted sequencing were mapped to UCSC human reference genome (GRCh37/hg19) using Burrows–Wheeler Aligner (Li and Durbin, 2010) “mem” mode with default options, followed by removal of polymerase chain reaction (PCR) duplicates and low-quality reads (BaseQ <20). The binary alignment map files were then sorted, indexed, and converted into the mpileup format by SAMtools (Li et al., 2009). Variant calling was implemented in VarScan (Koboldt et al., 2012) software¹ using the mpileup2snp and mpileup2indel modules.

Variant Annotation and Prioritization

The identified variants were annotated by ANNOVAR (Wang et al., 2010). The annotation information included minor allele frequency (MAF) in the Genome Aggregation Database (gnomAD) (Lek et al., 2016), variant pathogenicity scores by SIFT (Ng and Henikoff, 2003), PolyPhen2 (Adzhubei et al., 2013), MutationTaster (Schwarz et al., 2010), M-CAP (Jagadeesh et al., 2016), RefSeq gene and the consequences on protein, such as missense, frameshift, in-frameshift, stop-gain, and splicing. Rare variants (MAF < 0.01%) were filtered based on gnomAD (Lek et al., 2016).

Identification and Quantitative PCR Validation of *CHAT* Deletion

The *CHAT* deletion was firstly identified by targeted sequencing data as the loss of heterogeneity in the proband. The read depth for each site (base) within the exons of *CHAT* gene was calculated. The average read depth of each exon was then calculated by averaging the read depth for the sites within the exon (**Figure 3B**). Using *ALB* gene as the internal control, copy numbers of the 4th, 5th, and 6th exons in the *CHAT* gene were estimated by real-time fluorescence quantitative PCR (qPCR) in the patient and his parents.

¹<http://varscan.sourceforge.net/>

RESULTS

Clinical Features of the Patient

The proband, the second child of two healthy non-consanguineous parents, was a 10-month-old boy. He was born through cesarean section due to a scarred uterus after a full-term pregnancy, with a birth weight of 3.2 kg. Immediately after birth, the boy presented with a weak cry and feeding difficulties, such as slow swallowing, choking easily, and breathing difficulties with apnea if the feeding posture changed, especially when lying down. When he was 4 months old, eyelid ptosis developed. At the age of 9 months, episodic apnea occurred after fever, but improved through symptomatic treatment. The proband's development milestones were normal.

When the proband was 10 months old, respiratory insufficiency with cyanosis and apnea occurred after a respiratory tract infection, which required mechanical ventilation. A second apneic episode requiring intubation occurred soon thereafter, but the parents refused to permit endotracheal intubation, and thus he received Nasal Continuous Positive Airway Pressure respiratory support. However, repeated episodic apnea continued to occur, and eventually when intubation was deemed necessary. Intravenous immunoglobulin was administered, but was ineffective. The patient's respiratory function decreased gradually, and he died soon after stopping treatment. Overall, the length of hospital stay was 11 days.

Upon admission to our hospital, the proband's consciousness was clear. Fluctuating eyelid ptosis was observed, which was aggravated by fatigue. Eye movement in all directions was normal. Limb muscle strength and muscle tone decreased, while the tendon reflex was positive. Meningeal irritation signs and pathological signs were negative.

The neostigmine test was negative, and no anti-acetylcholine receptor or anti-muscle-specific kinase antibodies were detected in the serum. Biochemical examinations, including evaluations of the serum creatine kinase concentration, and serum and cerebrospinal fluid lactate levels, were normal. There were no specific changes in urinary organic acids or blood, as assessed by tandem mass spectrometry analyses. Brain magnetic resonance imaging (MRI) revealed deep sulci in the frontal and parietal lobes and a wide subarachnoid space (Figures 1A,B). Electroencephalography and echocardiography recordings were normal. Pulmonary computed tomography showed slight inflammation, and bronchoscopy did not reveal any abnormalities.

Family history investigation revealed that the proband's elder sister had similar manifestations and symptoms (Figure 1C), presenting with breathing and feeding difficulties soon after birth. She was hospitalized in the local hospital for nearly 50 days, without a diagnosis, and died of apnea after choking on milk.

Targeted Sequencing Analysis of the Proband

The proband was suspected mitochondrial disease previously, and targeted NGS (Jia and Shi, 2017; Ni and Shi, 2017) was

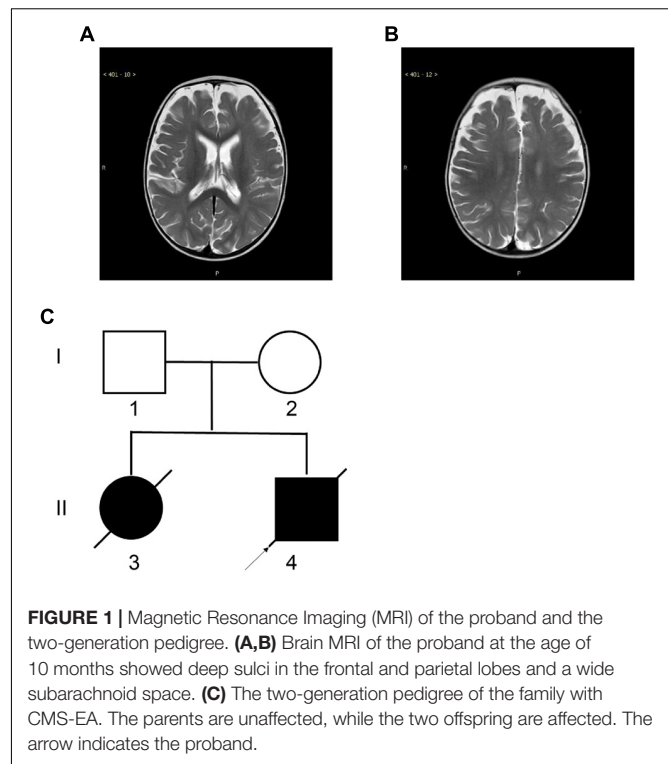


FIGURE 1 | Magnetic Resonance Imaging (MRI) of the proband and the two-generation pedigree. **(A,B)** Brain MRI of the proband at the age of 10 months showed deep sulci in the frontal and parietal lobes and a wide subarachnoid space. **(C)** The two-generation pedigree of the family with CMS-EA. The parents are unaffected, while the two offspring are affected. The arrow indicates the proband.

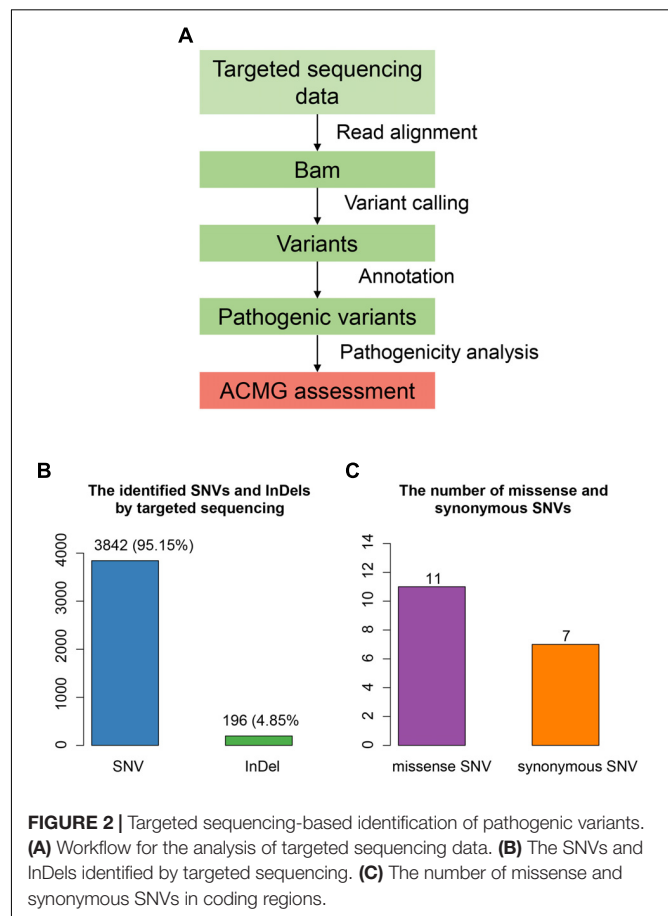


FIGURE 2 | Targeted sequencing-based identification of pathogenic variants. **(A)** Workflow for the analysis of targeted sequencing data. **(B)** The SNVs and InDels identified by targeted sequencing. **(C)** The number of missense and synonymous SNVs in coding regions.

performed on the proband (**Figure 2A**). In total, we identified 4,038 variants (**Figure 2B**), including 3,842 single nucleotide variants (SNVs), and 196 small insertions or deletions (InDels). We then annotated these variants using ANNOVAR, and found 979 rare variants (gnomAD MAF < 0.01%). After excluding

variants within non-coding regions, we identified only 11 missense variants (**Figure 2C**).

To evaluate the pathogenicity of these rare missense variants, we performed a systematic pathogenicity analysis using methods described in a previous study (Jin et al., 2018; Yu, 2018).

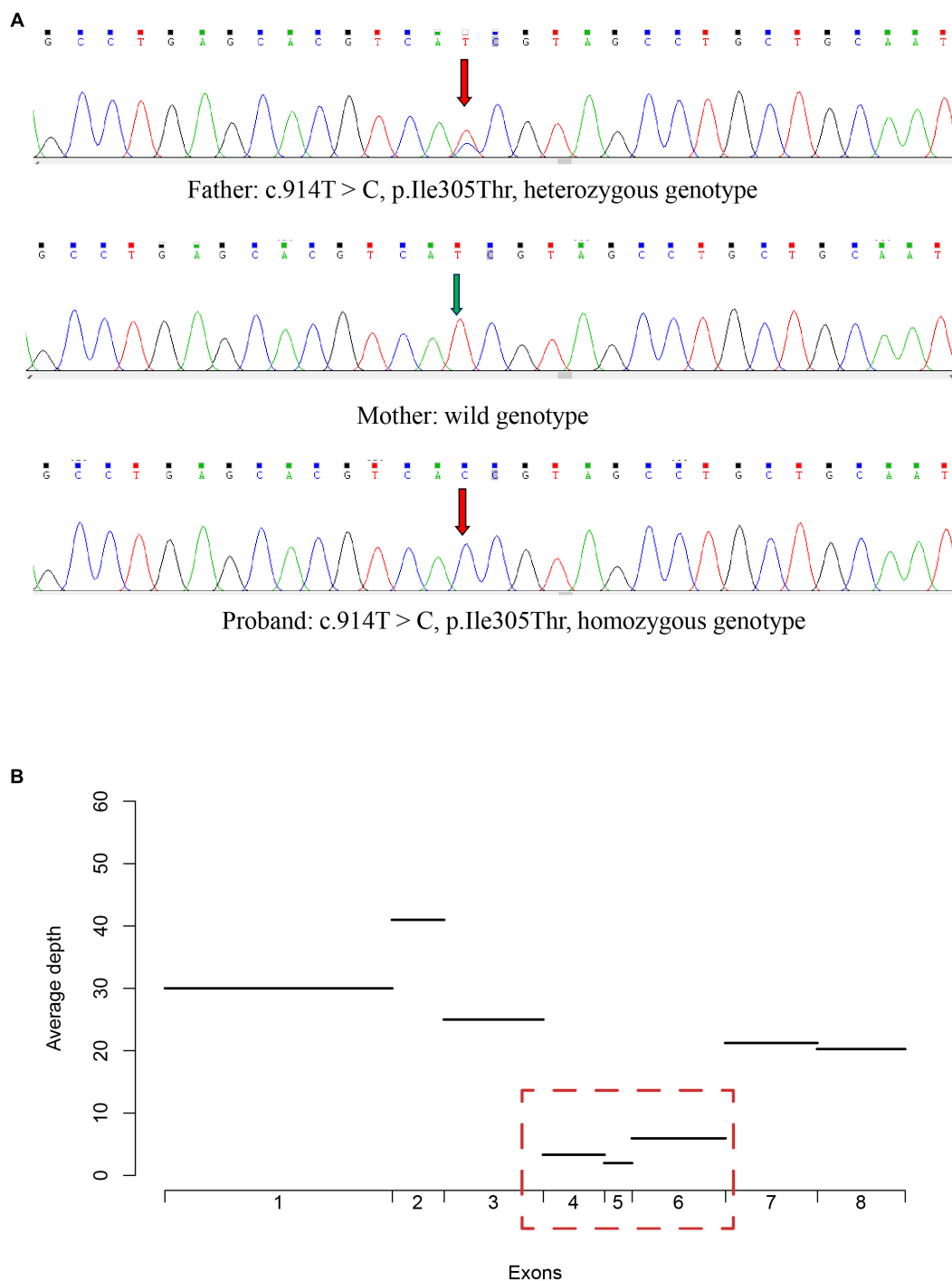
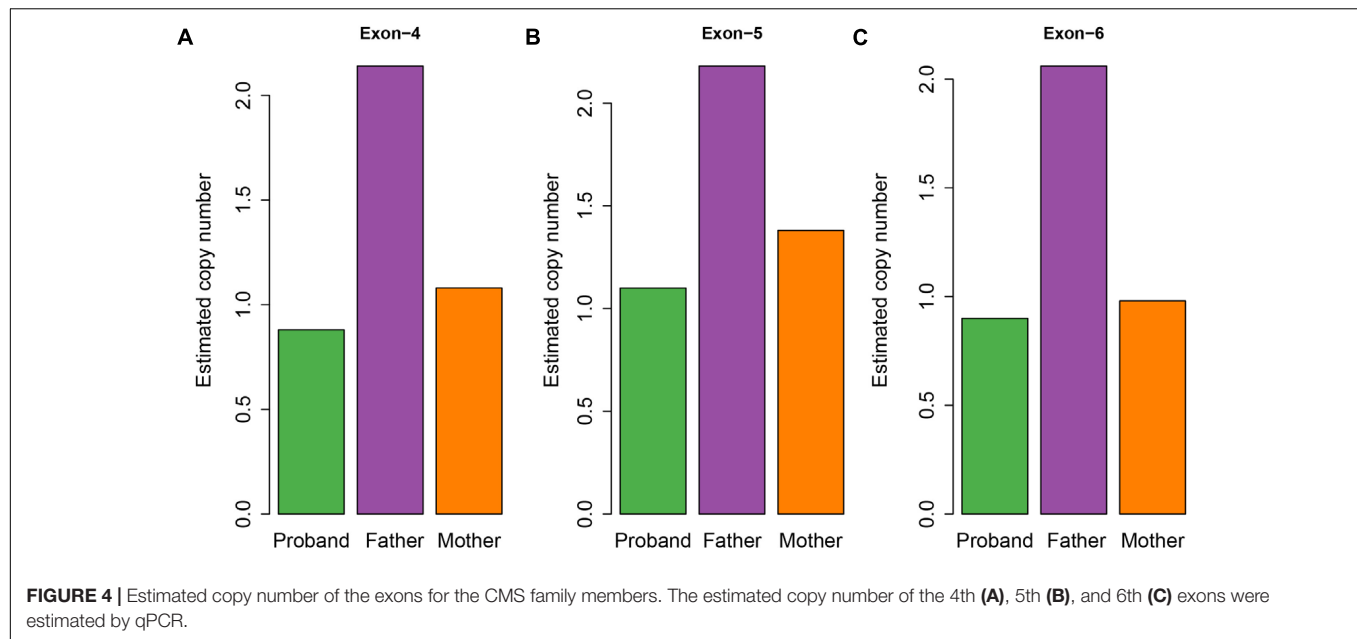


FIGURE 3 | Validation of the missense variant and the large deletion in *CHAT*. **(A)** The genotype of the missense variant in three members of the CMS-EA family, as identified by Sanger sequencing. **(B)** The average read depth for the first 8 exons for the proband was calculated with the targeted sequencing data.



For the missense variants, we identified a missense variant within *CHAT* (c.914T>C [p.Ile305Thr]) as the disease-causing variant (SIFT ≤ 0.05 , PolyPhen ≥ 0.957 , MutationTaster = disease causing, and M-CAP > 0.025), which had a homozygous genotype in the proband.

Validation of the Pathogenic Variants in *CHAT*

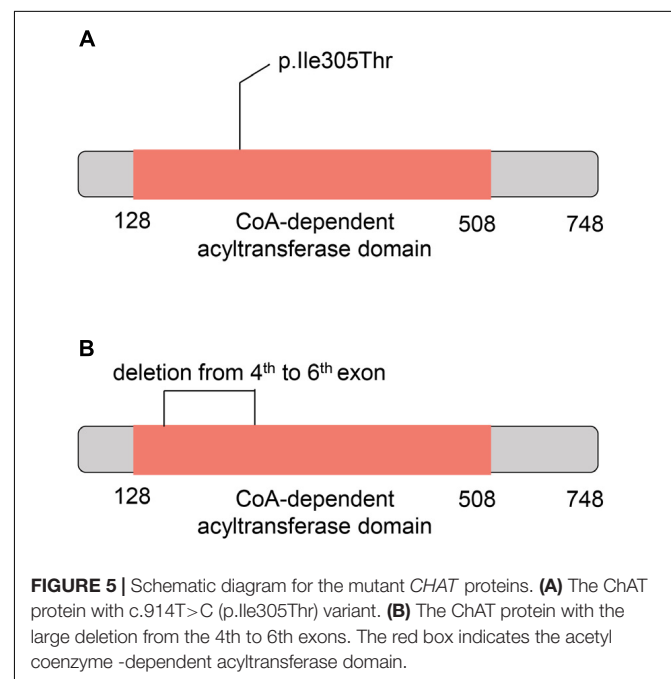
To validate the pathogenic variants, we performed Sanger sequencing on each of the family members. The missense variant in *CHAT* (c.914T>C [p.Ile305Thr]) was validated in the proband, who had a homozygous genotype, and in his father, who had a heterozygous genotype (Figure 3A). However, this variant was absent in the proband's mother, suggesting that loss of heterozygosity (LOH) may have led to the variant appearing homozygous in the proband.

To examine whether a large deletion was located within *CHAT*, we calculated the read depth for each exon of *CHAT*. As expected, the read depth was significantly reduced from the 4th to 6th exon of the transcript with RefSeq accession number NM_020549 (Figure 3B), indicating that a large deletion was located within these exons. To further validate the large deletion, we performed qPCR to estimate the copy number for exons 4, 5, and 6. In accordance with the read depth analysis, loss of heterozygosity was also observed within regions of the 4th to 6th exon in both the proband and his mother (Figure 4). Hence, the large deletion combined with the missense variant (c.914T>C [p.Ile305Thr]) led to the occurrence of CMS-EA in the proband.

Functional Characterization of the *CHAT* Variants

To further understand the role of the two *CHAT* variants in CMS, we investigated their potential consequences

on the ChAT protein. The missense variant in *CHAT* (c.914T>C [p.Ile305Thr]) was located within the CoA-dependent acyltransferase domain (the amino acid sites from 128 to 508) based on the SuperFamily (Pandurangan et al., 2019) annotation (Figure 5A). Moreover, the large deletion located within the 4th to 6th exons also overlapped with the CoA-dependent acyltransferase domain (Figure 5B). As reported in a previous study, the missense mutation markedly reduced ChAT expression in COS cells and had significantly impaired catalytic efficiencies



in kinetic studies (Ohno et al., 2001). These results indicated that the CoA-dependent acyltransferase domain might have been disrupted by the two variants in the patient of this study.

DISCUSSION

Since the first report in Ohno et al. (2001) showing that *CHAT* mutations cause CMS-EA, more than 40 mutations have been found to cause CMS-EA; of these, bi-allelic point mutations are the most frequent (HGMD® Professional version 2018.1). To our knowledge, the present study is the first study to report the case of a patient with CMS-EA caused by compound heterozygous exons deletion and missense mutation in the *CHAT* gene. Our functional characterization results indicated that the CoA-dependent acyltransferase domain might have been disrupted by the two variants in this study. The missense variant has previously been demonstrated to markedly reduce ChAT expression in COS cells and significantly impair catalytic efficiencies in kinetic studies (Ohno et al., 2001). In accordance with this previous report, the severe phenotype of the patient in our study may have been caused by the missense variant combined with the large deletion.

As mentioned earlier, CMS-EA usually presents at birth or in early infancy with hypotonia, ptosis, dysphagia due to severe bulbar weakness, and respiratory insufficiency with cyanosis and apnea. Occasionally, apnea crises may be mistaken for seizures and thus anticonvulsive drugs may be initiated, without positive effects (Schara et al., 2010); as such, EEG is essential for distinguishing apneic crises from seizures, especially video EEG. Several reported patients presented with mental retardation, and MRI showed brain atrophy. There are two possible explanations for this. First, the apnea may have led to brain hypoxemia. Second, mental retardation may be a symptom of ChAT deficiency in the brain (Schara et al., 2010). Indeed, ChAT deficiency has been reported in various developmental and neurodegenerative disorders, including Alzheimer's disease, Huntington's disease, amyotrophic lateral sclerosis, Schizophrenia, Rett syndrome, and Sudden Infant Death Syndrome (SIDS) (Oda, 1999). The cause of brain atrophy is undetermined, but normal development milestone in the present study support more that may be induced by brain hypoxemia.

Electrophysiological assessments are important for diagnosing CMS-EA, as repetitive nerve stimulation testing at a low frequency (10 stimuli at 2–3 Hz) might be normal, but prolonged subtetanic stimulation (10 Hz for 5 min) decreased the amplitude of the compound muscle action potential (CMAP) and endplate potential to 50% below baseline (normal decrease is <30%), followed by slow recovery. This slow recovery suggests the presence of a defect in acetylcholine resynthesis and previously aided in the discovery of mutations in *CHAT* (Ohno et al., 2001; Dilella et al., 2014; Engel et al., 2015). Prolonged subtetanic nerve stimulation is rarely used in infants, and thus performing genetic

testing as soon as possible is helpful for ensuring early diagnosis and treatment. Because of the severe condition of our patient, electrophysiological assessments were not conducted.

In general, CMS-EA is a treatable rare disease that responds positively to acetylcholinesterase (AChE) inhibitors, and thus early treatment with pyridostigmine is helpful for improving the clinical symptoms and prognosis (Engel, 2018). Although it was reported that midazolam may mitigate the severity of apnea episodes (Mallory et al., 2009; Barisic et al., 2011), further studies are needed. Unfortunately, the symptoms in our patient were so severe that he died before he could receive AChE inhibitor therapy. We think that the findings of this study will help clinicians identify such cases early, perhaps avoiding a negative outcome.

In addition, there are still some limitations in the present study. First, the association between *CHAT* large deletion and poor prognosis requires more cases to support. Moreover, the breakpoint of *CHAT* deletion was not accurately determined by Sanger sequencing due to lack of more blood sample. To conclude, we first reported compound heterozygous *CHAT* mutations consisting of a large deletion (exons 4, 5, and 6) and missense mutation (c.914T>C [p.Ile305Thr]) in a patient with severe CMS-EA.

DATA AVAILABILITY

All the clinical data and identified genetic variations have been deposited into the rare disease database, eRAM (Jia et al., 2018), at <http://www.pediascape.org/eram/>.

AUTHOR CONTRIBUTIONS

FF, CD, SQ designed the study. DS, XY, WZ, JwL, JD, SG, and JL collected the data and performed the research, ZL and LZ analyzed the data and wrote the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

This work was supported by Beijing Municipal Science and Technology Plan Projects (Z161100002616004).

ACKNOWLEDGMENTS

We would like to thank Editage (www.editage.com) for English language editing. We also acknowledge the financial support of the Open Access Publication Fund of Beijing Children's Hospital.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00259/full#supplementary-material>

TABLE S1 | Nuclear gene list of targeted NGS used in the present study.

REFERENCES

- Abicht, A., Müller, J. S., and Lochmüller, H. (1993–2019). “Congenital myasthenic syndromes,” in *Source® GeneReviews*, eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, et al. (Seattle, WA: University of Washington).
- Adzhubei, I., Jordan, D. M., and Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* 76, 7.20.1–7.20.41. doi: 10.1002/0471142905.hg0270s76
- Barisic, N., Chaouch, A., Muller, J. S., and Lochmüller, H. (2011). Genetic heterogeneity and pathophysiological mechanisms in congenital myasthenic syndromes. *Eur. J. Paediatr. Neurol.* 15, 189–196. doi: 10.1016/j.ejpn.2011.03.006
- Conomy, J. P., Levinsohn, M., and Fanaroff, A. (1975). Familial infantile myasthenia gravis: a cause of sudden death in young children. *J. Pediatr.* 87, 428–430. doi: 10.1016/S0022-3476(75)80651-2
- Dilena, R., Abicht, A., Sergi, P., Comi, G. P., Di Fonzo, A., Chidini, G., et al. (2014). Congenital myasthenic syndrome due to choline acetyltransferase mutations in infants: clinical suspicion and comprehensive electrophysiological assessment are important for early diagnosis. *J. Child Neurol.* 29, 389–393. doi: 10.1177/0883073812470000
- Engel, A. G. (2018). Genetic basis and phenotypic features of congenital myasthenic syndromes. *Handb. Clin. Neurol.* 148, 565–589. doi: 10.1016/B978-0-444-64076-5.00037-5
- Engel, A. G., Lambert, E. H., and Gomez, M. R. (1977). A new myasthenic syndrome with end-plate acetylcholinesterase deficiency, small nerve terminals, and reduced acetylcholine release. *Ann. Neurol.* 1, 315–330. doi: 10.1002/ana.410010403
- Engel, A. G., Shen, X. M., Sclen, D., and Sine, S. M. (2015). Congenital myasthenic syndromes: pathogenesis, diagnosis, and treatment. *Lancet Neurol.* 14, 420–434. doi: 10.1016/S1474-4422(14)70201-7
- Fang, F., Liu, Z., Fang, H., Wu, J., Shen, D., Sun, S., et al. (2017). The clinical and genetic characteristics in children with mitochondrial disease in China. *Sci. China Life Sci.* 60, 746–757. doi: 10.1007/s11427-017-9080-y
- Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., et al. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* 48, 1581–1586. doi: 10.1038/ng.3703
- Jia, J., An, Z., Ming, Y., Guo, Y., Li, W., Liang, Y., et al. (2018). eRAM: encyclopedia of rare disease annotations for precision medicine. *Nucleic Acids Res.* 46, D937–D943. doi: 10.1093/nar/gkx1062
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Jin, Y., Zhang, L., Ning, B., Hong, H., Xiao, W., Tong, W., et al. (2018). Application of genome analysis strategies in the clinical testing for pediatric diseases. *Pediatr. Investig.* 2, 72–81. doi: 10.1002/ped4.12044
- Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., et al. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* 22, 568–576. doi: 10.1101/gr.129684.111
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. doi: 10.1038/nature19057
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Mallory, L. A., Shaw, J. G., Burgess, S. L., Estrella, E., Nurko, S., Burpee, T. M., et al. (2009). Congenital myasthenic syndrome with episodic apnea. *Pediatr. Neurol.* 41, 42–45. doi: 10.1016/j.pediatrneurol.2009.02.017
- Ng, P. C., and Henikoff, S. (2003). SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814. doi: 10.1093/nar/gkg509
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60, 681–685. doi: 10.1007/s11427-017-9100-1
- Oda, Y. (1999). Choline acetyltransferase: the structure, distribution and pathologic changes in the central nervous system. *Pathol. Int.* 49, 921–937. doi: 10.1046/j.1440-1827.1999.00977.x
- Ohno, K., Tsujino, A., Brengman, J. M., Harper, C. M., Bajzer, Z., Udd, B., et al. (2001). Choline acetyltransferase mutations cause myasthenic syndrome associated with episodic apnea in humans. *Proc. Natl. Acad. Sci. U.S.A.* 98, 2017–2022. doi: 10.1073/pnas.98.4.2017
- Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B., and Gough, J. (2019). The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* 47, D490–D494. doi: 10.1093/nar/gky1130
- Rodriguez, C. P., Palace, J., and Beeson, D. (2018). The neuromuscular junction and wide heterogeneity of congenital myasthenic syndromes. *Int. J. Mol. Sci.* 19:E1677. doi: 10.3390/ijms19061677
- Schara, U., Christen, H. J., Durmus, H., Hietala, M., Krabetz, K., Rodolico, C., et al. (2010). Long-term follow-up in patients with congenital myasthenic syndrome due to CHAT mutations. *Eur. J. Paediatr. Neurol.* 14, 326–333. doi: 10.1016/j.ejpn.2009.09.009
- Schwarz, J. M., Rodelsperger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. doi: 10.1038/nmeth0810-575
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38:e164. doi: 10.1093/nar/gkq603
- Yu, D. (2018). Next-generation sequencing applied in pediatric diseases. *Pediatr. Investig.* 2, 82–82. doi: 10.1002/ped4.12045

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Liu, Zhang, Shen, Ding, Yang, Zhang, Li, Deng, Gong, Liu, Qian and Fang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Growth Pattern in Chinese Children With 5 α -Reductase Type 2 Deficiency: A Retrospective Multicenter Study

Xiu Zhao^{1,2†}, Yanning Song¹, Shaoke Chen^{3†}, Xiumin Wang⁴, Feihong Luo⁵, Yu Yang⁶, Linqi Chen⁷, Ruimin Chen⁸, Hui Chen⁹, Zhe Su², Di Wu¹ and Chunxiu Gong^{1*}

¹ Center of Endocrinology, Genetics and Metabolism, National Center for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China, ² Department of Endocrinology, Shenzhen Children's Hospital, Shenzhen, China, ³ Genetic and Metabolic Central Laboratory, Maternal and Children Health Hospital of Guangxi Zhuang Autonomous Region, Nanning, China, ⁴ Department of Endocrinology, Shanghai Children's Medical Center, Shanghai Jiao Tong University, Shanghai, China, ⁵ Department of Endocrinology, Children's Hospital of Fudan University, Fudan University, Shanghai, China, ⁶ Department of Endocrinology, Jiangxi Provincial Children's Hospital, Nanchang, China, ⁷ Department of Endocrinology, Children's Hospital of Soochow University, Suzhou, China, ⁸ Department of Endocrinology, Fuzhou Children's Hospital, Fuzhou, China, ⁹ Department of BME, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University, China

Reviewed by:

Michaël R. Laurent,
University Hospitals Leuven, Belgium
Chao Xu,
Shandong Qianfoshan Hospital, China
Michel Polak,
Necker-Enfants Malades Hospital,
France

*Correspondence:

Chunxiu Gong
chunxiugong@sina.com

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 13 July 2018

Accepted: 11 February 2019

Published: 15 March 2019

Citation:

Zhao X, Song Y, Chen S, Wang X,
Luo F, Yang Y, Chen L, Chen R,
Chen H, Su Z, Wu D and Gong C
(2019) Growth Pattern in Chinese
Children With 5 α -Reductase Type 2
Deficiency: A Retrospective
Multicenter Study.
Front. Pharmacol. 10:173.
doi: 10.3389/fphar.2019.00173

Background: 5 α -reductase type 2 deficiency (5 α RD) is an autosomal recessive hereditary disease of the group of 46, XY disorders of sex development (DSD).

Objective: To study the growth pattern in Chinese pediatric patients with 5 α RD.

Subjects: Data were obtained from 141 patients with 5 α RD (age: 0–16 years old) who visited eight pediatric endocrine centers from January 2010 to December 2017.

Methods: In this retrospective cohort study, height, weight, and other relevant data were collected from the multicenter hospital registration database. Baseline luteinizing hormone (LH), follicle stimulating hormone (FSH), testosterone (T), and dihydrotestosterone (DHT) after human chorionic gonadotropin (HCG) stimulation test were measured by enzyme enhanced chemiluminescence assay. Bone age (BA) was assessed using the Greulich-Pyle (G-P) atlas. Growth curve was constructed based on λ -median-coefficient of variation method (LMS).

Results: The height standard deviation scores (HtSDS) and weight standard deviation scores (WtSDS) in 5 α RD children were in the normal range as compared to normal boys. Significantly higher HtSDS was observed in patients with 5 α RD who were <1 year old ($t = 3.658$, 2.103 , $P = 0.002$, 0.048 , respectively), and higher WtSDS in those <6 months old ($t = 2.756$, $P = 0.012$). Then HtSDS and WtSDS decreased gradually and fluctuated near the median of the same age until 13 years. WtSDS in 5 α RD children from northern China were significantly higher than those from the south ($Z = -2.670$, $P = 0.008$). The variation tendency of HtSDS in Chinese 5 α RDs was consistent with the trend of stimulating T. HtSDS and stimulating T in the external masculinization score (EMS) <7 group were slightly higher than those in EMS ≥ 7 group without significant difference. Additionally, the ratio of BA over chronological age (BA/CA) was significantly <1 in children with 5 α RD.

Conclusion: Children with 5 α RD had a special growth pattern that was affected by high levels of T, while DHT played a very small role in it. Their growth accelerated at age <1 year, followed by slowing growth and fluctuating height near normal median boys' height. The BA was delayed in 5 α RD children. Androgen treatment, which may be considered anyway for male 5 α RD patients with a micropenis, may also be beneficial for growth.

HIGHLIGHTS

- The first growth curve in children with 5 α RD.
- Our study provided evidence for applying androgen products at specific time periods for 5 α RD patients with a micropenis.

Keywords: dihydrotestosterone, testosterone, 5 α -reductase type 2 deficiency, 46, XY DSD, height, children

INTRODUCTION

Sex hormones are synthesized in the gonads and adrenal glands (Shen, 2016) and have bioactivity within bone and other target tissues (Vanderschueren et al., 2014). Their biological effects on bone are mediated by different cell types and mechanisms (Almeida et al., 2017) and controlled by gonadotropins via hypothalamic-pituitary feedback. Before puberty, longitudinal bone growth shows no significant sex differences (Nishiyama et al., 2012). During puberty, estrogen shows the biphasic regulation of longitudinal bone growth and epiphyseal closure. Early in puberty, estrogen at low concentrations can stimulate longitudinal growth via indirect effects on GH and insulin-like growth factor I (IGF-I), both of which stimulate growth plate chondrocytes (Veldhuis et al., 2005b, 2011; Almeida et al., 2017). However, in late puberty, higher levels of estrogen can exert inhibitory effects on the growth plate via estrogen receptors in the chondrocytes. Androgens mainly have direct effects on GH and show the influence on circulating IGF-I via peripheral and central aromatization (Veldhuis et al., 2005a, 2009). Whether androgen receptor in chondrocytes contributes to sex differences in longitudinal growth remains unclear. Thus, sex hormones have an essential role for male and female growth.

Disorders of sex development are defined as congenital conditions associated with atypical development of gonadal, chromosomal, or anatomical sex, such as androgen insensitivity syndrome (AIS) (Wang et al., 2017) and 5 α RD. With the rapid development of next-generation sequencing and the popularity of precision medicine, DSD as one of rare diseases can be diagnosis earlier and more accurately (Jia and Shi, 2017; Ni and Shi, 2017).

Abbreviations: ACMG, American College of Medical Genetics and Genomics; BA, bone age; BL, birth length; BMI, body mass index; BWt, birth weight; CA, chronological age; CAIS, complete androgen insensitivity syndrome; CIs, confidence intervals; DHT, dihydrotestosterone; DSD, disorders of sex development; EMS, external masculinization score; 5 α RD, 5 α -reductase type 2 deficiency; FSH, follicle stimulating hormone; GATK, genome analysis toolkit; GH, growth hormone; G-P, Greulich-Pyle; HCG, human chorionic gonadotropin; HtSDS, height standard deviation score; IGF-I, insulin-like growth factor-I; L, coefficient of skewness; LH, luteinizing hormone; LMS, λ -median-coefficient of variation method; M, median; S, coefficient of variation; SD, standard deviation; SNVs, single nucleotide variations; T, testosterone; THtSDS, target height standard deviation score; WtSDS, weight standard deviation score.

Furthermore, different DSDs may have different characteristics and different growth patterns partly owing to different changes in sex hormones. For example, gonadal dysplasia impacts physical development throughout the prenatal period until adulthood (Hughes et al., 2002; Richter-Unruh et al., 2004). In addition, height in children with CAIS who had their gonads removed in the pre-pubertal stage has shown to be lower than those who had it done after the puberty (Han et al., 2008), thus addressing the effect of sex hormones on pre-pubertal height. In our previous study, we found that children with 46, XY DSD are shorter than the normal population (Di et al., 2013). In addition, DSD children with T <100 ng/dL after HCG test are shorter than those with T \geq 100 ng/dL (Wu et al., 2017). However, published literature on DSD growth is scarce.

5 α -reductase type 2 deficiency (OMIM 264600) is an autosomal recessive hereditary disease with an incidence of 11.2–15.5% among patients with 46, XY DSD (Veiga-Junior et al., 2012; Ittiwut et al., 2017) and is caused by loss-of-function mutations of the SRD5A2 gene on chromosome 2. The mutations make the enzyme defective and impair T to DHT conversion (Sultan et al., 2001); therefore, individuals with 5 α RD may develop malformation of external genitalia, including pseudovaginalis, ambiguous genitalia, hypospadias, micropenis, and cryptorchid or, in some cases, even a normal phenotype (Choi et al., 2008). Most of the patients with 5 α RD are born as females, and more than 50% undergo gender self-reassignment during puberty (Wilson et al., 1993; Kolesinska et al., 2014; Bertelloni et al., 2016; Deeb et al., 2016). While many research studies have focused on gender reassignment (Deeb et al., 2016; Byers et al., 2017; Raveenthiran, 2017), our study investigates the growth and development of 5 α RD patients. With the help of growth pattern, doctors can get more information for diagnosis and antidiastole of 5 α RD. Perhaps, this will also help in accurate genetic analysis. Thus far, only two studies have reported on a 5 α RD growth pattern in children. Ko et al. (2010) examined six Korean children with 5 α RD and found a height percentile of P95, P90, and P90th in three cases that did not receive hormone treatment or gonadectomy. Eren et al. (2016) found the HtSDS being -0.31 , $+0.24$, and $+0.48$ in three untreated Turkish children with 5 α RD. In this

systemic multicenter study, we aimed to determine the growth of 5 α RD children.

MATERIALS AND METHODS

Patients

All 187 patients with 5 α RD (age 0–16 years) and admitted to eight pediatric endocrine centers from January 2010 to December 2017 were included in this retrospective cohort study. Their clinical manifestations included ambiguous genitalia at birth, hypospadias, micropenis, and cryptorchid. The ratio of T over dihydrotestosterone (T/DHT) after HCG stimulation test in all patients fluctuated between 10.67 and 86.56 (M: 32.72). All patients were diagnosed as 5 α RD according to their manifestations and ratio of T/DHT > 8.5 (Avendano et al., 2018). Then, the diagnosis was confirmed by pathogenic *SRD5A2* gene mutations. All patients had 46, XY karyotype and did not require hormonal treatment. The serum T level was >100 ng/dL after HCG stimulation test.

We excluded patients with malformation, abnormal functioning of the liver and kidney, or other with systemic diseases which may affect the physical development. Patients with 17 β -hydroxysteroid dehydrogenase type 3, androgen insensitivity syndrome, and other types of 46, XY DSD were excluded by biochemical diagnosis and genetic confirmation. Informed consent for genetic testing was obtained from parents or caregivers of all study subjects.

In our study, we enrolled 141 cases and excluded 46 cases according to the inclusion and exclusion criteria. There were no significant differences in initial gender, age, geographical distribution, BWt, BL, BA, LH, FSH, T, and DHT levels after HCG stimulation test between inclusion and exclusion groups ($\chi^2 = 0.983$, $P = 0.321$, $t = 1.483$, $P = 0.140$, $\chi^2 = 3.678$, $P = 0.055$, $Z = -0.864$, $P = 0.388$, $Z = -0.288$, $P = 0.773$, $Z = -1.700$, $P = 0.095$, $Z = -1.101$, $P = 0.271$, $Z = -0.206$, $P = 0.837$, $Z = -0.612$, $P = 0.540$, $Z = -0.164$, $P = 0.870$, respectively). Also patients with EMS < 7 showed no significantly different from those with EMS ≥ 7 between inclusion and exclusion groups ($\chi^2 = 2.841$, $P = 0.092$).

Data Collection

Data on anthropometry, CA, native place, gestational term, BL, BWt, family history, external genitalia and EMS (Ahmed et al., 2000), baseline LH and FSH, BA, T, and DHT levels after HCG stimulation test (Ishii et al., 2015) were recorded. Height was measured as orthostatic height when patients were aged > 3 years. Patients aged < 3 years were measured based on supine length. For each case, the height and weight were measured three times by experienced nurses, and the average value was considered.

Outcome Measures

The following were main outcomes: HtSDS and WtSDS of different age groups were compared to the control subjects of the same age (Li et al., 2009).

We chose Chinese normal boys as the control group for the following reasons: All patients in the study had the

46, XY karyotype; about 60% of 5 α RD children who were assigned the female gender in the period of infancy had marked masculinization and were reassigned as males at puberty (Lee et al., 2006).

HCG Stimulation Test, BA, and Hormone Examination

Bone age, hormonal, and HCG stimulation test data were collected from the Beijing Children's Hospital. Hormones were tested by enzyme enhanced chemiluminescence assay (Siemens Immulite 2000, Munich, Germany). BA was assessed according to the G-P atlas, by the same endocrinologist based on radiographic films of the left hand.

Growth Curves Plotting

The internationally well accepted method (λ -median-coefficient of variation, LMS) for generating standard curves was adopted to calculate the M, S, and L (after converting the data into a normal distribution, using Box-Cox transformation) (Cole and Green, 1992), which described the growth index in each age band. L, M, and S of smooth curves and the required percentile were calculated using age as an independent variable. Growth curves (P3, P10, P25, P50, P75, P90, P97 percentile curves as well as $-2SD$, $-1SD$, $0SD$, $+1SD$, $+2SD$ standard deviation curves) for children in the age groups of 0–36 months and 3–13 years were constructed.

Gene Analysis

All patients underwent genetic testing. Five to ten milliliters of peripheral blood was collected in disposable vacuum tubes for genetic testing. Genomic DNA was isolated using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. After construction according to the standard protocol, whole exon sequencing (100X) of the libraries was performed with SureSelect Human All Exon V6 array on the Illumina HiSeq X Ten Platform with PE150 strategy. To detect the potential variants in the cases, we performed bioinformatic processing and data analysis after receiving the primary sequencing data. Sequence variants were carefully identified with the help of GATK (McKenna et al., 2010) software following the best practice guidelines recommended by GATK (DePristo et al., 2011; Van der Auwera et al., 2013), including local realignment around INDELs, base quality score recalibration, followed by SNVs. INDELs were called simultaneously with the default setting of GATK Unified Genotyper on the realigned and recalibrated reads, followed by SNV and INDEL filtering to eliminate false-positive calls. The pathogenic *SRD5A2* gene mutations were confirmed by homozygous or compound heterozygous mutations inherited from parents or *de novo*. Nucleotide sequences of *SRD5A2* gene were compared with the published data. The pathogenicity of unreported *SRD5A2* mutations was tested by *in silico* analysis using two software of "SIFT" (<https://sift.bii.a-star.edu.sg/www/code.html>) and "Polyphen" (<http://genetics.bwh.harvard.edu/pph2/>).

The interpretation of gene pathogenicity is based on the ACMG (Richards et al., 2015).

Data Analysis

The calculations for the growth curves were performed using LMS-chartmaker Pro software, and curves were drawn using the GraphPad Prism 6 software. SPSS 23.0 software was used for statistical analyses. Data pertaining to quantitative variables were expressed as mean \pm SD or quartiles. Intergroup differences between two groups were assessed using the Student's *t*-test for normally distributed data and Mann-Whitney *U*-test for non-normally distributed data. Multiple age group comparisons were assessed using the Kruskal-Wallis *H*-test and Bonferroni correction. The ratio of BA/CA was assessed using 95% CIs. Intergroup differences between inclusion and exclusion groups were assessed using the Chi-square test for qualitative data. $P < 0.05$ was considered statistically significant. Geographic difference analysis was based on north and south regions; Qinling Mountains and Huaihe River were selected as geographical boundaries between the south and north.

RESULTS

General Data of Chinese Cases With 5 α RD

The initial sex assignment was male in 143/187 (76.5%) and female in 44/187 (23.5%) cases. The clinical presentation was simple micropenis ($n = 64$, 34.2%); simple hypospadias ($n = 43$, 23.0%); ambiguous genitalia ($n = 37$, 19.8%); simple cryptorchid ($n = 19$, 10.2%); and others (micropenis and/or hypospadias and/or cryptorchid) ($n = 24$, 12.8%). Only two patients were siblings; the others were unrelated. The details of standard phenotype and gene mutations (Jia and Shi, 2017) were shown in the section of **Supplementary Tables 1, 2**.

Only 141 of 187 patients with 5 α RD who had the data of height and weight and were aged between 0.08 and 16 years (M age: 1.75 years) were enrolled in our study. All 141 children with 5 α RD came from 25 provinces and municipalities across China (**Figure 1**). Seven cases had presentations of puberty, with a testicular volume of ≥ 4 mL (**Table 1**). Four cases aged between 10 and 12 years (M age: 11.83 years) were in Tanner stage 2. The other 3 cases aged ≥ 13 years (M age: 15.16 years) were in

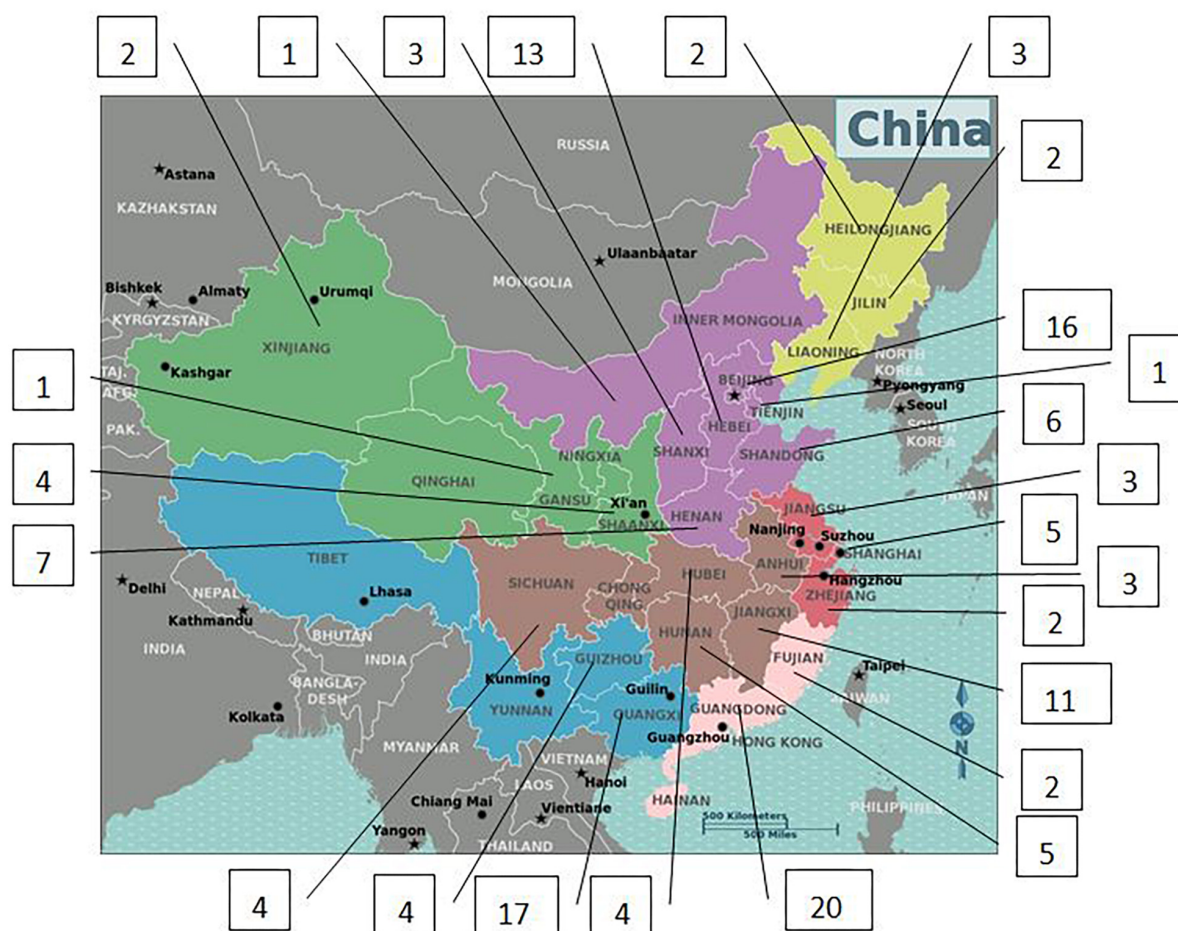


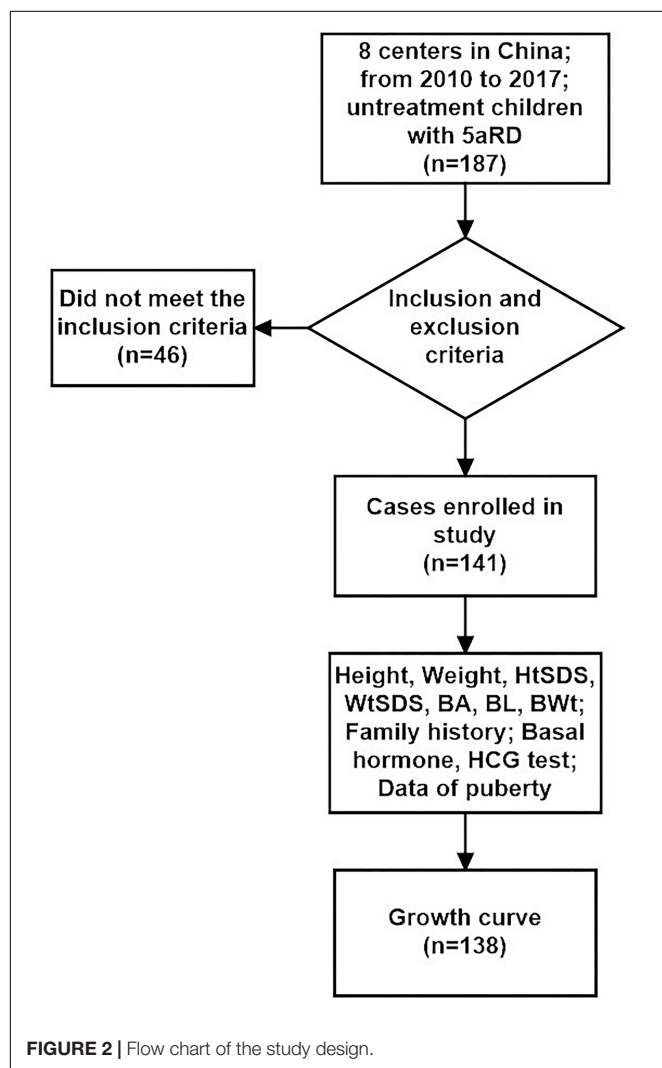
FIGURE 1 | Geographical distribution of children with 5 α RD in China.

TABLE 1 | Clinical parameters of children with 5 α RD in the pubertal period.

No.	Age (year)	Ht (cm)	HtSDS	Wt (cm)	WtSDS	BMI (kg/m ²)	BA (year)	Testes volume (ml)
1	11.17	143	−0.4	45.2	0.9	22.17	10.7	4
2	11.5	149	0	38	−0.3	17.16	12.6	4
3	12.17	148	−0.6	37	−0.7	16.94	–	5
4	12.83	152	−0.5	45.5	0	19.75	13	6
5	14.5	164	−0.6	49.5	−0.62	18.4	14	10
6	15	160	−1.5	44	−1.5	17.19	–	8
7	16	175	0.5	65	0.5	21.22	14	25

Data of cases 1–4 were included for the growth curve. Data of cases 5–7 were excluded for the growth curve. Ht, height; Wt, weight; HtSDS, height standard deviation score; WtSDS, weight standard deviation score; BMI, body mass index; BA, bone age.

Tanner stage 3–5. When plotting growth curve, we enrolled 138 cases aged <13 years because only 3 patients were aged between 13 and 16 years (Table 1, cases 5–7). Patients were divided into the following six groups: 0–5 months, 6–11 months, 1–2 years, 3–5 years, 6–9 years, and 10–12 years. The flow chart of the study is presented in Figure 2.



Physical Parameters of Chinese Children With 5 α RD (Table 2)

Height and weight curves are shown in Figures 3–6.

Height standard deviation scores and WtSDS of 138 cases with 5 α RD were 0.21 ± 1.25 and 0.10 ± 1.02 , respectively and in the normal range. The average BL of 5 α RD was 50.00 (49.00, 50.00) cm and the average BWt of 5 α RD was 3.40 (3.00, 3.70) kg, within the normal range, which was comparable to the normal reference standard for the Chinese population. Compared to the normal reference values for boys of the same age, HtSDS and WtSDS of 5 α RD patients were higher when they were younger than 2 years old (the mean of HtSDS and WtSDS ranged from 0SD to +1SD) with significant difference aged <1 year in HtSDS ($t = 3.658$, 2.103, $P = 0.002$, 0.048; respectively) and aged <6 months in WtSDS ($t = 2.756$, $P = 0.012$). After that, their height and weight plateaued gradually and fluctuated around the median for normal Chinese boys until the age of 13 years (Figure 7). To sum up, these data showed that the height in 5 α RD children increased faster than the normal population before 2 years of age, especially within the first 1 year of life, after which, the growth velocity gradually reduced and fluctuated near the median height of normal Chinese boys of the same age.

Physical Assessments and Hormone Levels of 5 α RD Among Different Age Groups

Physical Assessment for Chinese Children With 5 α RD Among Different Age Groups (Table 2)

The highest and lowest HtSDS were found in the 0–5 months and 3–5 years groups. HtSDS in the 0–5 months group had significant differences with the 1–2 years and 3–5 years groups ($P = 0.001$, 0.004, respectively), while no significant differences were observed in the other groups.

The highest WtSDS was observed in the 0–5 months group; however, no significant differences were observed among the different age groups.

Hormones in Different Age Groups of 5 α RD

The baseline levels of LH, FSH and the levels of T and DHT after HCG stimulation test are shown in Table 3. The lowest levels of LH, FSH, and T were found in the 3–5 years group.

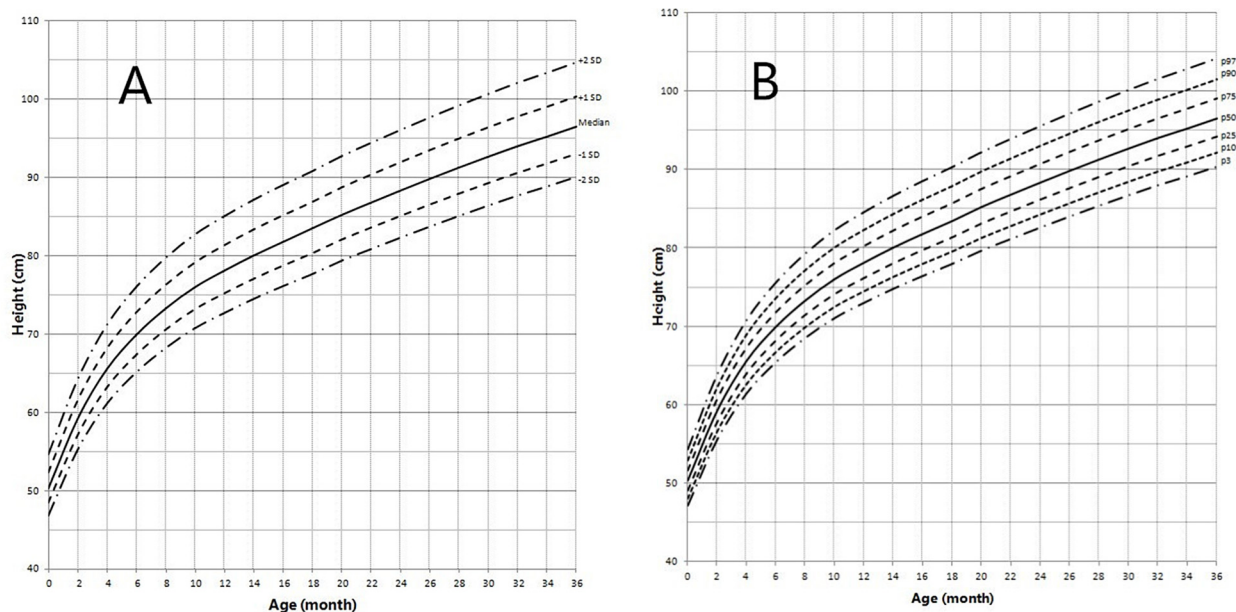


FIGURE 3 | Height in children with 5 α RD (0–36 months old). **(A)** Standard deviation curve; **(B)** Percentile curve.

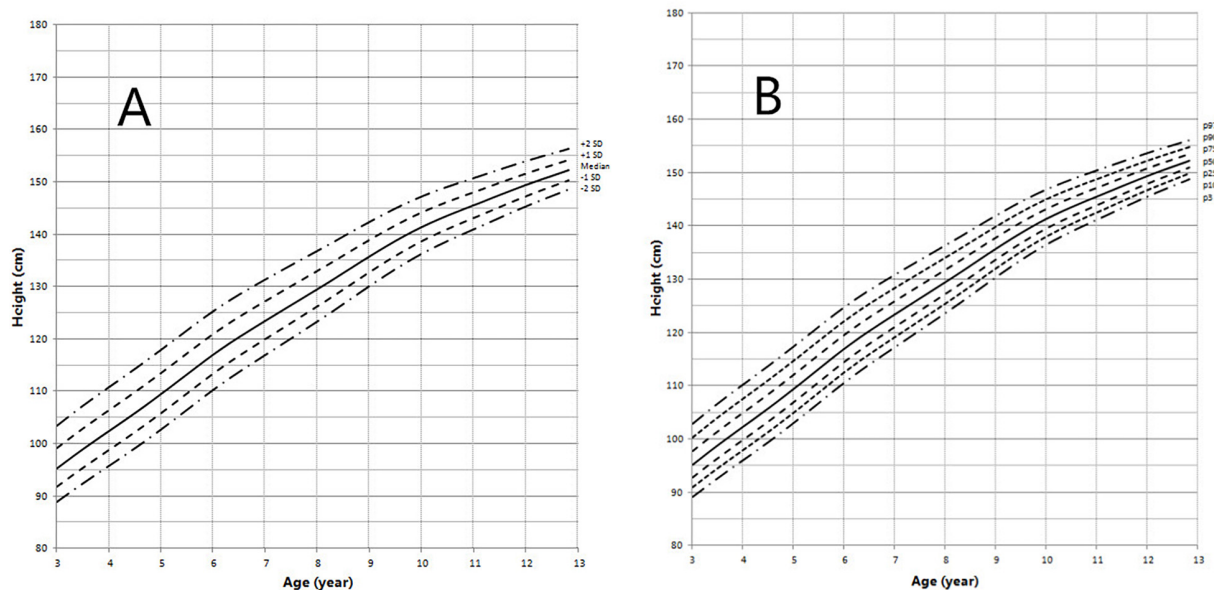


FIGURE 4 | Height in children with 5 α RD (3–13 years old). **(A)** Standard deviation curve; **(B)** Percentile curve.

LH and FSH in the 0–5 months and 10–12 years groups were higher than the other age groups with significant differences to those in the 3–5 years group ($P = 0.003, 0.048, 0.007, 0.009$, respectively). The top three levels of T were observed in the 0–5 months, 6–11 months, and 10–12 years groups. Further, T in the 0–5 months and 6–11 months groups were significantly higher than in those aged 1–9 years ($P = 0.001, 0.001, 0.003, 0.027, 0.002, 0.033$, respectively). The lowest DHT level was seen in the 6–9 years group. The highest DHT

was seen in the 0–5 months group and showed significant differences with those in the 6–9 years group ($P = 0.008$). The variation tendency of HtSDS in 5 α RDs was consistent with the trend of T.

Ratio of BA/CA in Children With 5 α RD (Table 4)

The ratio of BA/CA ($N = 68$) was 0.88 ± 0.23 with 95% CI < 1 . Additionally, 95% CI < 1 for BA/CA ratio were found in those aged 1–5 years, which meant that BA in 5 α RDs aged between 1

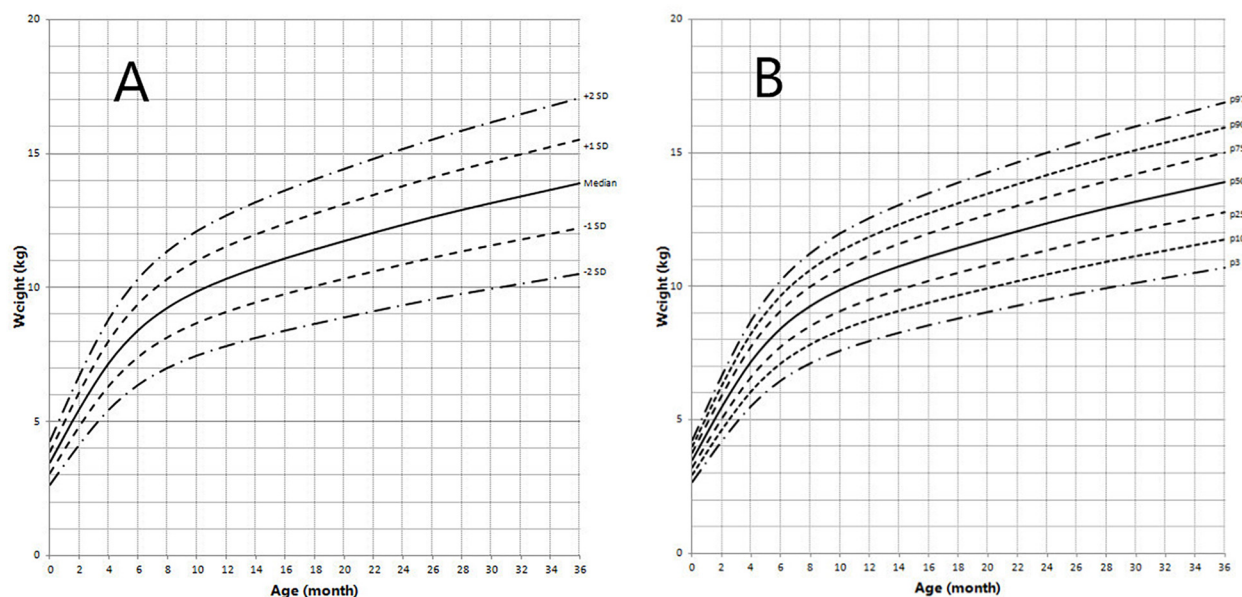


FIGURE 5 | Weight in children with 5 α RD (0–36 months old). **(A)** Standard deviation curve; **(B)** Percentile curve.

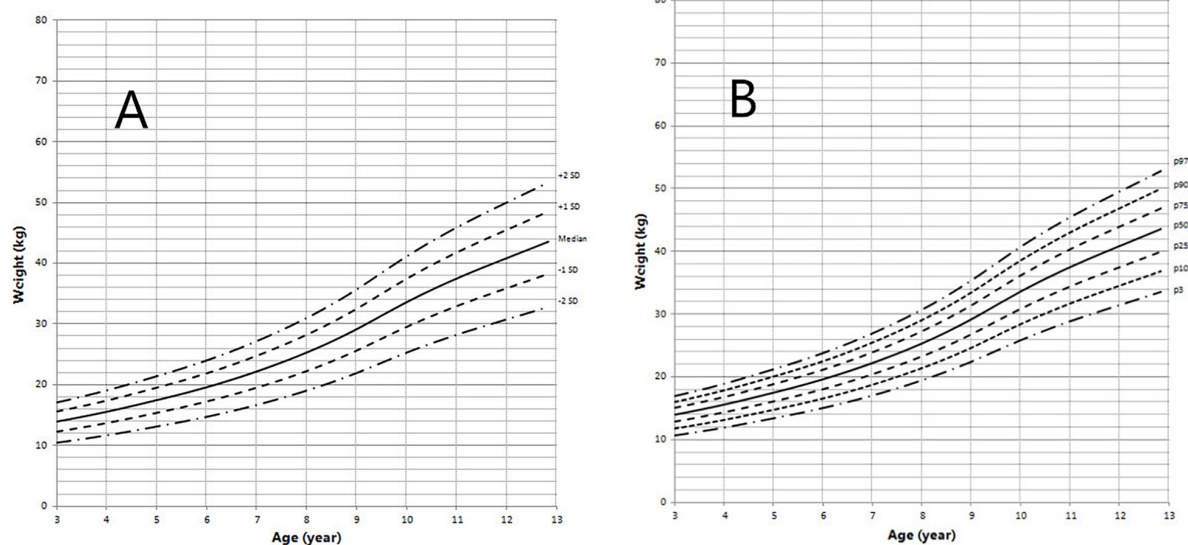


FIGURE 6 | Weight in children with 5 α RD (3–13 years old). **(A)** Standard deviation curve; **(B)** Percentile curve.

and 5 years was delayed, then BA was almost near to CA in the other age groups.

Physical Assessments and Hormone Levels of 5 α RD According to EMS (Table 5)

Children with 5 α RD were divided into two groups according to EMS ($n = 95$). One group was severely undervirilized with an EMS < 7 , and the other was mildly undervirilized with an EMS ≥ 7 . In the EMS < 7 group, HtSDS, WtSDS, BL, and T

after HCG stimulation test were higher, and DHT after HCG stimulation test was lower than those in the EMS ≥ 7 group without significant differences. BWt in EMS < 7 group was significantly higher ($Z = -2.191$, $P = 0.028$).

Geographical Difference in Children With 5 α RD (Table 6)

Data indicated that THtSDS, BWt, and WtSDS in children with 5 α RD from northern China were significantly higher than those from the southern region ($Z = -4.556$, -2.558 , -2.670 , $P = 0.001$,

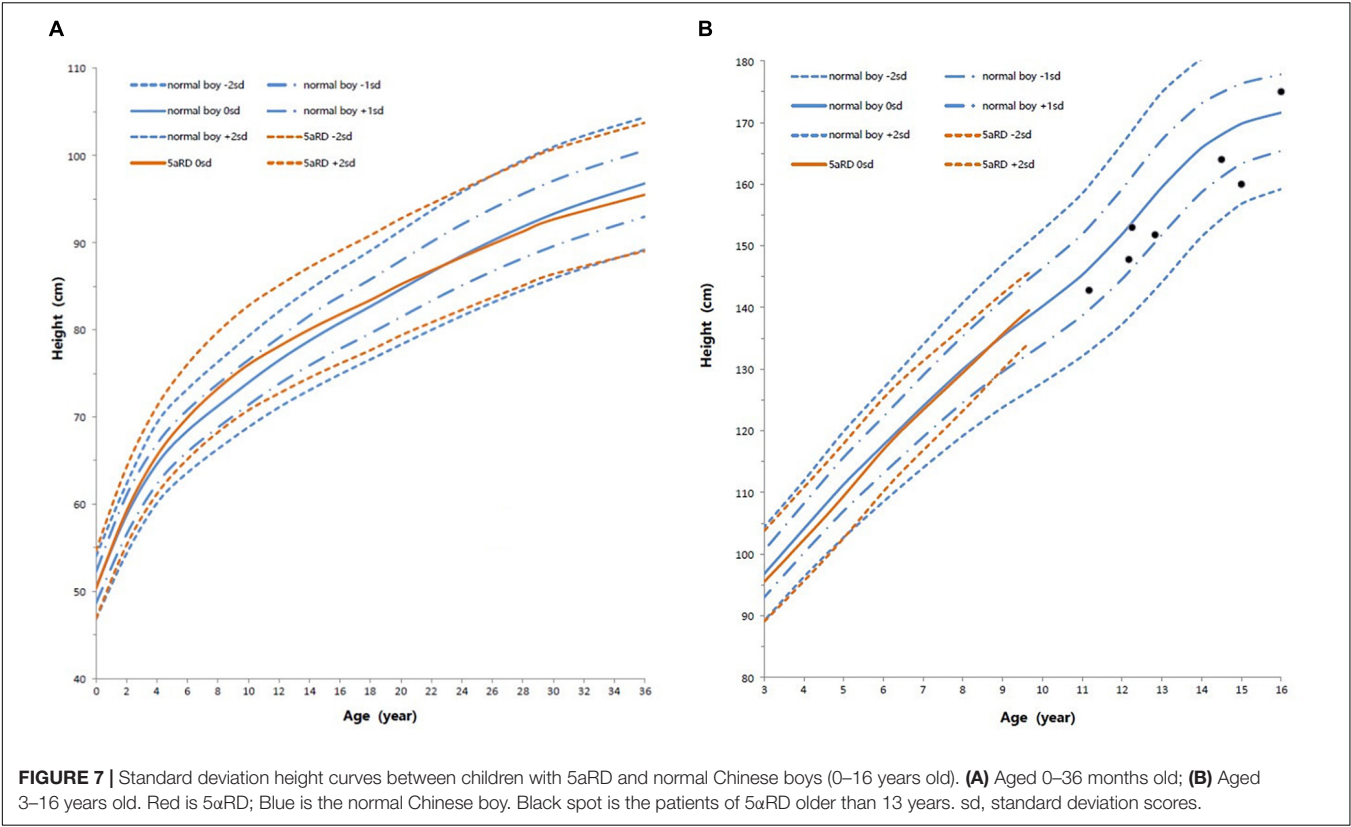


TABLE 2 | Physical assessments for Chinese children with 5 α RD among different age groups.

Age groups	<i>n</i>	Age (year)	THtSDS	BWt (kg)	BL (cm)	HtSDS	WtSDS	BMI (kg/m ²)
0–5 m	21	0.3 (0.3, 0.4)	0.1 ± 0.7	3.5 (3.2, 3.7)	50.0 (49.0, 50.0)	1.2 ± 1.4 ^a	0.6 ± 1.1 ^a	17.6 (16.9, 18.3)
6–11 m	23	0.7 (0.6, 0.9)	0.3 ± 0.6	3.3 (3.0, 3.8)	50.0 (49.0, 50.0)	0.5 ± 1.1 ^a	0.3 ± 1.0	17.1 (16.1, 19.0)
1–2 y	48	1.8 (1.4, 2.2)	0.1 ± 0.8	3.5 (3.0, 3.8)	50.0 (49.0, 50.0)	−0.1 ± 1.2 ^b	−0.1 ± 1.0	16.5 ± 1.4
3–5 y	24	3.8 (3.5, 4.8)	−0.1 (−0.6, 0.3)	3.2 (2.8, 3.5)	50.0 (49.0, 50.0)	−0.2 ± 1.3 ^b	−0.2 ± 1.3	15.8 ± 1.8
6–9 y	11	7.1 (6.9, 9.0)	−0.4 (−0.9, 0.1)	3.4 (2.9, 3.6)	50.0 (48.0, 50.0)	0.2 ± 0.8	−0.1 ± 1.0	15.3 (13.4, 17.8)
10–12 y	11	11.3 (10.0, 12.2)	−0.1 ± 0.9	3.5 (3.1, 3.7)	50.0 (48.0, 50.5)	0.1 ± 0.5	−0.0 ± 0.7	17.9 ± 2.2
Total	138	1.8 (0.8, 3.8)	−0.08 (−0.50, 0.50)	3.4 (3.0, 3.7)	50.0 (49.0, 50.0)	0.2 ± 1.3	0.1 ± 1.0	16.9 (15.5, 17.8)
<i>P</i>	^a					0.02, 0.048	0.012	

P < 0.05, ^a vs. median of normal Chinese boys' general reference values in the same age groups. ^b vs. 0–5 m group. THtSDS, target height standard deviation score; BWt, birth weight; BL, birth length; HtSDS, height standard deviation score; WtSDS, weight standard deviation score; BMI, body mass index; m, months; y, years.

0.011, 0.008, respectively), and no differences were found in age, BL, HtSDS, BMI, BA/CA, T, and DHT between the two groups.

DISCUSSION

5 α -reductase type 2 deficiency is a 5 α -reductase isoenzyme 2 formation defect caused by the SRD5A2 gene mutation. Its clinical profile ranges from 46, XY presented as complete female external genitalia to under masculinized male external genitalia (Sinnecker et al., 1996; Mendonca et al., 2016) such as enlarged clitoris, hypospadias, and micropenis (Ng et al., 1990; Gad et al., 1997; Silver and Russell, 1999; Nicoletti et al., 2005; Avendano et al., 2018). With the help of gene analysis, more and more 5 α RD can be diagnosed in some cases with the mild phenotype

(Wang et al., 2004; Nie et al., 2011). These give clinicians a chance to profile the growth of children with 5 α RD. The growth pattern of 5 α RD is becoming a concerned focus by clinicians.

Height Features in Chinese Children With 5 α RDs

In the present study, the plotted growth chart showed that the average HtSDS in 5 α RD patients was within the range for normal reference values of Chinese boys. Children with 5 α RD grow faster than normal boys before the age of 2 years, particularly when <1 year old. Between the ages of 2 and 13 years, their growth velocity decreases gradually and height fluctuates around the median height of normal Chinese boys. This trend was concordant with stimulating T fluctuation levels

TABLE 3 | Hormonal characterization for Chinese children with 5 α RD among different age groups.

Agegroups	n	LH (mIU/ml)	FSH (mIU/ml)	T (ng/dl)	DHT (ng/ml)
0–5 m	14	1.88 (1.34, 3.12)	2.12 (1.32, 3.50)	489.68 \pm 103.44	149.69 \pm 35.56
6–11 m	14	0.34 (0.14, 3.0)	1.22 (0.91, 2.10)	448.18 \pm 99.36	134.00 (101.14, 179.25)
1–2 y	36	0.27 (0.12, 0.95)	0.78 (0.43, 1.31)	331.01 \pm 85.79 ^{a,c}	127.30 (97.43, 164.08)
3–5 y	15	0.12 (0.10, 0.41) ^{a,b}	0.85 (0.65, 1.56) ^{a,b}	282.93 \pm 70.91 ^{a,c}	100.315 (70.83, 133.70)
6–9 y	9	0.29 (0.12, 1.44)	1.85 (0.50, 3.10)	287.63 \pm 138.97 ^{a,c}	69.90 (37.19, 101.05) ^a
10–12 y	7	1.20 (0.64, 3.79)	3.15 \pm 1.57	360.03 \pm 147.01	118.16 \pm 51.75
Total	95	0.36 (0.13, 1.56)	1.27 (0.71, 2.30)	362.10 \pm 121.60	123.00 (95.10, 162.32)
P	a	0.003	0.048	0.001, 0.001, 0.003	0.008
	b	0.007	0.009		
	c			0.027, 0.002, 0.033	

P < 0.05, a vs. 0–5 m group, b vs. 10–12 years group, and c vs. 6–11 m group. Testosterone (T) and dihydrotestosterone (DHT) were data after human chorionic gonadotropin stimulation test. LH, luteinizing hormone; FSH, follicle stimulating hormone; m, months; y, years.

TABLE 4 | BA/CA in children with 5 α RD among different age groups.

Agegroups	BA/CA	95%CI	
		Lower	Upper
0–5 m	1.0 \pm 0.32	0.7045	1.3197
6–11 m	0.77 \pm 0.25	0.4256	1.0700
1–2 y	0.76 \pm 0.23 ^c	0.6611	0.8495
3–5 y	0.93 \pm 0.21 ^c	0.7962	0.9495
6–9 y	0.99 \pm 0.04	0.8852	1.0882
10–12 y	0.93 \pm 0.04	0.8279	1.0341
Total	0.88 \pm 0.23 ^c	0.8226	0.9359

P < 0.05, c vs. 1. BA, bone age; CA, chronological age; m, months; y, years.

in 5 α RD children in the growing age. 5 α RD had decreased 5 α -reductase type 2 enzymatic activity caused by SRD5A2 gene mutation. In 5 α RDs, the lower the residual activity of 5 α -reductase type 2 isoenzyme, the greater the severity of the manifestation and higher accumulation of T. Patients with severe undervirilized male external genitalia (EMS < 7) had slightly higher HtSDS, WtSDS, BL, and T after the HCG stimulation test and significantly higher BWt. All these hint at T having an impact on their growth. In 5 α RD children, increased activity of the hypothalamic-pituitary-testicular axis during infancy would result in higher T levels because of minipuberty and diminished conversion to DHT. Thereafter, the slowed growth rate is a consequence of lower T during the quiescence of childhood before puberty. Previous studies have shown that growth in DSD children is associated with androgens (Hughes et al., 2002; Richter-Unruh et al., 2004; Han et al., 2008; Di et al., 2013; Wu et al., 2017). For example, Han et al. (2008) assessed the

height in patients with CAIS and found that patients who underwent gonadectomy after adolescence or during adulthood were taller than those who underwent the same surgery in the pre-pubertal phase. The same conclusion was drawn when comparing with testicular dysfunction 46, XY DSD (Wu et al., 2017). This suggests that androgen affects growth in the pre-pubertal stage despite being undetectable in childhood. T and DHT are two types of androgen. T increases growth in association with a direct elevation of GH and an indirect elevation of IGF-I, respectively; the latter occurs due to estrogen by peripheral and central aromatization (Veldhuis et al., 2005a,b, 2009; Perry et al., 2008). DHT, on the other hand, could not up-regulate the function of hypothalamo-somatotrope-IGF-I axis directly (Veldhuis et al., 1997) and could not be aromatized to estrogen. There are two 5-alpha-reductase isoenzymes (Wilson, 1972). The type 1 isoenzyme is distributed in the bone, skeletal muscle, osteoblast-like cells, and a few other tissues (Wilson, 1972; Thigpen et al., 1993; Issa et al., 2002; Tria et al., 2004), whereas the type 2 isoenzyme is distributed in the prostate, seminal vesicle epididymis, medulla oblongata, and other tissues (Thigpen et al., 1993; Tria et al., 2004). The capability of T to induce biologic actions in bone depends on localized intraskeletal sex steroid hormone metabolism via type-1 5-alpha-reductase isoenzyme specially expressed in bone tissue (Yarrow et al., 2015). Furthermore, T surges in perinatal periods and consequent imprinting of the GH/IGF-I axis has important positive effects on the growth plate (Sims et al., 2006; Vanderschueren et al., 2014). In conclusion, growth of children with 5 α RD was affected by high levels of T, while DHT plays a very minor role in their growth. Individuals with 5 α RD have exhibited different residual activity of 5 α -reductase. T treatment, which may be considered

TABLE 5 | Physical assessments for Chinese children with 5 α RD between two EMS groups.

EMS	n	BWt (kg)	BL (cm)	HtSDS	WtSDS	BMI (kg/m ²)	T (ng/dL)	DHT (ng/mL)
<7	45	3.5 \pm 0.5	50.0 (49.0, 50.0)	0.5 \pm 1.5	0.2 \pm 0.9	16.9 (15.4, 17.8)	363.3 (296.7, 415.4)	112.3 (73.1, 160.0)
\geq 7	50	3.3 (2.9, 3.7) ^a	49.5 (48.3, 50.0)	0.2 \pm 1.0	0.0 \pm 1.1	16.6 (15.2, 17.8)	350.7 \pm 121.4	128.6 (99.4, 166.9)
P	a	0.028	0.083	0.350	0.388	0.758	0.371	0.166

P < 0.05, a vs. 5 α RDs with EMS < 7. Testosterone (T) and dihydrotestosterone (DHT) were data after human chorionic gonadotropin stimulation test. EMS, external masculinization score; BWt, birth weight; BL, birth length; HtSDS, height standard deviation score; WtSDS, weight standard deviation score; BMI, body mass index.

TABLE 6 | Clinical parameters of children with 5 α RD in the north and south region of China.

Region	n	Age (year)	THtSDS	BWt (kg)	BL (cm)	HtSDS	WtSDS	BMI (kg/m ²)	BA/CA	T (ng/dL)	DHT (ng/mL)
South	77	2.00 (0.86, 4.96)	-0.47 \pm 1.74	3.30 \pm 0.43	50 (49, 50)	0.10 \pm 1.24	-0.20 (-0.80, 0.50)	16.66 \pm 2.27	0.89 \pm 0.23	352.01 \pm 151.20	103.00 (63.61, 259.52)
North	61	1.42 (0.68, 3.07)	0.25 (-0.17, 0.75) ^a	3.44 \pm 0.67 ^a	50 (49, 50)	0.20 (-0.60, 1.00)	0.33 \pm 1.08 ^a	17.08 (15.80, 17.78)	0.91 (0.67, 1.03)	346.50 (235.75, 491.50)	131.70 (67.90, 240.18)
P	a	0.058	0.001	0.011	0.504	0.485	0.008	0.356	0.943	0.766	0.440

P < 0.05, a South region vs. north region using Student's t-test for normal distribution data and Mann-Whitney U-test for non-normal distribution data. T (testosterone) and DHT (dihydrotestosterone) were data after human chorionic gonadotropin stimulation test. THtSDS, target height standard deviation score; BWt, birth weight; BL, birth length; HtSDS, height standard deviation score; WtSDS, weight standard deviation score; BMI, body mass index; BA, bone age; CA, chronological age.

anyway for 5 α RDs patients with micropenis, may also have an extra benefit on their growth.

Bone Age Feature in 5 α RDs

Bone age is the best index of growth potential ability. Sex hormones promote bone maturation, which, in different types of DSDs show special characteristics (Bertelloni et al., 2010). Short-term treatment with testosterone undecanoate (Andriol) did not change the BA in our previous study (Chen et al., 2012). The average ratio of BA/CA in this study of 5 α RD children was lower than 1, particularly in the ages of 1–5 years. Perhaps, physiological androgen resistance during early childhood could explain the BA delay in presence of high T concentration among 5 α RD children. This research addressed concerns related to treatment with androgen products on bone maturation and provided some supporting evidence for androgen therapy in childhood. Suitable T therapy may be beneficial for either micropenis or height growth among children with 5 α RD.

Puberty Feature in Chinese 5 α RDs

Four of 11 cases aged 10–12 years (M age: 11.83 years) had the signs of puberty with Tanner stage 2. Another 3 cases aged >13 years (M age: 15.16 years) who were excluded from the growth curve were in Tanner stage 3–5. Perhaps these data showed that the onset time of puberty in children with 5 α RD was somewhat more delayed than those of normal Chinese boys (M age: 10.55 years) (Pubertal Study Group of the Subspeciality Group of Endocrinologie, Hereditary and Matabolic Diseases, Society of Pediatrics, and Chinese Medical Association, 2010). Growth deceleration before puberty may partly explain the growth retardation in 5 α RD during the age of 10–12 years.

Study Limitation

We acknowledge that our study has some limitations. There were relatively few patients older than 10 years; thus, the growth pattern in children with 5 α RD during the puberty period was limited. Seven children aged >10 years were in the pubertal stage with Tanner stage 2–5. These data partly reflected the onset time of puberty in children with 5 α RD. We need more cases to validate the therapeutic opinions in the follow-up study. Owing to technical constraints in most Chinese labs hereto, serum LH, FSH, T, and DHT were measured by enzyme enhanced chemiluminescence assay and not as per the gold standard of mass spectrometry. Nonetheless, our hormone results had the value to assess the patients' sex hormones, because there are several publications that have used the same testing methods, and enzyme enhanced chemiluminescence assay is still widely used to measure hormones in many countries currently. The levels of GH, IGF-I, and estrogens were not mentioned in the research due to lack of sufficient data in patients' medical records.

CONCLUSION

The growth curve of children with 5 α RD revealed the special pattern affected by T, while DHT plays a very minor role in it. Our growth curve can provide the reference for clinical judgment

of Chinese children with 5 α RD. In addition, patients showed lagging BA. Furthermore, androgen treatment, which may be considered anyway for 5 α RD patients with micropenis, may also be beneficial for their growth.

NOTE

Number of cases with 5 α RD in the present study

No.	Center	Case number	Enrolled number
1	Beijing Children's Hospital	104	95
2	Maternal and Children Health Hospital of Guangxi Zhuang Autonomous Region	23	13
3	Shenzhen Children's Hospital	36	18
4	Children's Hospital of Soochow University	6	2
5	Fuzhou Children's Hospital	2	2
6	Shanghai Children's Medical Center	6	4
7	Jiangxi Province Children's Hospital	5	3
8	Children's Hospital of Fudan University	5	4
	Total	187	141

ETHICS STATEMENT

This study was approved by the ethical committee of Beijing Children's Hospital. This study was approved by

the ethical committee of Beijing Children's Hospital. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

XZ contributed to data collection, data interpretation, and writing of the report. CG contributed to study design and reviewed the paper. YS, SC, XW, FL, YY, LC, RC, DW, and ZS contributed to data collection. HC contributed to statistical analysis.

FUNDING

This work was funded by the Public Health Project for Residents in Beijing (Z151100003915103) and the National Key Research and Development Program of China (2016YFC0901505).

ACKNOWLEDGMENTS

The authors would like to thank all patients and their families for participation in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00173/full#supplementary-material>

REFERENCES

- Ahmed, S. F., Khwaja, O., and Hughes, I. A. (2000). The role of a clinical score in the assessment of ambiguous genitalia. *BJU Int.* 85, 120–124.
- Almeida, M., Laurent, M. R., Dubois, V., Claessens, F., O'Brien, C. A., Bouillon, R., et al. (2017). Estrogens and androgens in skeletal physiology and pathophysiology. *Physiol. Rev.* 97, 135–187. doi: 10.1152/physrev.00033.2015
- Avendano, A., Paradisi, I., Cammarata-Scalisi, F., and Callea, M. (2018). 5-alpha-Reductase type 2 deficiency: is there a genotype-phenotype correlation? A review. *Hormones* 17, 197–204. doi: 10.1007/s42000-018-0013-9
- Bertelloni, S., Baldinotti, F., Russo, G., Ghirri, P., Dati, E., Michelucci, A., et al. (2016). 5alpha-reductase-2 deficiency: clinical findings, endocrine pitfalls, and genetic features in a large italian cohort. *Sex. Dev.* 10, 28–36. doi: 10.1159/000445090
- Bertelloni, S., Baroncelli, G. I., and Mora, S. (2010). Bone health in disorders of sex differentiation. *Sex. Dev.* 4, 270–284. doi: 10.1159/000315961
- Byers, H. M., Mohnach, L. H., Fechner, P. Y., Chen, M., Thomas, I. H., Ramsdell, L. A., et al. (2017). Unexpected ethical dilemmas in sex assignment in 46,XY DSD due to 5-alpha reductase type 2 deficiency. *Am. J. Med. Genet. C Semin. Med. Genet.* 175, 260–267. doi: 10.1002/ajmg.c.31560
- Chen, J. J., Gong, C. X., Cao, B. Y., Di, W. U., and Yu-chuan, L. I. (2012). Clinical observation of short-term oral testosterone undecanoate treatment for 46, XY DSD Chinese boys with small penis: a self-comparison study. *Clin. J. Evid. Based Pediatr.* 7, 167–171. doi: 10.3969/j.issn.1673-5501.2012.03.002
- Choi, J. H., Kim, G. H., Seo, E. J., Kim, K. S., Kim, S. H., and Yoo, H. W. (2008). Molecular analysis of the AR and SRD5A2 genes in patients with 46,XY disorders of sex development. *J. Pediatr. Endocrinol. Metab.* 21, 545–553.
- Cole, T. J., and Green, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Stat. Med.* 11, 1305–1319.
- Deeb, A., Al, S. H., Ibukunoluwa, F., and Attia, S. (2016). Phenotype, sex of rearing, gender re-assignment, and response to medical treatment in extended family members with a novel mutation in the SRD5A2 gene. *J. Clin. Res. Pediatr. Endocrinol.* 8, 236–240. doi: 10.4274/jcrpe.2782
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Di, W. U., Gong, C. X., and Qin, M. (2013). Analysis of the clinical characteristics and body height in 153 disorders of sex development children without known cause. *Chin. J. Evid. Based Pediatr.* 8, 46–49.
- Eren, E., Edgunlu, T., Asut, E., and Karakas, C. S. (2016). Homozygous Ala65Pro mutation with V89L polymorphism in SRD5A2 deficiency. *J. Clin. Res. Pediatr. Endocrinol.* 8, 218–223. doi: 10.4274/jcrpe.2495
- Gad, Y. Z., Nasr, H., Mazen, I., Salah, N., and El-Ridi, R. (1997). 5 alpha-reductase deficiency in patients with micropenis. *Inherit. J. Metab. Dis.* 20, 95–101.
- Han, T. S., Goswami, D., Trikudanathan, S., Creighton, S. M., and Conway, G. S. (2008). Comparison of bone mineral density and body proportions between women with complete androgen insensitivity syndrome and women

- with gonadal dysgenesis. *Eur. Endocrinol. J.* 159, 179–185. doi: 10.1530/EJE-08-0166
- Hughes, I. A., Northstone, K., and Golding, J. (2002). Reduced birth weight in boys with hypospadias: an index of androgen dysfunction? *Arch. Dis. Child Fetal Neonatal Ed.* 87, F150–F151.
- Ishii, T., Matsuo, N., Sato, S., Ogata, T., Tamai, S., Anzo, M., et al. (2015). Human chorionic gonadotropin stimulation test in prepubertal children with micropenis can accurately predict leydig cell function in pubertal or postpubertal adolescents. *Horm. Res. Paediatr.* 84, 305–310. doi: 10.1159/000439234
- Issa, S., Schnabel, D., Feix, M., Wolf, L., Schaefer, H. E., Russell, D. W., et al. (2002). Human osteoblast-like cells express predominantly steroid 5 α -reductase type 1. *J. Clin. Endocrinol. Metab.* 87, 5401–5407. doi: 10.1210/jc.2001-011902
- Ittiwut, C., Pratuangdejikul, J., Supornsilchai, V., Muensri, S., Hiranras, Y., Sahakitrungruang, T., et al. (2017). Novel mutations of the SRD5A2 and AR genes in Thai patients with 46, XY disorders of sex development. *J. Pediatr. Endocrinol. Metab.* 30, 19–26. doi: 10.1515/jpem-2016-0048
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Ko, J. M., Cheon, C. K., Kim, G. H., Kim, S. H., Kim, K. S., and Yoo, H. W. (2010). Clinical characterization and analysis of the SRD5A2 gene in six Korean patients with 5 α -reductase type 2 deficiency. *Horm. Res. Paediatr.* 73, 41–48. doi: 10.1159/000271915
- Kolesinska, Z., Ahmed, S. F., Niedziela, M., Bryce, J., Molinska-Glura, M., Rodie, M., et al. (2014). Changes over time in sex assignment for disorders of sex development. *Pediatrics* 134, e710–e715. doi: 10.1542/peds.2014-1088
- Lee, P. A., Houk, C. P., Ahmed, S. F., and Hughes, I. A. (2006). Consensus statement on management of intersex disorders. International consensus conference on intersex. *Pediatrics* 118, e488–e500. doi: 10.1542/peds.2006-0738
- Li, H., Ji, C. Y., Zong, X. N., and Zhang, Y. Q. (2009). [Height and weight standardized growth charts for Chinese children and adolescents aged 0 to 18 years]. *Zhonghua Er Ke Za Zhi* 47, 487–492.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Mendonca, B. B., Batista, R. L., Domenice, S., Costa, E. M., Arnhold, I. J., Russell, D. W., et al. (2016). Steroid 5 α -reductase 2 deficiency. *J. Steroid Biochem. Mol. Biol.* 163, 206–211. doi: 10.1016/j.jsbmb.2016.05.020
- Ng, W. K., Taylor, N. F., Hughes, I. A., Taylor, J., Ransley, P. G., and Grant, D. B. (1990). 5 α -reductase deficiency without hypospadias. *Arch. Dis. Child.* 65, 1166–1167.
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60, 681–685. doi: 10.1007/s11427-017-9100-1
- Nicoletti, A., Baldazzi, L., Balsamo, A., Barp, L., Pirazzoli, P., Gennari, M., et al. (2005). SRD5A2 gene analysis in an Italian population of under-masculinized 46,XY subjects. *Clin. Endocrinol.* 63, 375–380. doi: 10.1111/j.1365-2265.2005.02348.x
- Nie, M., Zhou, Q., Mao, J., Lu, S., and Wu, X. (2011). Five novel mutations of SRD5A2 found in eight Chinese patients with 46,XY disorders of sex development. *Mol. Hum. Reprod.* 17, 57–62. doi: 10.1093/molehr/gaq072
- Nishiyama, K. K., Macdonald, H. M., Moore, S. A., Fung, T., Boyd, S. K., and McKay, H. A. (2012). Cortical porosity is higher in boys compared with girls at the distal radius and distal tibia during pubertal growth: an HR-pQCT study. *Bone J. Miner. Res.* 27, 273–282. doi: 10.1002/jbmr.552
- Perry, R. J., Farquharson, C., and Ahmed, S. F. (2008). The role of sex steroids in controlling pubertal growth. *Clin. Endocrinol.* 68, 4–15. doi: 10.1111/j.1365-2265.2007.02960.x
- Pubertal Study Group of the Subspecialty Group of Endocrinologie, Hereditary and Metabolic Diseases, Society of Pediatrics, and Chinese Medical Association (2010). Testis volume, pubic hair development and spermarcheal age in urban Chinese boys. *Chinese J. Pediatr.* 48, 418–424. doi: 10.3760/cma.j.issn.0578-1310.2010.06.005
- Raveenthiran, V. (2017). Controversies of sex re-assignment in genetic males with congenital inadequacy of the penis. *Indian Pediatr. J.* 84, 700–708. doi: 10.1007/s12098-017-2412-3
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30
- Richter-Unruh, A., Knauer-Fischer, S., Kaspers, S., Albrecht, B., Gillesen-Kaesbach, G., and Hauffa, B. P. (2004). Short stature in children with an apparently normal male phenotype can be caused by 45,X/46,XY mosaicism and is susceptible to growth hormone treatment. *Eur. Pediatr. J.* 163, 251–256. doi: 10.1007/s00431-004-1406-0
- Shen, Z. (2016). What are the biological roots of sexual orientation? *Chin. Sci. Bull.* 61, 1733–1747. doi: 10.1360/N972015-01212
- Silver, R. I., and Russell, D. W. (1999). 5 α -reductase type 2 mutations are present in some boys with isolated hypospadias. *J. Urol.* 162(3 Pt 2), 1142–1145.
- Sims, N. A., Brennan, K., Spaliviero, J., Handelsman, D. J., and Seibel, M. J. (2006). Perinatal testosterone surge is required for normal adult bone size but not for normal bone remodeling. *Am. J. Physiol. Endocrinol. Metab.* 290, E456–E462. doi: 10.1152/ajpendo.00311.2005
- Sinnecker, G. H., Hiort, O., Dibbelt, L., Albers, N., Dorr, H. G., Hauss, H., et al. (1996). Phenotypic classification of male pseudohermaphroditism due to steroid 5 α -reductase 2 deficiency. *Am. J. Med. Genet.* 63, 223–230.
- Sultan, C., Paris, F., Terouanne, B., Balaguer, P., and Georget, V. (2001). Disorders linked to insufficient androgen action in male children. *Hum. Reprod. Update* 7, 314–322. doi: 10.1093/humupd/7.3.314
- Thigpen, A. E., Silver, R. I., Guileyardo, J. M., Casey, M. L., McConnell, J. D., and Russell, D. W. (1993). Tissue distribution and ontogeny of steroid 5 α -reductase isozyme expression. *Clin. J. Invest.* 92, 903–910. doi: 10.1172/JCI116665
- Tria, A., Hiort, O., and Sinnecker, G. H. (2004). Steroid 5 α -reductase 1 polymorphisms and testosterone/dihydrotestosterone ratio in male patients with hypospadias. *Horm. Res.* 61, 180–183. doi: 10.1159/000076279
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33. doi: 10.1002/0471250953.bi1110s43
- Vanderschueren, D., Laurent, M. R., Claessens, F., Gielen, E., Lagerquist, M. K., Vandenput, L., et al. (2014). Sex steroid actions in male bone. *Endocr. Rev.* 35, 906–960. doi: 10.1210/er.2014-1024
- Veiga-Junior, N. N., Medaets, P. A., Petroli, R. J., Calais, F. L., de Mello, M. P., Castro, C. C., et al. (2012). Clinical and laboratorial features that may differentiate 46,XY DSD due to partial androgen insensitivity and 5 α -reductase type 2 deficiency. *Int. Endocrinol. J.* 2012:964876. doi: 10.1155/2012/964876
- Veldhuis, J. D., Anderson, S. M., Iranmanesh, A., and Bowers, C. Y. (2005a). Testosterone blunts feedback inhibition of growth hormone secretion by experimentally elevated insulin-like growth factor-I concentrations. *J. Clin. Endocrinol. Metab.* 90, 1613–1617. doi: 10.1210/jc.2004-1303
- Veldhuis, J. D., Frystyk, J., Iranmanesh, A., and Orskov, H. (2005b). Testosterone and estradiol regulate free insulin-like growth factor I (IGF-I), IGF binding protein 1 (IGFBP-1), and dimeric IGF-I/IGFBP-1 concentrations. *J. Clin. Endocrinol. Metab.* 90, 2941–2947. doi: 10.1210/jc.2004-1314
- Veldhuis, J. D., Erickson, D., Wigham, J., Weist, S., Miles, J. M., and Bowers, C. Y. (2011). Gender, sex-steroid, and secretagogue-selective recovery from growth hormone-induced feedback in older women and men. *J. Clin. Endocrinol. Metab.* 96, 2540–2547. doi: 10.1210/jc.2011-0298
- Veldhuis, J. D., Metzger, D. L., Martha, J. P. M., Mauras, N., Kerrigan, J. R., Keenan, B., et al. (1997). Estrogen and testosterone, but not a nonaromatizable androgen, direct network integration of the hypothalamo-somatotrope (growth hormone)-insulin-like growth factor I axis in the human: evidence from pubertal pathophysiology and sex-steroid hormone replacement*. *J. Clin. Endocrinol. Metab.* 82, 3414–3420.
- Veldhuis, J. D., Mielke, K. L., Cosma, M., Soares-Welch, C., Paulo, R., Miles, J. M., et al. (2009). Aromatase and 5 α -reductase inhibition during an exogenous

- testosterone clamp unveils selective sex steroid modulation of somatostatin and growth hormone secretagogue actions in healthy older men. *J. Clin. Endocrinol. Metab.* 94, 973–981. doi: 10.1210/jc.2008-2108
- Wang, Y., Gong, C., Wang, X., and Qin, M. (2017). AR mutations in 28 patients with androgen insensitivity syndrome (Prader grade 0-3). *Sci. China Life Sci.* 60, 700–706. doi: 10.1007/s11427-017-9084-9
- Wang, Y., Li, Q., Xu, J., Liu, Q., Wang, W., Lin, Y., et al. (2004). Mutation analysis of five candidate genes in Chinese patients with hypospadias. *Eur. Hum. J. Genet.* 12, 706–712. doi: 10.1038/sj.ejhg.5201232
- Wilson, J. D. (1972). Recent studies on the mechanism of action of testosterone. *N. Engl. J. Med.* 287, 1284–1291. doi: 10.1056/NEJM197212212872508
- Wilson, J. D., Griffin, J. E., and Russell, D. W. (1993). Steroid 5 alpha-reductase 2 deficiency. *Endocr. Rev.* 14, 577–593. doi: 10.1210/edrv-14-5-577
- Wu, D., Chen, H., and Gong, C. (2017). Physical assessment and reference growth curves for children with 46,XY disorders of sex development. *Pediatr. Invest.* 1, 16–22.
- Yarrow, J. F., Wronski, T. J., and Borst, S. E. (2015). Testosterone and adult male bone: actions independent of 5alpha-reductase and aromatase. *Exerc. Sport Sci. Rev.* 43, 222–230. doi: 10.1249/JES.0000000000000056

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhao, Song, Chen, Wang, Luo, Yang, Chen, Chen, Su, Wu and Gong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Immunohistochemical Evaluation of Histological Change in a Chinese Milroy Disease Family With Venous and Skin Abnormalities

Sijia Zhang^{1,2†}, Xihui Chen^{1†}, Lijuan Yuan^{1,3}, Shuyan Wang², Dangzhi Moli¹, Shujuan Liu^{4*} and Yuanming Wu^{1*}

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Liyuan Zhu,
National Center for Toxicological
Research (FDA), United States
Ting Li,
University of Arkansas at Little Rock,
United States

*Correspondence:

Shujuan Liu
hanliu@fmmu.edu.cn
Yuanming Wu
wuyym@fmmu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 12 November 2018

Accepted: 26 February 2019

Published: 19 March 2019

Citation:

Zhang S, Chen X, Yuan L, Wang S,
Moli D, Liu S and Wu Y (2019)
Immunohistochemical Evaluation of
Histological Change in a Chinese
Milroy Disease Family With Venous
and Skin Abnormalities.
Front. Genet. 10:206.
doi: 10.3389/fgene.2019.00206

¹ Department of Biochemistry and Molecular Biology, Center for DNA Typing, Air Force Medical University, Xi'an, China,

² State Key Laboratory of Military Stomatology and National Clinical Research Center for Oral Diseases and Shaanxi
Engineering Research Center for Dental Materials and Advanced Manufacture, Department of Implant Dentistry, School of
Stomatology, Air Force Medical University, Xi'an, China, ³ Department of General Surgery, Tangdu Hospital, Air Force Medical
University, Xi'an, China, ⁴ Department of Obstetrics and Gynecology, Xijing Hospital, Air Force Medical University, Xi'an, China

Background: Milroy disease (MD) is rare and autosomal dominant resulting from mutations of the vascular endothelial growth factor receptor-3 (*VEGFR-3* or *FLT4*), which leads to dysgenesis of the lymphatic system.

Methods: Here we report a Chinese MD family with 2 affected members of two generations. We identified the mutation of c.3075G>A in one allele of *FLT4* in Chinese population firstly. The father and child presented lymphedema under knees both. Unfortunately, the child was premature delivered for a car accident of the mother and then died of asphyxia. Then we gathered the tissue of the lower-limb from the child with permission from the parents and ethic committee. We stained the tissue with lymphatic marker D2-40 and hematoxylin-eosin to explore the histological changes. Afterwards, we compared the results with a normal child who unfortunately died of premature delivery also.

Results: It is firstly identified the mutation of *FLT4*: c.3075G>A in Chinese population, and the mutation inherited in the lineage. The histological evaluation indicated: (1) The number of lymphatic vessels decreased; (2) The morphology and structure of lymphatic vessels was abnormal. And what is added to our knowledge: (1) Capillary hyperemia and phlebectasia is severe; (2) Vascular malformations; (3) The number of vascular endothelial cells and vascular smooth muscle cells decreased; (4) Large sheets of epidermis desquamated; (5) The numbers of cutaneous appendages reduced in MD.

Conclusions: Based on the new findings, we assume that mutation of *FLT4* not only affect the lymphogenesis, but also the angiogenesis, and epidermis structure.

Keywords: Milroy disease, lymphedema, D2-40, fetus, *FLT4*

BACKGROUND

Milroy disease (MD; *hereditary lymphedema type I*; MIM# 153100) is first described by Milroy in 1892. It is known caused by the dysfunction of the lymphatic system so far, with the key features of congenital onset and primary lymphedema (Milroy, 1892). The estimated incidence is 1/6,000 worldwide, with a male/female ratio of 1:2.3 generally (Gezginc et al., 2012). Typical MD patients usually exhibit lymphedema at birth with swelling of the lower-limb, most times it is bilateral. The patients often have a brawny texture of the skin and hyperkeratosis is reported from the laboratory. The swelling is confined to the dorsum of the foot with deep skin creases which could be detected on the toes.

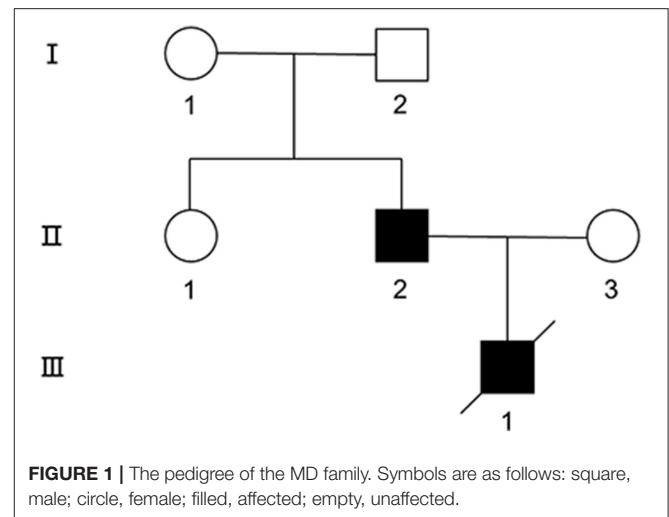
The gene locus of MD was first reported by Ferrell and Evans at chromosomal location 5q35 (Ferrell et al., 1998; Evans et al., 1999). The mutated gene (*FLT4*) in this region encodes a tyrosine kinase receptor for vascular endothelial growth factors C (*VEGFR-3*). The protein is believed to be involved in the process of lymphangiogenesis and maintenance of the lymphatic endothelium (Brice et al., 2005). Mutations of *FLT4* is responsible for about 75% of the diagnosed MD cases which has been published (Connell et al., 2009). Recent studies showed that the patients had large caliber great saphenous veins while presented no cutaneous signs of venous disease (Gordon et al., 2013). However, the superficial venous valve reflux indicates the venous development of MD patients might also be abnormal (Mellor et al., 2010).

For years, the limitations in the methods restrict us to present lymphatic vessels in MD patients, and the pathogenesis was mainly based on the evidence of lymphatic imaging, animal and cell studies. It is reported that the D2-40 monoclonal antibody could selectively detect lymphatic vessels (Bai et al., 2013), for it can specifically combine to a fixation-resistant epitope on a 40 kDa O-linked sialoglycoprotein which is expressed in lymphatic endothelium but not in blood vessels (Yonemura et al., 2006). So, it could help us present the histological change in MD patients clearly.

Following, we will report one Chinese MD family with typical symptoms. Helping by the next-generation sequencing (NGS), immune-histo-chemical staining with D2-40 and hematoxylin-eosin (HE), we represent histological change of the lower-limb and the feet dorsum from the dead child patient (*premature death*, 28 weeks + 2 days) in this family, with the mutation in the *FLT4* (*c.3075G>A*, exon 22). Then, we compared the results with a normal child who unfortunately died of premature delivery also. The results offered us a new perspective to understand the occurrence and development of lymphedema in MD patients.

CASE PRESENTATION

All the operations and tests were given full authorization by the members from two families. And the study was full authorized by the ethic committee of the Air force medical university. All the operations were under supervision by one member of the families and one officer of the ethic committee. Written and informed consent was obtained from the guardians of the patient for publication of this case report.



Clinical Characterization of the Family (Figure 1)

Child Patient

Male, died of neonatal asphyxia post-delivery. The B-ultrasound scanner showed bilateral lower limb edema in regular pregnancy test, which was confirmed post-delivery. The skin color of the lower limbs was slightly purple (**Figure 2a**). The father (*proband*) developed edema in both lower limbs from birth, and had no other physical dysfunction. The mother denied any history of illness or medication during pregnancy. The grandparents (*paternal*) also had no physical dysfunction.

Mother

Gestational age 28 weeks + 2 days, premature delivery for car accident, number and structure of Chromosome showed nothing abnormal post amniocentesis.

Father (Proband)

The skin temperature was normal, there was no limitation of movement, and the muscle tension of the limbs was normal. There was no obvious abnormality in blood gas, blood biochemistry, blood routine, thyroid function, blood lead, urine routine, urine, and blood metabolism. Microvirus B19 DNA and nucleic acid test resulted negative. Quantitative detection of CMV- PP65 antigen and CMV- DNA was also negative.

The father was diagnosed as MD since birth and received plastic surgery, the symptoms of the left lower-limb was reduced while no significant improvement showed in the right. The father was also advised to order stretch socks to relieve symptoms (**Figures 2c,d**) (**Supplementary Video 1**).

Other Family Members (Paternal)

Aunt and grandparents showed nothing abnormal.

Clinical Characterization of the Matched Child

Mother

The mother was diagnosed of gastric carcinoma 25 weeks since menelipsis, and the family asked for termination



FIGURE 2 | (a) MD child patient, the skin color of the lower limbs was slightly purple, edema in both lower limbs could be detected; (b) matched child, no obvious abnormality; (c,d) father patient, edema in right lower limb could be detected; the left lower limb became normal post plastic surgery.

of pregnancy to save the mother. The gestational age was 26 weeks + 5 days, and the number and structure of Chromosome showed nothing abnormal post amniocentesis.

Matched Child

Male, died of labor induction. The structure and skin color of the lower limbs was normal, and no other physical dysfunction was detected (Figure 2b).

DESCRIPTION OF LABORATORY INVESTIGATIONS AND DIAGNOSTIC TESTS

D2-40 Staining and HE Staining

The tissue was gathered from the swollen situs of the lower-limb and the feet dorsum. Then we preserved them in 10%

formalin, paraffin-embedded after dehydration. Afterwards, we used the standard staining method to stain the tissue slice. We examined the sections under stereomicroscope (DMI6000 B, Leica Microsystems, Shanghai, China).

Results of D2-40 Staining (Figure 3)

1. The number of lymphatic vessels decreased in Milroy disease child.
2. The abnormal morphology and structure of lymphatic vessels could be detected in MD child.

Results of HE Staining

Venous abnormalities (Figure 4)

1. Capillary hyperemia is severe in Milroy disease child.
2. Phlebectasia could be detected in MD child.
3. Vascular malformations could be detected in Milroy disease child.

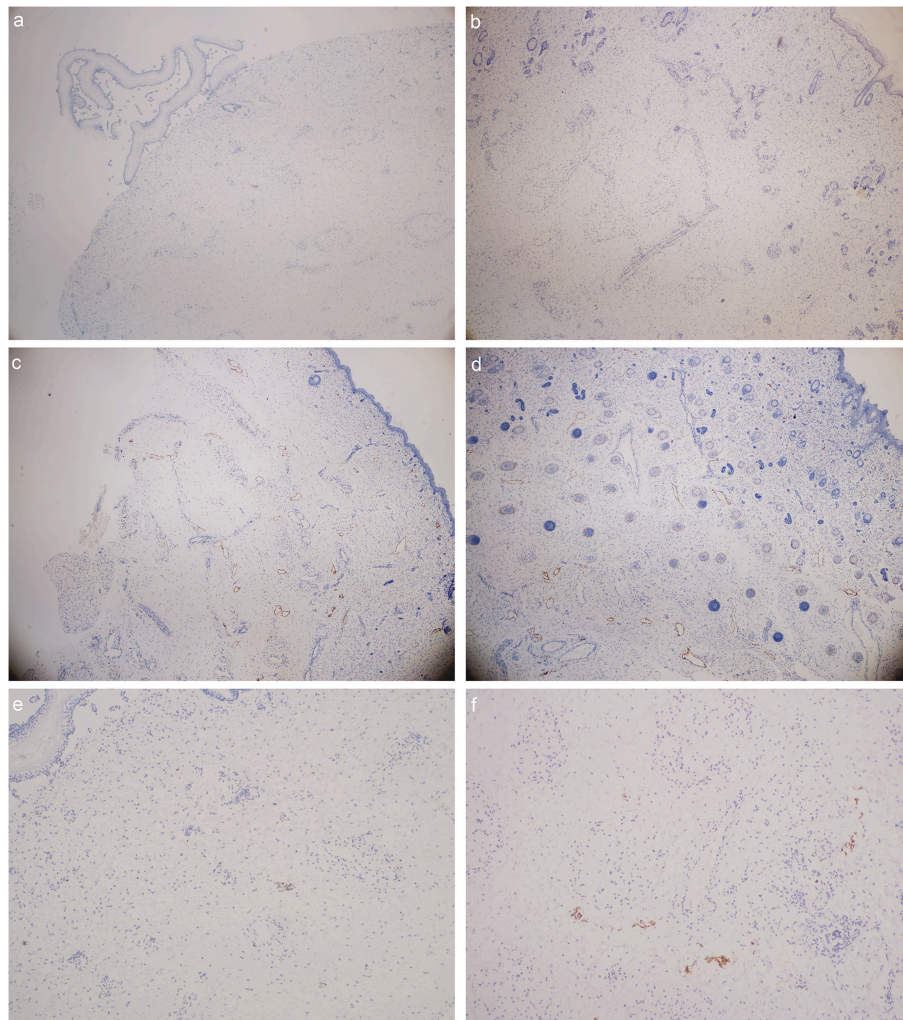


FIGURE 3 | D2-40 stain (lymphatic vessels were presented in orange color): **(a)** foot dorsum of MD child (40 times magnification); **(b)** lateral low-limb skin of MD child (40 times magnification); **(c)** foot dorsum of matched child (40 times magnification); **(d)** lateral low-limb skin of matched child (40 times magnification). **(e,f)** lymphatic vessels found within the field of vision **(e)** foot dorsum of MD child (100 times magnification); **(f)** lateral low-limb skin of MD child (100 times magnification).

4. The number of vascular endothelial cells and vascular smooth muscle cells decreased in MD child.

Skin abnormalities (Figure 4)

1. Large sheets of desquamated epidermis could be detected in MD child.
2. The numbers of cutaneous appendages reduced in MD child.

Genotyping by Next-Generation Sequencing (NGS), Sequence Analysis, and Cosegregation in the Family

Genomic DNA was isolated from peripheral blood samples from the family members using standard methods. NGS was applied to the MD child. Then, the other members in the family were verified by Sanger sequencing and cosegregation. The base pair

numbers of mutation sites were determined according to the GenBank mRNA reference sequences.

NGS revealed that the MD child has one heterozygous missense mutation in the FLT4: c.3075G>A (*p.M1025I*) in exon 22, a mutation which have not been described in Chinese ethnic. The result of the matched child showed normal in the FLT4 (Figure 5).

The mutation was subsequently confirmed by sanger sequencing of the family. The following cosegregation of the FLT4 alleles in the family pedigree confirmed that the c.3075G>A allele was paternal (II-2).

DISCUSSION

MD is characterized by heredity, painless, slow progression of disease, and it is limited to lower limb edema. MD could be diagnosed according to the clinical characteristics and genetic

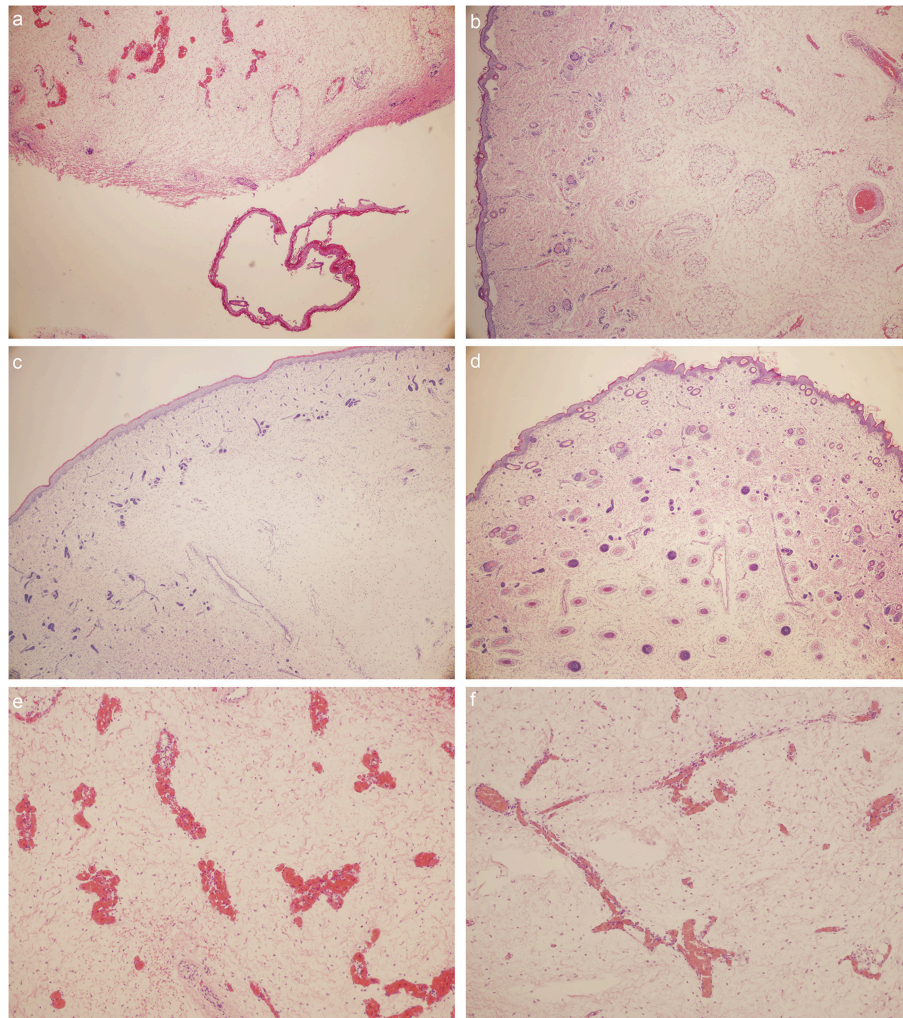


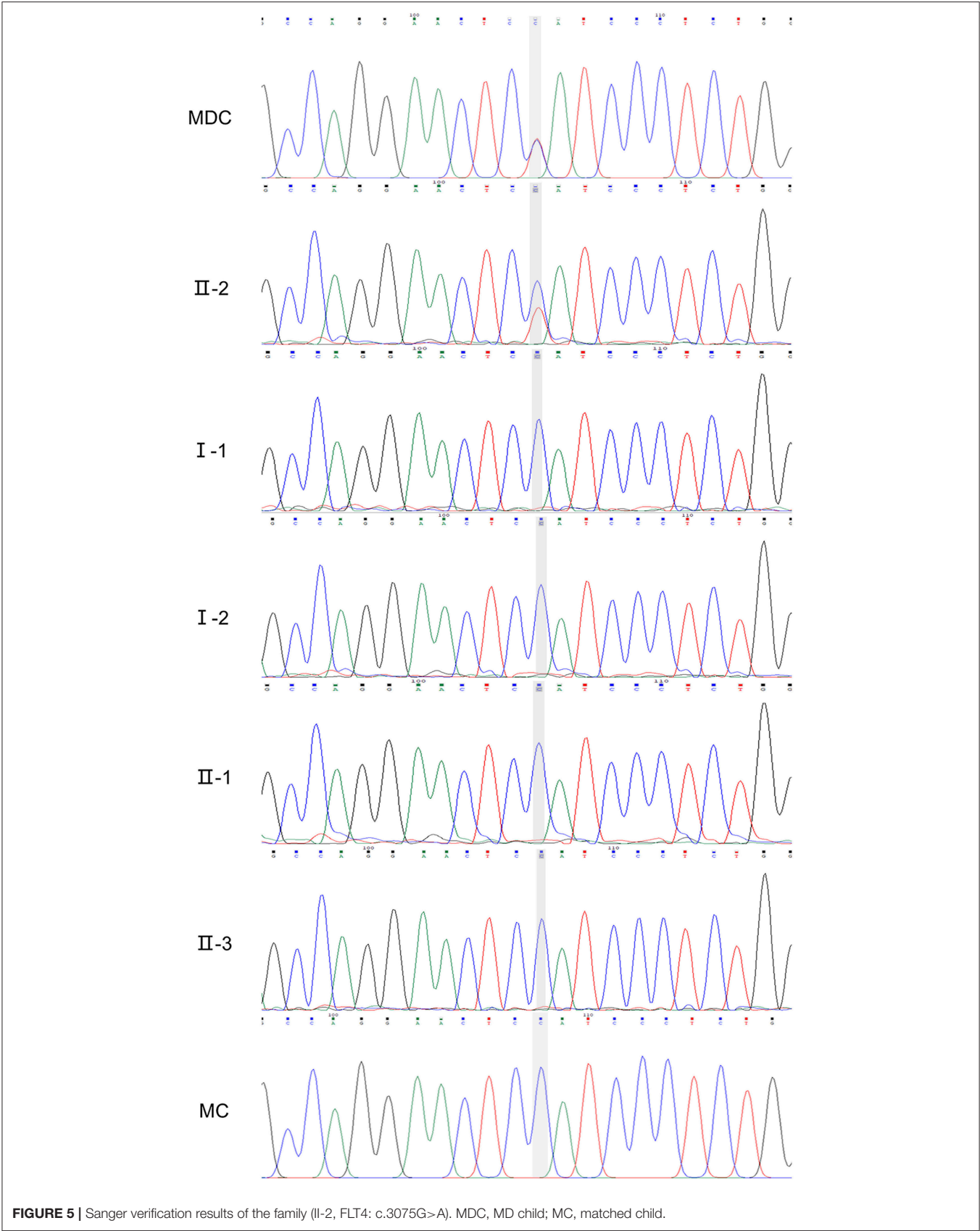
FIGURE 4 | HE stain: **(a)** foot dorsum of MD child (40 times magnification); **(b)** lateral low-limb skin of MD child (40 times magnification); **(c)** foot dorsum of matched child (40 times magnification); **(d)** lateral low-limb skin of matched child (40 times magnification). **(e,f)** lymphatic vessels found within the field of vision **(e)** foot dorsum of MD child (100 times magnification); **(f)** lateral low-limb skin of MD child (100 times magnification).

analysis. The edema usually occurs from (*or before*) birth. In neonates the swelling tends to affect primarily the dorsum of the feet. The amount of edema varies from individuals. Other features associated with MD sometimes include: Hydrocele (37% of males); Prominent veins (23%); Upslanting toenails (14%); Papillomatosis (10%); Urethral abnormalities in males (4%) (Brice et al., 1993).

The mutation of VEGFR-3 (*FLT4*) is believed responsible for the lymphedema in MD. The protein is a member of the tyrosine kinase receptor family, which plays an important role in lymphangiogenesis. In MD, there is believed an abnormal accumulation of interstitial protein-rich fluid caused by congenital malformation of the lymphatic vessels (Tammela and Alitalo, 2010; Gezginc et al., 2012). Similar results have been confirmed through lymphangiography that there is local lymphangiogenesis insufficiency in the extremities of MD children (Rooke, 2003).

For human, the development of the lymphatic vascular system begins in the sixth to seventh week of embryonic life. Malformation of the lymphatic vessels could trigger an increase of the interstitial protein rich fluid, which subsequently results in insufficient lymphatic transport and drainage (Kitsiou-Tzeli et al., 2010). As a result, large amount of protein-rich fluid accumulates in tissue interstitial spaces, which makes skin, subcutaneous tissue, fibrous tissue hyperplasia, and oppression of lymphatics more difficult for lymphatic reflux, thus forming a vicious cycle. The skin thickened, hardened, getting rough and bulky, forming “elephant skin” swelling over time.

In our findings, the numbers of the lymphatic vessels in MD child decreased. The morphology and structure of lymphatic vessels was abnormal. These phenomena confirmed the evidences given by previous animal and cell studies of the *FLT4* mutation (Rauniyar et al., 2018). But out of our expectation, the capillary and skin were also influenced in fact.



Capillary hyperemia and phlebectasia is severe in MD child. Vascular malformations could also be detected. And the number of vascular endothelial cells and vascular smooth muscle cells decreased in MD child. All the results indicated the mutation of FLT4 did not only affect the lymphatic formation, but also affect the blood vessels in human. Furthermore, large sheets of epidermis were desquamated. And the numbers of cutaneous appendages were also reduced. So concerning the connection between phenotype and genotype, we assume that mutation of FLT4 affect the angiogenesis and epidermis structure in human. While, this assumption still needs further evidence.

At present, radionuclide lymphatic imaging is the preferred method to observe the lymphatic system. Meanwhile, the b-ultrasonography, as a convenient and fast examination method, has still been widely used in the diagnosis and evaluation of lymphoedema in gravaida (Matter et al., 2002). Unfortunately, there is no effective treatment for lymphedema until now. Conservative treatment methods include roasting therapy, intermittent compression therapy and so on. If the edema and fibrosis is aggravated, surgical treatment might be required (Becker et al., 2012). And the key to treat MD is figuring out how the lymphedema come into being, and we still need more evidences.

REFERENCES

- Bai, B., Ma, W., Wang, K., Ha, S., Wang, J. B., Tan, B. X., et al. (2013). Detection of D2-40 monoclonal antibody-labeled lymphatic vessel invasion in esophageal squamous cell carcinoma and its clinicopathologic significance. *Cancer Biol. Med.* 10, 81–85. doi: 10.7497/j.issn.2095-3941.2013.02.003
- Becker, C., Arrive, L., Saaristo, A., Germain, M., Fanzio, P., Batista, B. N., et al. (2012). Surgical treatment of congenital lymphedema. *Clin. Plast. Surg.* 39, 377–384. doi: 10.1016/j.cps.2012.08.001
- Brice, G., Child, A. H., Evans, A., Bell, R., Mansour, S., Burnand, K., et al. (2005). Milroy disease and the VEGFR-3 mutation phenotype. *J. Med. Genet.* 42, 98–102. doi: 10.1136/jmg.2004.024802
- Brice, G. W., Mansour, S., Ostergaard, P., Connell, F., Jeffery, S., and Mortimer, P. (1993). "Milroy disease," in *GeneReviews* (R), eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, and A. Amemiya (Seattle, WA: University of Washington).
- Connell, F. C., Ostergaard, P., Carver, C., Brice, G., Williams, N., Mansour, S., et al. (2009). Analysis of the coding regions of VEGFR3 and VEGFC in milroy disease and other primary lymphoedemas. *Hum. Genet.* 124, 625–631. doi: 10.1007/s00439-008-0586-5
- Evans, A. L., Brice, G., Sotirova, V., Mortimer, P., Beninson, J., Burnand, K., et al. (1999). Mapping of primary congenital lymphedema to the 5q35.3 region. *Am. J. Hum. Genet.* 64, 547–555. doi: 10.1086/302248
- Ferrell, R. E., Levinson, K. L., Esmen, J. H., Kimak, M. A., Lawrence, E. C., Barmada, M. M., et al. (1998). Hereditary lymphedema: evidence for linkage and genetic heterogeneity. *Hum. Mol. Genet.* 7, 2073–2078.
- Gezginc, K., Yazici, F., and Gok, D. (2012). Prenatal diagnosis of milroy's primary congenital lymphedema. *Taiwan. J. Obstet. Gynecol.* 51, 280–282. doi: 10.1016/j.tjog.2011.01.041
- Gordon, K., Spiden, S. L., Connell, F. C., Brice, G., Cottrell, S., Short, J., et al. (2013). FLT4/VEGFR3 and milroy disease: novel mutations, a review of published variants and database update. *Hum. Mutat.* 34, 23–31. doi: 10.1002/humu.22223

AUTHOR CONTRIBUTIONS

SZ and YW planned the study. SZ and XC conducted the surgery. SZ and LY conducted the histological staining. SZ and SW wrote the article. DM organized the photographs. SL and YW supervised the study.

FUNDING

This study was supported by National Natural Science Foundation of China (No. 81671476).

ACKNOWLEDGMENTS

The authors thank Guangwei Shi, Guangyuan Shi, Li Wang, Jisen Shi, Ziyun Yang, Rui Li, and Kun Chen for their understanding, supporting, and valuable sacrifice in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00206/full#supplementary-material>

Supplementary Video 1 | The edema does not affect normal walking to the father patient.

- Kitsiou-Tzeli, S., Vrettou, C., Leze, E., Makrythanasis, P., Kanavakis, E., and Willems, P. (2010). Milroy's primary congenital lymphedema in a male infant and review of the literature. *In vivo* 24, 309–314.
- Matter, D., Grosshans, E., Muller, J., Furderer, C., Mathelin, C., Warter, S., et al. (2002). [Sonographic imaging of lymphatic vessels compared to other methods]. *J. Radiol.* 83, 599–609. doi: 10.1097/00004424-200205000-00008
- Mellor, R. H., Hubert, C. E., Stanton, A. W., Tate, N., Akhras, V., Smith, A., et al. (2010). Lymphatic dysfunction, not aplasia, underlies milroy disease. *Microcirculation*. 17, 281–296. doi: 10.1111/j.1549-8719.2010.00030.x
- Milroy, W. F. (1892). An undescribed variety of hereditary edema. *New York Med. J.* 56, 505–508.
- Rauniyar, K., Jha, S. K., and Jeltsch, M. (2018). Biology of vascular endothelial growth factor C in the morphogenesis of lymphatic vessels. *Front. Bioeng. Biotechnol.* 6:7. doi: 10.3389/fbioe.2018.00007
- Rooke, T. (2003). Diseases of the lymphatics: Sir Norman Browse, Kevin, G. Burnand, and Peter, S. Mortimer; London, England; Arnold; 336 pages; £198.5. *J. Vasc. Surg.* 38:1145. doi: 10.1016/S0741-5214(03)01026-7
- Tammela, T., and Alitalo, K. (2010). Lymphangiogenesis: molecular mechanisms and future promise. *Cell* 140, 460–476. doi: 10.1016/j.cell.2010.01.045
- Yonemura, Y., Endou, Y., Tabachi, K., Kawamura, T., Yun, H. Y., Kameya, T., et al. (2006). Evaluation of lymphatic invasion in primary gastric cancer by a new monoclonal antibody, D2-40. *Hum. Pathol.* 37, 1193–1199. doi: 10.1016/j.humpath.2006.04.014

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhang, Chen, Yuan, Wang, Moli, Liu and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comparative Analysis for the Performance of Variant Calling Pipelines on Detecting the *de novo* Mutations in Humans

Yu Liang^{1†}, Li He^{2†}, Yiru Zhao³, Yinyi Hao¹, Yifan Zhou¹, Menglong Li¹, Chuan Li³, Xuemei Pu^{1*} and Zhining Wen^{1*}

¹ College of Chemistry, Sichuan University, Chengdu, China, ² Biogas Appliance Quality Supervision and Inspection Center, Biogas Institute of Ministry of Agriculture, Chengdu, China, ³ College of Computer Science, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:

Zhichao Liu,
The National Center for Toxicological
Research (FDA), United States

Reviewed by:

Dan Li,
University of Arkansas at Little Rock,
United States
Arun Samidurai,
Virginia Commonwealth University,
United States

*Correspondence:

Xuemei Pu
xmpuscu@scu.edu.cn
Zhining Wen
w_zhining@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 11 December 2018

Accepted: 21 March 2019

Published: 11 April 2019

Citation:

Liang Y, He L, Zhao Y, Hao Y,
Zhou Y, Li M, Li C, Pu X and Wen Z
(2019) Comparative Analysis
for the Performance of Variant Calling
Pipelines on Detecting the *de novo*
Mutations in Humans.
Front. Pharmacol. 10:358.
doi: 10.3389/fphar.2019.00358

Despite of the low occurrence rate in the entire genomes, *de novo* mutation is proved to be deleterious and will lead to severe genetic diseases via impacting on the gene function. Considering the fact that the traditional family based linkage approaches and the genome-wide association studies are unsuitable for identifying the *de novo* mutations, in recent years, several pipelines have been proposed to detect them based on the whole-genome or whole-exome sequencing data and were used for calling them in the rare diseases. However, how the performance of these variant calling pipelines on detecting the *de novo* mutations is still unexplored. For the purpose of facilitating the appropriate choice of the pipelines and reducing the false positive rate, in this study, we thoroughly evaluated the performance of the commonly used trio calling methods on the detection of the *de novo* single-nucleotide variants (DNSNVs) by conducting a comparative analysis for the calling results. Our results exhibited that different pipelines have a specific tendency to detect the DNSNVs in the genomic regions with different GC contents. Additionally, to refine the calling results for a single pipeline, our proposed filter achieved satisfied results, indicating that the read coverage at the mutation positions can be used as an effective index to identify the high-confidence DNSNVs. Our findings should be good support for the committees to choose an appropriate way to explore the *de novo* mutations for the rare diseases.

Keywords: *de novo* mutation, rare diseases, variant calling pipelines evaluation, gene function, whole-exon sequencing

INTRODUCTION

The genomic structural variations, such as single-nucleotide variants (SNVs), copy-number variants (CNV) and the indels, play important roles in the genetic diseases. The researches in the past decade have discovered the landscape of SNVs in human and the strong causality between the SNVs and the genetic diseases (Ku et al., 2012; Veltman and Brunner, 2012; Boycott et al., 2013). Among the SNVs, the occurrence frequency of *de novo* SNVs (DNSNVs) in germline is as

Abbreviations: AJ, ashkenazi jewish; CNV, copy-number variants; DNSNVs, *de novo* single-nucleotide variants; GATK, genomeanalysis toolkit; GIAB, genome in a bottle; OMIM, online mendelian inheritance in man; PCR, polymerase chain reaction; SNVs, single-nucleotide variants; Ti/Tv, the transition/transversion ratio; WES, whole-exome sequencing; WGS, whole-genomes sequencing.

low as $1.0\sim 3.0 \times 10^{-8}$ SNVs per site per generation (Conrad et al., 2012; Veltman and Brunner, 2012), but this type of mutation is proved to be deleterious and will lead to severe genetic diseases via impacting on the different gene functions. It had been reported in recent studies that the rare sporadic malformation syndromes (Hoischen et al., 2010a,b; Ng et al., 2010) as well as the neurodevelopmental diseases (Hamdan et al., 2014; Turner et al., 2017) were primarily caused by the DNSNVs in single specific genes or a set of genes, elaborating the fundamentality of the *de novo* mutations in the genetic diseases despite of the unclear underlying mechanisms. Therefore, accurately identifying the *de novo* mutations located in the rare-disease-causing genes can be great helpful not only for improving the clinical diagnostics, but also for better understanding the mechanisms in the rare genetic diseases.

Due to the fact that the traditional family based linkage approaches and the genome-wide association studies were unsuited to the detection of *de novo* mutations, the emerging next-generation sequencing technologies such as the whole-genomes sequencing (WGS)/whole-exome sequencing (WES) began to be applied in the researches of genetic diseases (Boycott et al., 2013; Peters et al., 2015; Jin et al., 2017; Turner et al., 2017; Hyrenius-Wittsten et al., 2018). A number of bioinformatics pipelines have subsequently been proposed to call the *de novo* mutations based on the WGS/WES data (McKenna et al., 2010; Li et al., 2012; Koboldt et al., 2013; Kojima et al., 2013; Peng et al., 2013; Ramu et al., 2013; Cleary et al., 2014; Salzberg et al., 2014; Santoni et al., 2014; He et al., 2015; Wei et al., 2015; Francioli et al., 2017; Gomez-Romero et al., 2018; Zhou et al., 2018). The heated discussions have been carried on in recent years about applying these approaches in the diagnostics of rare diseases and the potential clinical implementations (Yang et al., 2013; Lee et al., 2014; Jamuar and Tan, 2015; Bacchelli and Williams, 2016; Krier et al., 2016; Thiffault and Lantos, 2016). However, the occurrence of germline *de novo* mutations is much lower than that of the inherited variations, resulting in the difficulty of discriminating these variants from the noise derived from the procedures of sequencing, reads mapping as well as the variant calling and annotation. So far how the performance of these variant calling pipelines on detecting the *de novo* mutations is still unexplored.

Therefore, in our study, we thoroughly investigated three commonly used trio calling pipelines named GATK (McKenna et al., 2010; Francioli et al., 2017), RTG (Cleary et al., 2014) and VarScan (Koboldt et al., 2013) on the detection of the DNSNVs and found that GATK can detect the DNSNVs in the low GC-content region with a relative low error rate while RTG and VarScan are more suitable for detecting the DNSNVs in the high GC-content region. In refining the calling results of a single pipeline, our proposed filter not only effectively excluded the redundant DNSNVs, but also ensured the transitions/transversions ratio of the results, indicating that the read coverage at the mutation positions of the son's genome and the parents' genomes can be an important index for evaluating the quality of DNSNVs.

MATERIALS AND METHODS

Dataset

The WES data of the Ashkenazi Jewish (AJ) trio set (NA12878) were applied in our study (Zook et al., 2014, 2016). The preprocessed BAM files of the mother (HG004), the father (HG003) and the son (HG002), for which the data preprocessing steps including the reads alignment and duplicates marking had been conducted beforehand, were directly downloaded from the Genome in a Bottle (GIAB) website¹ and used for calling the *de novo* mutations. In order to facilitate the comparison of the similarities and differences of the pipelines and make the point clear, in our study, we only focused on the *de novo* SNVs on the autosomes. It is worth noting that the structure variations, e.g., indels, as well as the mutations on the chromosomes X and Y are also important for the genetic diseases, but it will make the problem more complicated when involving them in the comparisons.

Trio Calling Pipelines

Three pipelines, namely GenomeAnalysisToolKit (version 4.0.5.2) (McKenna et al., 2010; Francioli et al., 2017), RTG (non-commercial version 3.9.1) (Cleary et al., 2014) and VarScan (version 2.3.9) (Koboldt et al., 2013), were applied in this study to call the DNSNVs. In GATK pipeline, the gVCF files for the trio samples were firstly generated by using *HaplotypeCaller* separately and combined into a multi-sample gVCF file through *CombineGVCFs*. Then, the raw SNVs were called by *GenotypeGVCFs* and recalibrated by *VariantRecalibrator* and *ApplyVQSR*. Finally, after deriving the posterior probabilities of the genotypes by *CalculateGenotypePosteriors*, the low quality genotypes were filtered out by *VariantFiltration* and the *de novo* SNVs were annotated by *VariantAnnotator*. In RTG pipeline, the quality calibration files for the BAM files of trio samples were firstly prepared and applied in the subsequent calling procedures. The *de novo* SNVs were called and extracted by RTG with two parameters of *family* and *vcfilter*. As to the VarScan pipeline, the BAM files of trio samples were firstly sorted and combined into a three-sample pileup file. Subsequently, the *de novo* SNVs were called by running the *VarScan trio* command.

Metrics for the Comparison of the Pipelines

We used three metrics namely GC content (Shin et al., 2013), substitution type (Zook et al., 2014) and SNV density (Choi et al., 2018), to evaluate the differences in the calling results generated by three pipelines. A genomic sequence of 100 bases centered on a DNSNV was extracted from the reference genome and the GC content can be calculated via the following equ.

$$\text{GC content (\%)} = \frac{\text{number of bases G and C}}{100} \times 100\% \quad (1)$$

The substitution type includes the point mutations of transition and transversion. The transition refers to the nucleotide changes

¹ftp://ftp-trace.ncbi.nlm.nih.gov/giab/ftp/data/AshkenazimTrio/

from a purine to another purine (A↔G) or a pyrimidine to another pyrimidine (C↔T), which occurs more frequently than the transversion in the SNPs. The transversion refers to the changes from purine to pyrimidine, or vice versa (A↔T, A↔C, G↔T, and C↔G). By counting the total number of SNVs in the genomic sequence of 100 bases centered on a DNSNV, the SNV density can be calculated via the equ. (2):

$$\text{SNV density (\%)} = \frac{\text{number of SNVs}}{100} \times 100\% \quad (2)$$

Definition of DNSNV Filter

To distinguish the high-confidence DNSNVs from the noise, the read coverage at the mutation point is an important indicator for this purpose. Based on the concept of signal-to-noise ratio, we not only expect that the number of reads mapped to the reference genome at the mutation point for the parents' datasets is high than that mapped to the mutated sequence as much as possible, but also expect that the number of reads mapped to the mutated sequence for the son's dataset is higher than that mapped to the reference genome. Therefore, to refine the calling results from the pipelines, we proposed a DNSNV filter by considering both read coverage at the mutation point in the son's dataset and the parents' datasets. The filter was defined as the following equ:

$$\text{Score} = \log \left(\frac{|F_{\text{ref}} - F_{\text{alt}}| + |M_{\text{ref}} - M_{\text{alt}}|}{2 \times |S_{\text{alt}} - S_{\text{ref}}|} \right) \quad (3)$$

Where the F_{ref} , M_{ref} , and S_{ref} indicate the number of reads mapped to the reference genome at the i th mutation point for the datasets of father, mother and son, respectively. The F_{alt} , M_{alt} , and S_{alt} indicate the number of reads mapped to the mutated sequence at the i th mutation point for the datasets of father, mother and son, respectively. We assigned the scores for the DNSNVs and rank them by their scores. In this study, we took a non-stringent score (score = 0) as the threshold to investigate the improvement of the calling results.

Genotype Quality

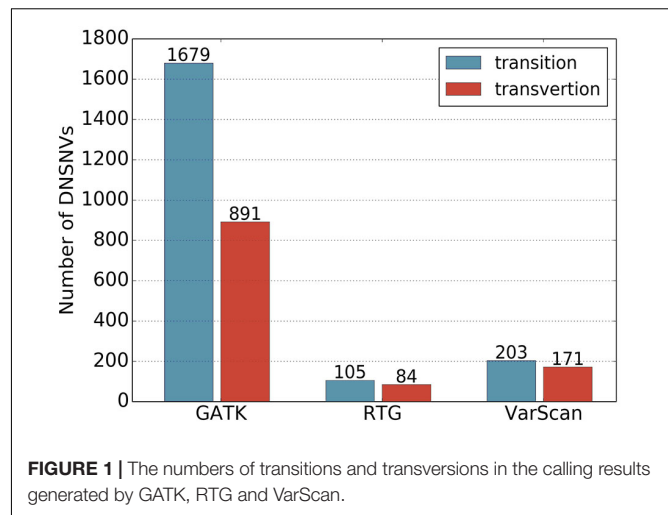
Genotype Quality (GQ) (Zhang et al., 2013) is used to evaluate the filtering results of DNSNVs, which indicates the quality value of the most likely genotype. The quality value refers the possibility of the genotype being present at the site. The larger value means the greater the likelihood of the genotype.

RESULTS

To facilitate the appropriate choice of the trio calling pipelines for detecting the DNSNVs, in our study, we firstly evaluated the results of three commonly used pipelines named GATK, RTG, and VarScan by using the WES data of a father-mother-child trio set from Ashkenazi Jews, and then proposed a filter for removing the redundant DNSNVs from the calling results.

The Landscape of the Calling Results From Three Pipelines

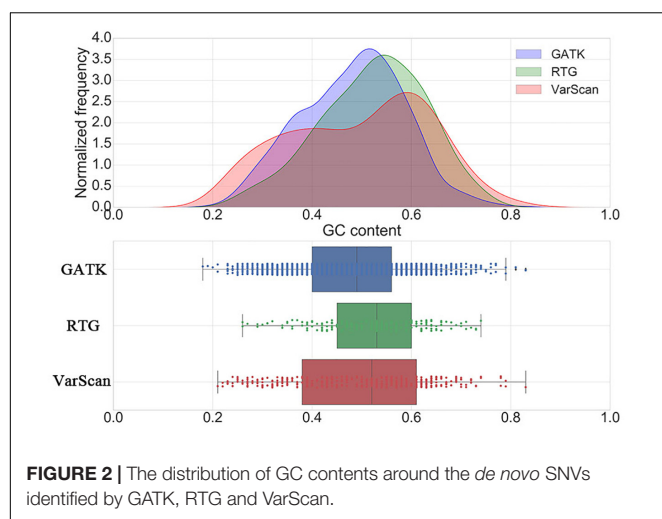
In total, 2570, 189 and 374 DNSNVs in the autosomes were identified by GATK, RTG and VarScan, respectively. In



terms of the quantity of DNSNVs, GATK exhibited higher detecting sensitivity than RTG and VarScan. For the point of mutation types, **Figure 1** showed clear difference between the numbers of transitions and transversions in the calling results of three pipelines. Generally the number of transitions was greater than that of the transversions for all the pipelines. The transition/transversion ratio (Ti/Tv) had been suggested in GIAB study (Zook et al., 2014) to be a metric for evaluating the quality of the novel calling variants based on the assumption that Ti/Tv in novel variants would be similar to that in the common variants. In our results, Ti/Tv for GATK was 1.88 (1679/891), which was higher than those for the other two pipelines [1.25 (105/84) and 1.19 (203/171) for RTG and VarScan, respectively], indicating a lower error rate in the GATK calling results. In addition, we mapped the DNSNVs to the dbSNP database² and separately calculated the overlap rates between the DNSNVs and all the variations, as well as between the DNSNVs and the common variations for the three pipelines (**Supplementary Figure S1**). The overlap rate was calculated by dividing the number of overlapped DNSNVs by the total number of DNSNVs identified by the pipeline. From the figure we can see that GATK achieved the highest overlap rate among three pipelines, indicating the least proportion of DNSNVs involved in the calling results. On the contrary, VarScan achieved the lowest overlap rate, indicating the largest proportion of DNSNVs involved in the calling results. Meanwhile, the Ti/Tv rates of three pipelines indicated the lowest error rate in the calling results of GATK and highest error rate in the results of VarScan. It suggested that VarScan and RTG tend to reveal more DNSNVs in their reported results, but the reliability of them needs to be more carefully validated by further experiments.

Considering the fact that with the increase of GC content, the difficulty of polymerase chain reaction (PCR) amplification in sequencing procedure will increase, which will result in an increase in the error rate of the calling results, we further investigated the distribution of GC contents around the DNSNVs

²<https://www.ncbi.nlm.nih.gov/snp/>



for three pipelines (**Figure 2**). The majority of the DNSNVs called by GATK located in the regions with GC contents less than 50% while the DNSNVs called by RTG mainly located in the regions with relatively high GC contents (> 50%). This may be the reason that GATK can yield the higher Ti/Tv ratio (1.88) than that (1.25) obtained by RTG. Interestingly, although VarScan yielded a relatively low Ti/Tv ratio (1.19), the distribution of the calling results from VarScan had two peaks that covered the regions with both low (~40%) and high (~60%) GC content. We separately inspected the point mutation types of the DNSNVs called by VarScan in the genome regions with GC content $\geq 50\%$ and GC content $< 50\%$. The Ti/Tv ratios for the high ($\geq 50\%$) and low ($< 50\%$) GC-content regions were 1.32 (116/88) and 1.05 (87/83), respectively. It may indicate that, for the high GC-content region, the VarScan can identify the DNSNVs with even lower error rate than RTG (Ti/Tv = 1.25). As to calling the DNSNVs in the low GC-content region, the performance of VarScan (Ti/Tv = 1.05) is inferior to GATK (Ti/Tv = 1.88).

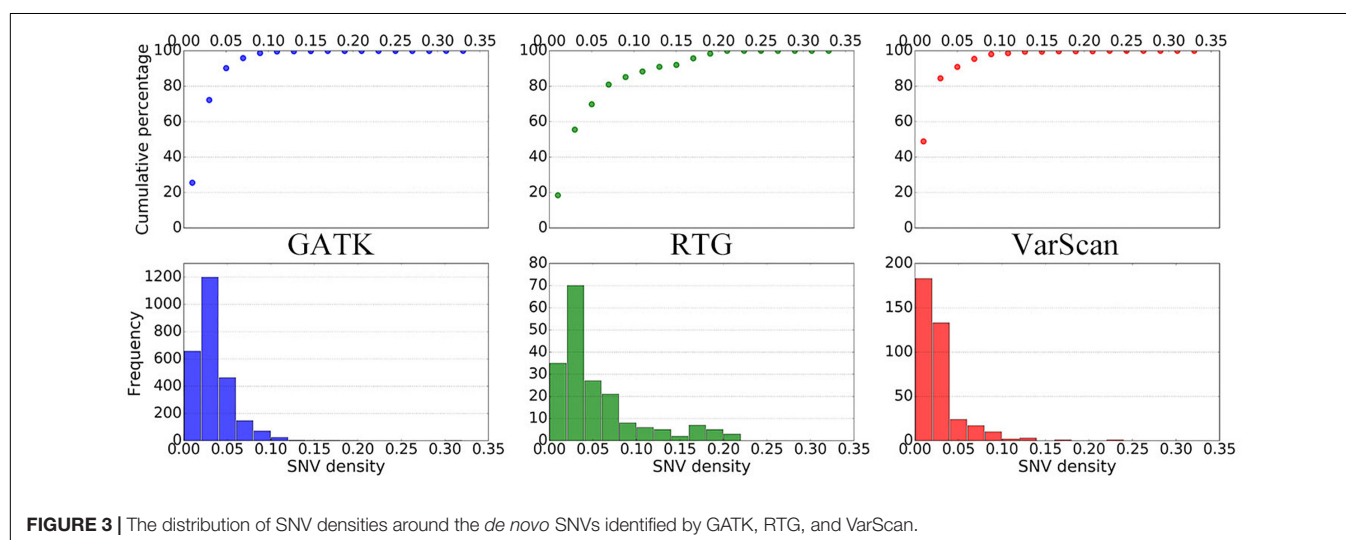
To further elucidate the performance of the three pipelines, we subsequently summarized the distributions of SNV densities

around the DNSNVs (**Figure 3**). As the SNV density increases, the error rate for calling DNSNVs may increase. It can be seen from the figure that, for the RTG pipeline, only ~70% DNSNVs located in the regions with the SNV density less than 5%, and the accumulated percentage gradually increased to 90% when counting the DNSNVs in the regions with the SNV density less than 15%, indicating a larger error rate may exist in the call results. Compared to RTG, over 90% DNSNVs called by GATK and VarScan located in the regions with the SNV density less than 5%, which may suggest the better quality of the identified DNSNVs.

Performance of Our Proposed Filter on Refining the DNSNVs

To obtain the high-confidence DNSNVs, we suggested further refining the calling results of three pipelines by using a proposed filter, which was taken both read coverage at the mutation sites of the son's genome and the parents' genomes into consideration to identify the DNSNVs and can be directly applied to an individual pipeline without reference to the information of other pipelines. **Figure 4** showed the numbers of DNSNVs kept by the filter when applying different cut-offs. The number of DNSNVs called by all the pipelines took on a tendency of descension when the cut-off became more stringent, especially for the GATK pipeline. The number of the left DNSNVs called by GATK dramatically decreased from over 1,500 to 15 as the value of the cut-off increased from -3 to 3.

In this study, we just took a non-stringent cut-off (score = 0) as an example for the comparison of the three pipelines. When filtering the DNSNVs with the score > 0 , the numbers of DNSNVs detected by GATK decreased from 2570 to 630. For RTG and VarScan, the numbers of DNSNVs decreased from 189 and 374 to 124 and 350, respectively. **Figure 5A** showed the number of transitions and transversions in the filtered calling results. The GATK pipeline still yielded the highest ratio (Ti/Tv = 1.96) among three pipelines, which was improved after filtering a number of DNSNVs from the calling results. For the pipelines of RTG and VarScan, since only a small number



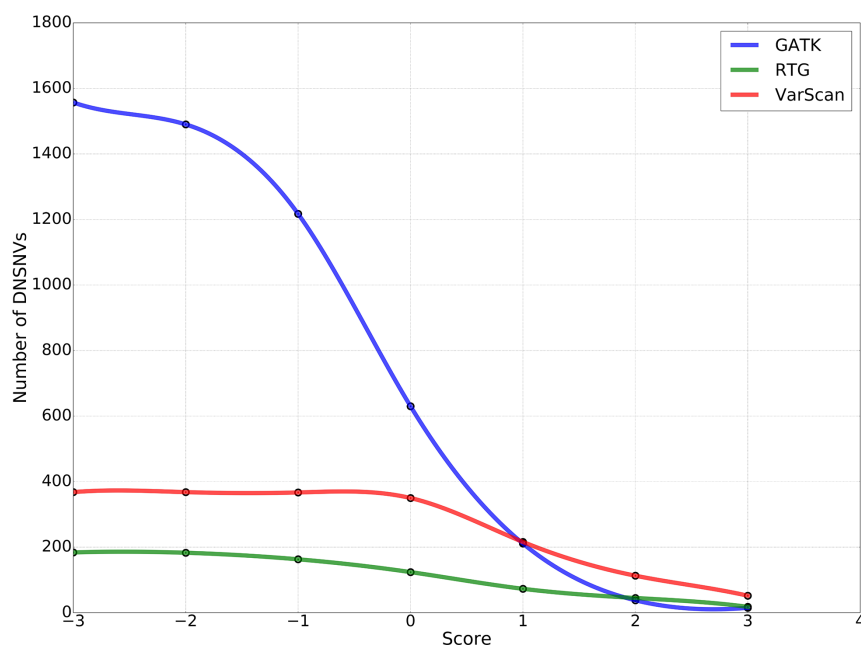


FIGURE 4 | The number of *de novo* SNVs kept in the calling results when filtered by using different cut-offs in our proposed filter.

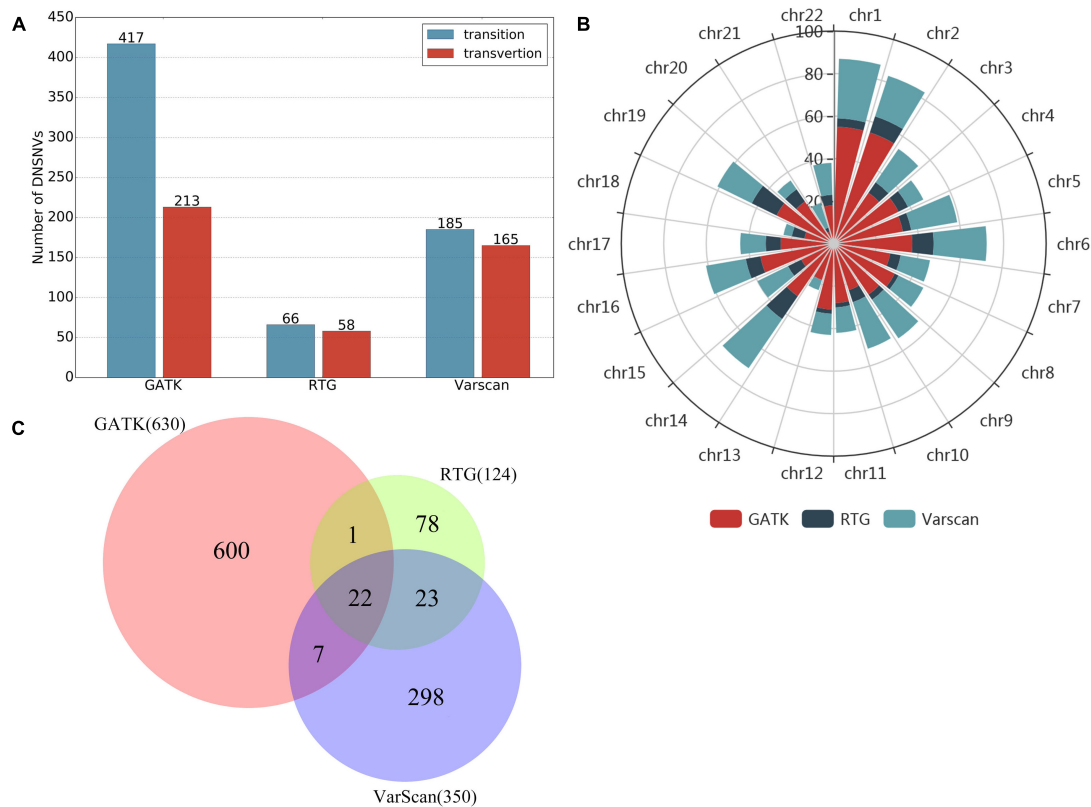
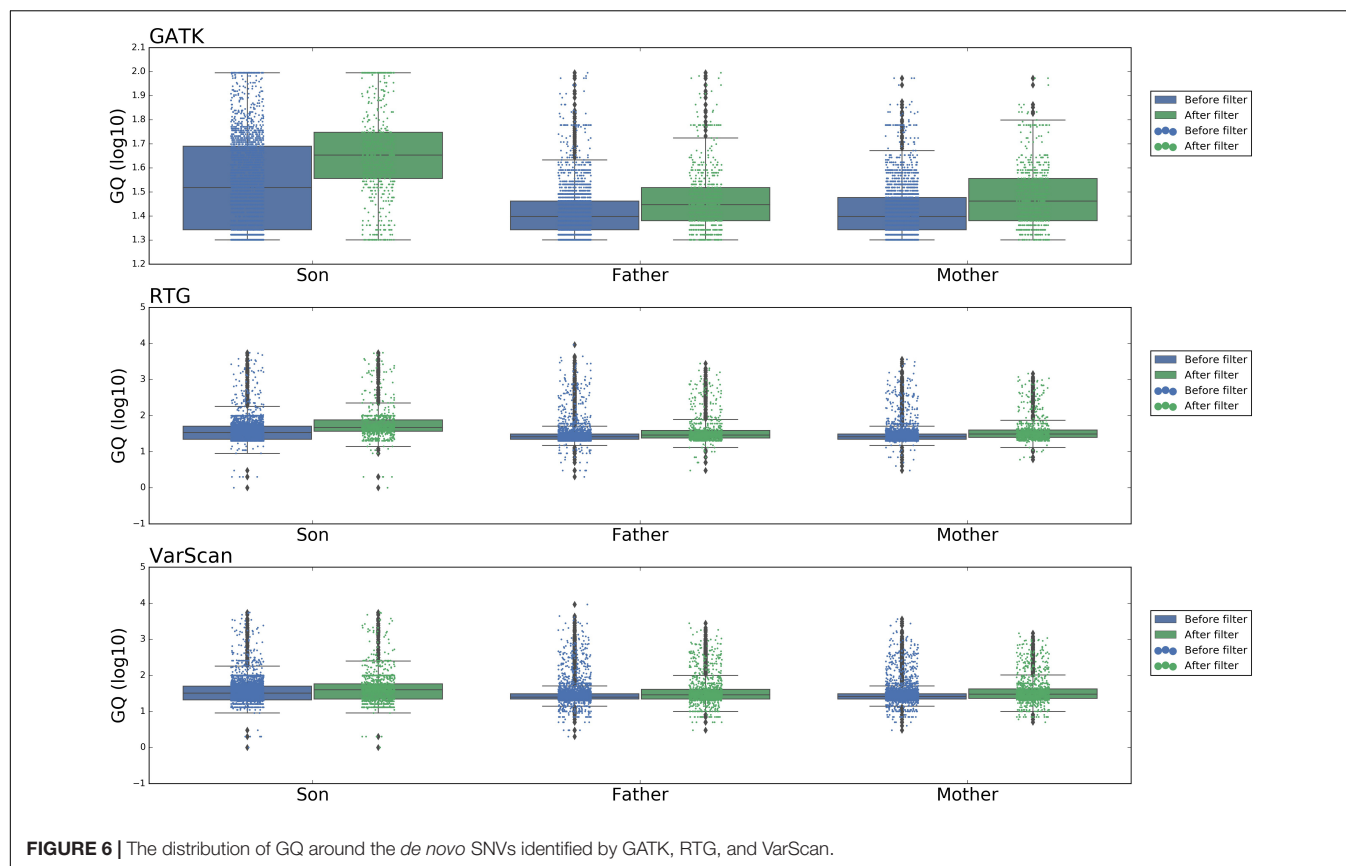


FIGURE 5 | Results on the filtered *de novo* SNVs identified by GATK, RTG and VarScan. **(A)** The number of transitions and transversions in the filtered calling results. **(B)** The chromosomal distribution of the filtered *de novo* SNVs. **(C)** The overlaps of the filtered *de novo* SNVs among three pipelines.



of DNSNVs were removed, the quality of final DNSNVs were comparable to that of the original calling results, with a slight decrease in the ratio ($Ti/Tv = 1.14$ and 1.12 for RTG and VarScan, respectively). Our results may suggest that our proposed filter can efficiently reduce the error rate in the DNSNVs from the redundant calling results. In addition, for the calling results with less redundancy, our filter can well maintain the quality of the original results.

The chromosomal distribution of the filtered DNSNVs was shown in **Figure 5B**. The DNSNVs identified by three calling pipelines mainly located on the chromosomes 1, 2, 6, 14, 16, and 19, which indicates that the DNSNVs are more likely to occur on these six chromosomes. Conversely, the DNSNVs rarely occur on chromosomes 13, 18, and 21. **Figure 5C** showed the overlaps of the filtered DNSNVs among three pipelines. Because both RTG and VarScan were adapted for detecting the DNSNVs in the high GC-content region with relative low error rate, close to 36% (45/124) DNSNVs identified by RTG can be found in the results called by VarScan even if 65 out of 189 DNSNVs identified by RTG had been removed out by the filter. When comparing GATK with RTG and VarScan, only about 19% (23/124) and 8% (29/350) DNSNVs called by RTG and VarScan, respectively, were involved in the calling results of GATK. The reason may be that GATK was fit for calling the DNSNVs in the low GC-content region while the ability of RTG and VarScan to identify the DNSNVs in this region is inferior. This should be further validated by using the real disease

samples. Eventually, a total of 22 DNSNVs were detected by all the pipelines.

Figure 6 showed the distribution of DNSNVs GQ of the trio before and after filtering. For GATK, significant difference exists between the calling results before and after filtering. However, for RTG and VarScan, the difference is slight. But for the latter two pipelines, the filtered DNSNVs GQ distribution is still slightly better than that before filtering. In general, the GQ distributions of DNSNVs for the three pipelines after filtering are better than those before filtering.

Biological Relevance of the Overlapped DNSNVs Among Three Pipelines

The 22 overlapped DNSNVs can be detected by three pipelines simultaneously, and should be of relatively high confidence. To investigate the biological relevance of these DNSNVs, we firstly mapped them to the corresponding genes and then explored the associations of the genes with genetic diseases by the Online Mendelian Inheritance in Man (OMIM) database searching and literature survey. The description of 22 DNSNVs as well as the corresponding genes finally identified by all pipelines was listed in **Table 1**. In the results, six out of 22 genes has been proved to be directly associated with the rare genetic diseases. For example, the gene *DTNB* in chromosome 2 is an important protein-coding gene of beta-dystrobrevin. This protein is found to interact directly with dystrophin and the low expression level of it will

TABLE 1 | The 22 overlapped DNSNVs identified by all the trio calling methods and the corresponding genes associated with the diseases.

Chromosome	Position	Substitution	Gene symbol	phenotype
2	25457155	C→T	DTNB*	Muscular dystrophy
2	197791183	C→T	—	—
2	241835203	G→T	—	—
3	4669342	A→T	ITPR1*	Gillespie syndrome Spinocerebellar ataxia 15 Spinocerebellar ataxia 29, congenital non-progressive
3	48603870	G→A	UQCRC1	Predisposition of Alzheimer's disease
3	52547912	C→G	PBRM1*	Clear cell renal cell carcinoma
6	109740508	T→C	FIG4*	Polymicrogyria, bilateral temporooccipital Amyotrophic lateral sclerosis 11 Charcot-Marie-Tooth disease, type 4J Yunis-Varon syndrome
5	140730969	A→G	—	—
8	30703949	C→G	GSR*	Hemolytic anemia due to glutathione reductase deficiency
9	131041047	G→T	LAMC3*	Cortical malformations, occipital
9	131851319	C→G	—	—
11	82698724	T→C	—	—
14	77245340	C→T	TMEM63C	—
14	93170535	G→A	—	—
14	96180196	C→G	—	—
14	106330036	T→A	—	—
14	106993919	A→G	—	—
15	79852462	G→T	MTHFS	Heart defects; lung cancer
17	53638886	G→A	—	—
19	45720935	C→T	FBXO46	—
22	23029544	A→C	—	—
22	23029596	A→T	—	—

*The genes were reported to be directly associated with genetic diseases in the OMIM database.

cause severe Duchenne muscular dystrophy (Blake et al., 1998). The genes *FIG4* and *LAMC3* were reported to be highly correlated with the polymicrogyria and cortical malformations, respectively. Campeau et al. (2013) demonstrated that the inactivation of *FIG4* would result in the central nervous system dysfunction and extensive skeletal anomalies. The research by Barak et al. (2011) exhibited an important role of the gene *LAMC3* in cortical organization. It can be seen that to a certain extent our proposed filter can be helpful for removing the redundant DNSNVs from the calling results.

DISCUSSION

Nowadays, it has been proved that the *de novo* mutations played an important role in human genetic diseases. Based on the next-generation sequencing technology, more and more researches focused on detecting the *de novo* mutations in the rare genetic diseases as well as the potential applications in the clinics for the purpose of improving the clinical diagnosis and better understanding the mechanisms of the genetic diseases. However, it is still a tough work and remains challenging to accurately identify the genomic variants because of the complexity of the sequencing experiments and the variants calling procedures. In

the previous study, Reumers et al. (2012) suggested that an optimized filtering procedure would be helpful for reducing the error rate when detecting the genomic variants with the short-read sequencing data. Considering the fact that the occurrence rate of *de novo* mutations is much lower than those of inherited variations and somatic variants, it is more difficult to distinguish them from the errors in the whole genome. At present, a number of trio calling methods were proposed to detect the *de novo* mutations based on the WGS/WES data, but how the performance of these pipelines on detecting the *de novo* mutations is still unexplored. Therefore, by carefully comparing the results from three commonly used trio calling pipelines, we elucidated that the performance of the three pipelines on calling DNSNVs in high or low GC-content region was different. In addition, based on the read coverage, our proposed filter can be well applied to the single pipeline for refining the calling results.

In this study, we analyzed the calling results from three pipelines named GATK, RTG and VarScan. Generally speaking, GATK can identify the DNSNVs in the low GC-content region with the lowest error rate among the three pipelines while RTG tends to detect the DNSNVs in the high GC-content region. Considering the effect of high GC-content on experimental and computational results of SNV detection, the Ti/Tv ratio achieved by RTG was lower than that achieved by GATK, indicating

a higher error rate in RTG calling results. Therefore, when using single pipeline to identify the DNSNVs, people should pay more attention to the DNSNVs in the GATK calling results that fell into the high GC-content region, or the DNSNVs in the RTG results that fell into the low GC-content region. For VarScan, although the calling results covered a broad region of the GC-content, the Ti/Tv ratio of the DNSNVs in the low GC-content region was only 1.05, which is much lower than that achieved by GATK (Ti/Tv = 1.88). The Ti/Tv ratio of the DNSNVs in the high GC-content region (Ti/Tv = 1.32) was comparable to that achieved by RTG (Ti/Tv = 1.25). So, people still need to carefully validate the DNSNVs detected in the low GC-content region when using VarScan for calling.

For the purpose of removing the redundant DNSNVs, we proposed a filter to refine the calling results for the single pipeline by considering the read coverage at the mutation sites of the son's genome and the parents' genomes. Our results showed that a number of DNSNVs were removed from the GATK calling results when applying a non-stringent cut-off (score = 0) and the Ti/Tv ratio of the left DNSNVs increased, indicating an improvement of the calling results. For the less redundant results, e.g., the DNSNVs detected by VarScan, only a small number of DNSNVs were filtered out and the Ti/Tv ratio did not changed significantly. Our findings indicated that the proposed filter might be benefit to the refinement of the DNSNVs identified by current pipelines.

It is worth noting that, considering the fact that the size of the input data is small, the filtering algorithm can be further improved by giving a confidence index score and a statistical test index, (e.g., *p*-value) for the score of a DNSNV, which can be estimated from the mapping probability of the reads and the confidence level of the read coverage. It is helpful for increasing the confidence level of the findings. Moreover, due to the limitation of samples, we only used normal samples for the comparative analysis in this study. The work could be further improved by using genetic disease samples, which would make the evaluation of the variant calling error rate of the specific pipeline more accurate. When constructing the filtering algorithm, we only used the exome sequence as an input. The proposed algorithm can be broadened to the analysis of whole genome sequence by integrating additional steps to trim the whole genome. Additionally, we only discussed the difference of *de novo* SNVs detected by different pipelines. In fact, other types of *de novo* structural variations, such as indels, also play important roles in the biological processes in the genetic diseases. While taking more types of structural variations into account will

make the comparative analysis more complex, it does contribute to a more comprehensive assessment of the performance of existing pipelines. These issues will be further addressed in our future study.

CONCLUSION

In this study, we demonstrated that different pipelines have a specific tendency to detect the DNSNVs in the genomic regions with different GC contents. GATK performed better on detecting the DNSNVs in the low GC-content region while RTG and VarScan are better suited for detecting the DNSNVs in the high GC-content region. To refine the calling results for single pipeline, the read coverage at the mutation positions of the son's genome and the parents' genomes can be considered as an effective index to identify DNSNVs with high confidence. Our findings would be useful for the community to choose the appropriate pipelines and obtain the calling results with high confidence when discovering the *de novo* mutations for the genetic diseases.

AUTHOR CONTRIBUTIONS

ZW designed the experiments. YL, LH, YZ, and YH performed the data analysis. ZW wrote the initial version of manuscript. YZ, YL, and ML prepared all the figures. ZW, CL, and XP discussed the results and revised the manuscript. All authors contributed to discussions regarding the results and the manuscript.

FUNDING

This project was supported by grants from the National Natural Science Foundation of China (Nos. 21575094 and 21573151) and NSAF (No. U1730127).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00358/full#supplementary-material>

FIGURE S1 | The overlap rates between the DNSNVs and the variants in the dbSNP database.

REFERENCES

- Bacchelli, C., and Williams, H. J. (2016). Opportunities and technical challenges in next-generation sequencing for diagnosis of rare pediatric diseases. *Expert Rev. Mol. Diagn.* 16, 1073–1082. doi: 10.1080/14737159.2016.1222906
- Barak, T., Kwan, K. Y., Louvi, A., Demirbilek, V., Saygi, S., Tuysuz, B., et al. (2011). Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat. Genet.* 43, 590–594. doi: 10.1038/ng.836
- Blake, D. J., Nawrotzki, R., Loh, N. Y., Gorecki, D. C., and Davies, K. E. (1998). beta-dystrobrevin, a member of the dystrophin-related protein family. *Proc. Natl. Acad. Sci. U.S.A.* 95, 241–246. doi: 10.1073/pnas.95.1.241
- Boycott, K. M., Vanstone, M. R., Bulman, D. E., and MacKenzie, A. E. (2013). Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* 14, 681–691. doi: 10.1038/nrg3555
- Campeau, P. M., Lenk, G. M., Lu, J. T., Bae, Y., Burrage, L., Turnpenny, P., et al. (2013). Yunis-Varón syndrome is caused by mutations in FIG4, encoding a phosphoinositide phosphatase. *Am. J. Hum. Genet.* 92, 781–791. doi: 10.1016/j.ajhg.2013.03.020
- Choi, Y., Chan, A. P., Kirkness, E., Telenti, A., and Schork, N. J. (2018). Comparison of phasing strategies for whole human genomes. *PLoS Genet.* 14:e1007308. doi: 10.1371/journal.pgen.1007308
- Cleary, J. G., Braithwaite, R., Gaastra, K., Hilbush, B. S., Inglis, S., Irvine, S. A., et al. (2014). Joint variant and *de novo* mutation identification on pedigrees from

- high-throughput sequencing data. *J. Comput. Biol.* 21, 405–419. doi: 10.1089/cmb.2014.0029
- Conrad, D. F., Keebler, J. E., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., et al. (2012). Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* 43, 712–714. doi: 10.1038/ng.862
- Francioli, L. C., Cretu-Stancu, M., Garimella, K. V., Fromer, M., Kloosterman, W. P., Samocha, K. E., et al. (2017). A framework for the detection of *de novo* mutations in family-based sequencing data. *Eur. J. Hum. Genet.* 25, 227–233. doi: 10.1038/ejhg.2016.147
- Gomez-Romero, L., Palacios-Flores, K., Reyes, J., Garcia, D., Boege, M., Davila, G., et al. (2018). Precise detection of *de novo* single nucleotide variants in human genomes. *Proc. Natl. Acad. Sci. U.S.A.* 115, 5516–5521. doi: 10.1073/pnas.1802244115
- Hamdan, F. F., Srour, M., Capo-Chichi, J. M., Daoud, H., Nassif, C., Patry, L., et al. (2014). *De novo* mutations in moderate or severe intellectual disability. *PLoS Genet.* 10:e1004772. doi: 10.1371/journal.pgen.1004772
- He, M., Person, T. N., Hebring, S. J., Heinzen, E., Ye, Z., Schrod, S. J., et al. (2015). SeqHBase: a big data toolset for family based sequencing data analysis. *J. Med. Genet.* 52, 282–288. doi: 10.1136/jmedgenet-2014-102907
- Hoischen, A., van Bon, B. W., Gilissen, C., Arts, P., van Lier, B., Stehouwer, M., et al. (2010a). *De novo* mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* 42, 483–485. doi: 10.1038/ng.581
- Hoischen, A., van Bon, B. W., Rodriguez-Santiago, B., Gilissen, C., Vissers, L. E., de Vries, P., et al. (2010b). *De novo* nonsense mutations in ASXL1 cause Bohring-Opitz syndrome. *Nat. Genet.* 43, 729–731. doi: 10.1038/ng.868
- Hyrenius-Wittsten, A., Pilheden, M., Stureson, H., Hansson, J., Walsh, M. P., Song, G., et al. (2018). *De novo* activating mutations drive clonal evolution and enhance clonal fitness in KMT2A-rearranged leukemia. *Nat. Commun.* 9:1770. doi: 10.1038/s41467-018-04180-1
- Jamuar, S. S., and Tan, E. C. (2015). Clinical application of next-generation sequencing for Mendelian diseases. *Hum. Genomics* 9:10. doi: 10.1186/s40246-015-0031-5
- Jin, Z. B., Wu, J., Huang, X. F., Feng, C. Y., Cai, X. B., Mao, J. Y., et al. (2017). Trio-based exome sequencing arrests *de novo* mutations in early-onset high myopia. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4219–4224. doi: 10.1073/pnas.1615970114
- Koboldt, D. C., Larson, D. E., and Wilson, R. K. (2013). Using VarScan 2 for germline variant calling and somatic mutation detection. *Curr. Protoc. Bioinformatics* 44, 15.14.11–15.14.17. doi: 10.1002/0471250953.bi1504s44
- Kojima, K., Nariiai, N., Mimori, T., Takahashi, M., Yamaguchi-Kabata, Y., Sato, Y., et al. (2013). A statistical variant calling approach from pedigree information and local haplotyping with phase informative reads. *Bioinformatics* 29, 2835–2843. doi: 10.1093/bioinformatics/btt503
- Krier, J. B., Kalia, S. S., and Green, R. C. (2016). Genomic sequencing in clinical practice: applications, challenges, and opportunities. *Dialogues Clin. Neurosci.* 18, 299–312.
- Ku, C. S., Vasilou, V., and Cooper, D. N. (2012). A new era in the discovery of *de novo* mutations underlying human genetic disease. *Hum. Genomics* 6:27.
- Lee, H., Deignan, J. L., Dorrani, N., Strom, S. P., Kantarci, S., Quintero-Rivera, F., et al. (2014). Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA* 312, 1880–1887. doi: 10.1001/jama.2014.14604
- Li, B., Chen, W., Zhan, X., Busonero, F., Sanna, S., Sidore, C., et al. (2012). A likelihood-based framework for variant calling and *de novo* mutation detection in families. *PLoS Genet.* 8:e1002944. doi: 10.1371/journal.pgen.1002944
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Ng, S. B., Bigham, A. W., Buckingham, K. J., Hannibal, M. C., McMillin, M. J., Gildersleeve, H. I., et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* 42, 790–793. doi: 10.1038/ng.646
- Peng, G., Fan, Y., Palculict, T. B., Shen, P., Ruteshouser, E. C., Chi, A.-K., et al. (2013). Rare variant detection using family-based sequencing analysis. *Proc. Natl. Acad. Sci. U.S.A.* 110, 3985–3990. doi: 10.1073/pnas.1222158110
- Peters, B. A., Kermani, B. G., Alferov, O., Agarwal, M. R., McElwain, M. A., Gulbahce, N., et al. (2015). Detection and phasing of single base *de novo* mutations in biopsies from human in vitro fertilized embryos by advanced whole-genome sequencing. *Genome Res.* 25, 426–434. doi: 10.1101/gr.181255.114
- Ramu, A., Noordam, M. J., Schwartz, R. S., Wuster, A., Hurles, M. E., Cartwright, R. A., et al. (2013). DeNovoGear: *de novo* indel and point mutation discovery and phasing. *Nat. Methods* 10, 985–987. doi: 10.1038/nmeth.2611
- Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., et al. (2012). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat. Biotechnol.* 30, 61–68. doi: 10.1038/nbt.2053
- Salzberg, S. L., Pertea, M., Fahrner, J. A., and Sobreira, N. (2014). DIAMUND: direct comparison of genomes to detect mutations. *Hum. Mutat.* 35, 283–288. doi: 10.1002/humu.22503
- Santoni, F. A., Makrythanasis, P., Nikolaev, S., Guipponi, M., Robyr, D., Bottani, A., et al. (2014). Simultaneous identification and prioritization of variants in familial, *de novo*, and somatic genetic disorders with VariantMaster. *Genome Res.* 24, 349–355. doi: 10.1101/gr.163832.113
- Shin, S. C., Ahn, D. H., Kim, S. J., Lee, H., Oh, T. J., Lee, J. E., et al. (2013). Advantages of single-molecule real-time sequencing in high-GC content genomes. *PLoS One* 8:e68824. doi: 10.1371/journal.pone.0068824
- Thiffault, I., and Lantos, J. (2016). The challenge of analyzing the results of next-generation sequencing in children. *Pediatrics* 137(Suppl. 1), S3–S7. doi: 10.1542/peds.2015-3731C
- Turner, T. N., Coe, B. P., Dickel, D. E., Hoekzema, K., Nelson, B. J., Zody, M. C., et al. (2017). Genomic patterns of *de novo* mutation in simplex autism. *Cell* 171, 710–722.e12. doi: 10.1016/j.cell.2017.08.047
- Veltman, J. A., and Brunner, H. G. (2012). *De novo* mutations in human genetic disease. *Nat. Rev. Genet.* 13, 565–575. doi: 10.1038/nrg3241
- Wei, Q., Zhan, X., Zhong, X., Liu, Y., Han, Y., Chen, W., et al. (2015). A Bayesian framework for *de novo* mutation calling in parents-offspring trios. *Bioinformatics* 31, 1375–1381. doi: 10.1093/bioinformatics/btu839
- Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* 369, 1502–1511. doi: 10.1056/NEJMoa1306555
- Zhang, L., Zhang, J., Yang, J., Ying, D., Lau, Y. L., and Yang, W. (2013). PriVar: a toolkit for prioritizing SNVs and indels from next-generation sequencing data. *Bioinformatics* 29, 124–125. doi: 10.1093/bioinformatics/bts627
- Zhou, X., Batzoglu, S., Sidow, A., and Zhang, L. (2018). HAPDeNovo: a haplotype-based approach for filtering and phasing *de novo* mutations in linked read sequencing data. *BMC Genomics* 19:467. doi: 10.1186/s12864-018-4867-7
- Zook, J. M., Catoe, D., McDaniel, J., Vang, L., Spies, N., Sidow, A., et al. (2016). Extensive sequencing of seven human genomes to characterize benchmark reference materials. *Sci. Data* 3:160025. doi: 10.1038/sdata.2016.25
- Zook, J. M., Chapman, B., Wang, J., Mittelman, D., Hofmann, O., Hide, W., et al. (2014). Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* 32, 246–251. doi: 10.1038/nbt.2835

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Liang, He, Zhao, Hao, Zhou, Li, Li, Pu and Wen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Novel Nonsense Mutation in *FERMT3* Causes LAD-III in a Pakistani Family

Saba Shahid^{1*}, Samreen Zaidi², Shariq Ahmed¹, Saima Siddiqui¹, Aiysha Abid³, Shabbir Malik² and Tahir Shamsi⁴

¹ Department of Genomics and Clinical Genetics, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan, ² Department of Pediatrics, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan, ³ Center of Human Genetics and Molecular Medicine, Sindh Institute of Urology and Transplantation, Karachi, Pakistan, ⁴ Department of Clinical Hematology, National Institute of Blood Diseases and Bone Marrow Transplantation, Karachi, Pakistan

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Fan Jin,
Zhejiang University, China
Yan-Qing Ma,
Bloodcenter of Wisconsin,
United States
Joshua Xu,
National Center for Toxicological
Research (FDA), United States

*Correspondence:

Saba Shahid
sabashahid_dbt@yahoo.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 07 December 2018

Accepted: 04 April 2019

Published: 24 April 2019

Citation:

Shahid S, Zaidi S, Ahmed S,
Siddiqui S, Abid A, Malik S and
Shamsi T (2019) A Novel Nonsense
Mutation in *FERMT3* Causes LAD-III
in a Pakistani Family.
Front. Genet. 10:360.
doi: 10.3389/fgene.2019.00360

Leukocyte adhesion deficiency-III (LAD3) is an extremely rare primary immunodeficiency disorder, transmitted with autosomal-recessive inheritance. It is caused by genetic alteration in the *FERMT3* gene, which leads to abnormal expression of kindlin-3. This cytoplasmic protein is highly expressed in leukocytes and platelets, and acts as an important regulator of integrin activation. LAD3 has features like bleeding syndrome of Glanzmann-type and leukocyte adhesion deficiency. *FERMT3* mutation(s) have not been well characterized in Pakistani patients with LAD3. In this study, an infant and his family of Pakistani origin, presenting with clinical features of LAD, were investigated to determine the underlying genetic defect. Targeted next generation sequencing (TGS) and Sanger sequencing were performed to identify and confirm the causative mutations, respectively, and their segregation within the family. A novel, homozygous *FERMT3* nonsense mutation (c.286C > T, p.Q96*) was found in the proband, and its co-segregation with LAD3 phenotype within the family was consistent with an autosomal recessive inheritance. Both parents were carriers of the same mutation. This family was offered prenatal diagnosis during first trimester of the subsequent pregnancy; the fetus carried the variant. In conclusion, our study is the first report to identify the novel homozygous variant c.286C > T, p.Q96* in the *FERMT3* gene, which might be the causative mutation for LAD3 patients of Pakistani origin.

Keywords: primary immunodeficiency, leukocyte adhesion deficiency type III, targeted next generation sequencing, *FERMT3* gene, mutation screening

INTRODUCTION

Leukocyte adhesion deficiency (LAD) is a primary immunodeficiency disorder caused by a defect in neutrophil adhesion to the vessel endothelium. There are three different types of this disease, and LAD3, also known as LAD1 variant (LAD1V), is the most rare form (AlmarzaNovoa et al., 2008). Additional manifestations of this disease include bleeding diathesis similar to what occurs in the Glanzmann thrombasthenia, which can, however, be excluded by normal platelet aggregation tests (Boudreaux et al., 2010). LAD3 is caused by a genetic defect in the *FERMT3* gene. This defect

leads to abnormal expression of kindlin-3, a protein whose major role is the regulation of integrin activation, which is essential for the adhesion of leukocytes and platelets (Robert et al., 2011).

Genetic mutations in the *FERMT3* gene (OMIM 607901) run in autosomal recessive pattern in LAD-3 (OMIM 612840) families. *FERMT3* also known, as *KIND3*, *MIG2B*, *UNC112C*, *URP2*, or *URO2SF*, is located on chromosome 11q13.1. It encodes kindlin-3, a cytoskeleton protein involved in the stabilization and activation of the glycoprotein receptor integrin through attachment to its beta subunit. These interactions are responsible for maintaining a stable integrin conformation and to activate its subunits (Ley et al., 2007). Genetic alterations in the *FERMT3* gene cause disruption of the adherent property of integrin on both leukocytes and platelets, possibly due to defect in integrin; its structure is intact, but activation (and thus binding) is not appropriate (Svensson et al., 2009; Zimmerman, 2009).

LAD3 and LAD1 have similar clinical manifestations i.e., leukocytosis, delay in the detachment of the umbilical cord, and critical life-threatening bacterial infections. In addition, there is platelet aggregation dysfunction, which results in severe bleeding episodes. This disorder has mostly been reported in patients of Turkish, Arab Maltese or African American origin. In the present study, we used targeted next-generation sequencing (TGS) technology, the advance methodology (Zhu et al., 2017), and found a novel homozygous mutation in the *FERMT3* gene in a Pakistani family with autosomal recessive LAD3. Sanger sequencing-based prenatal diagnosis was offered to the family for the successive pregnancy, and it confirmed the co-segregation of this genetic mutation with the phenotype in this family.

CASE PRESENTATION

Clinical Report

The index patient is a seven-month-old boy born to first cousins parents, presenting with a prolonged history of fever and recurrent infections for 4 months. Parents reported intermittent bleeding episodes from the nose, mouth, and anus that, during patient hospitalization, were unsuccessfully treated with broad-spectrum antibiotics and packed red cells and platelets transfusion. Examination revealed a failure to thrive in the child, with both height and body weight below the 3rd percentile. He had severe pallor, bruises all over the body, and there were bilateral anterior and posterior cervical palpable lymph nodes, which were firm and tender. The liver was also palpable; it was 9 cm in span, soft and non-tender, while a firm spleen was also palpable 3 cm in its longitudinal axis. The previous record had shown bicytopenia and leukocytosis, growth of multiple microorganisms in blood, including *Burkholderia cepacia* and *Staphylococcus aureus*, and persistently high inflammatory markers. Extensive investigations done during this admission confirmed the anemia, thrombocytopenia, and leukocytosis. Bone marrow aspiration and trephine biopsy showed cellular marrow. Basic primary immunodeficiency workup showed normal immunoglobulin, while flow cytometry revealed normal CD18 expression. There was strong suspicion of primary

immunodeficiency due to the persistent leukocytosis and recurrent infections.

METHODS

Ethics Statement, Consent Statement, and Proband

The study protocol was in accordance with the Institutional Review Board (ERC/IRB) and conformed to the tenets of the Declaration of Helsinki. Written informed consent was obtained from the parent of the patient for the publication of this case report. This study consisted of the proband and three closely related family members from two generations, with history of consanguineous marriage, described by their genetic workup and pedigree analysis in **Figure 1**.

Targeted Next Generation Sequencing

Peripheral blood samples of the family were drawn and DNA was extracted using a QIAamp DNA Blood Mini Kit (Qiagen), following manufacturer's instructions. The targeted next generation sequencing was performed using the Illumina TruSight™. Inherited Disease sequencing panel, a disease targeted sequencing research panel focusing on 552 genes in regions known to harbor recessive pediatric pathogenic mutations. This panel targets 2.25 Mb of the human genomic content, with fragments of ~500 bp. The medium coverage of the sample was >95% of amplicons at >100× coverage. Library was constructed by capturing targeted region using TruSight™ rapid capture. Enriched libraries were loaded onto flow cell (Illumina, CA, United States) and paired-end sequencing runs were processed on a MiSeq (Illumina™) genome sequencer. Data analysis alignment was performed with on-instrument MiSeq reporter software. The mutations identified as pathogenic were confirmed using Sanger method according to the standard protocol (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems®).

Sanger Sequencing

Polymerase chain reaction (PCR) amplification and Sanger sequencing of the TGS-identified variant was performed to confirm TGS results. Primers listed in **Table 1** surrounding the identified mutation were used to amplify a product of 332 bp, which was then sequenced by Sanger sequencing on an ABI-3500 sequencer instrument (Applied Biosystems Inc., Foster City, CA, United States).

RESULTS

Mutation Screening by Targeted Next Generation Sequencing

The identification of the severe immunodeficiency-causing gene mutations, through targeted inherited diseases sequencing panel, was performed on the index patient gDNA sample. Disease causing mutations were identified by the VariantStudio software

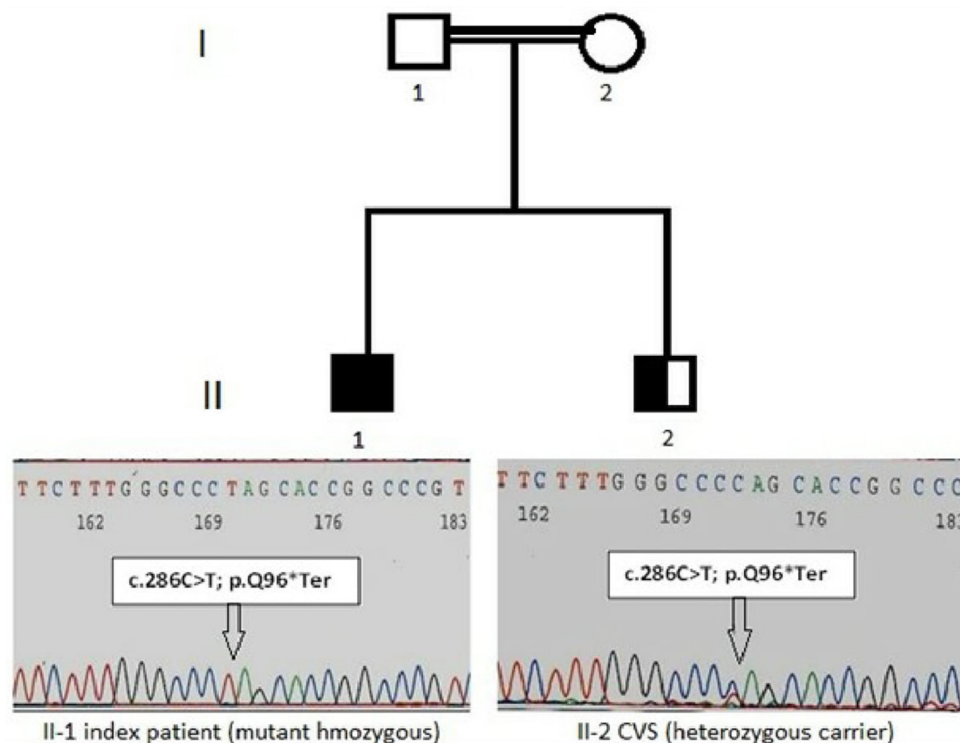


FIGURE 1 | Pedigree with LAD-3 diseases in the index patient. The patient (II-1) indicates the homozygous nonsense variant of the *FERMT3* gene in an electropherogram: NM_178443:c.286C > T and CVS (II-2) indicate with nonsense mutation as heterozygous carrier.

and Variant Interpreter tool (Illumina). These interpretation modules can call variants automatically; options are available to apply stringent filters and for the annotation of NGS data (Hu et al., 2017). The *in silico* prediction tools SIFT, Polyphen 2, MutationTaster, MutationAssessor, dbSNP, and COSMIC were applied to filter pathogenic, benign and variant of uncertain significance. Some of the newly identified variants were not present in any of the above-mentioned public databases, and would need further verification. Interestingly, a single homozygous nucleotide substitution (c.C286T) in the exon 3 of the *FERMT3* gene (NM_178443) was identified in this index patient. This mutation leads to an amino acid change from glutamine (Gln, Q) to a stop codon in position 96 of the kindlin-3 protein (p.Q96*). Additional details of the identified sequence variant c.C286T (p.Q96*), together with associated pathogenic effects, are mentioned in **Table 2** and **Figure 1**. This variant was seemed to be novel, as it could not be identified by searching in other databases like dbSNP, COSMIC, and HGMD (**Table 2**). Variants identified in other genes for *EVC*, *DPYD*, *COL4A3*,

and *TSPYL1* by NGS were excluded as not having a damaging effect assessed by prediction tools on proteins. Finally, this family was offered prenatal diagnosis during subsequent pregnancy, which was performed by chorionic villous sampling done during the 11th week of gestation and Sanger sequencing; the fetus was found to be heterozygote for the same mutation c.C286T (p.Q96*) (**Figure 1**).

DISCUSSION

Leukocyte adhesion deficiency-III (LAD3) is a rare and recently identified primary immunodeficiency, which has different genetic mutations than the ones present in the other two LAD types. In the current study, we found a novel homozygous, stop codon variant c.C286T (p.Q96*) in the *FERMT3* gene in a Pakistani family. The protein structure of *FERMT3* comprises of a FO

TABLE 1 | Primer sequences.

Primer	Sequence 5'-3'	Product size	TM
Forward (FERMT3-F)	CTGAATCCTGGGGTTGTGCT	332 bp	62°C
Reverse (FERMT3-R)	GAATCAGCCGGGCACTTAC		

TABLE 2 | *FERMT3* variant identified in a LAD3 patient.

Gene	Exon	Nucleotide	Protein	Variation		
				Type	Status	dbSNP
<i>FERMT3</i>	03	c.C286T	p.Q96*	Homozygous	Nonsense	Novel

FERMT3, leukocyte adhesion deficiency-III; c, variation at cDNA level; p, variation at protein level; *stop codon.

TABLE 3 | List of reported *FERMT3* mutations in LAD3.

S. No	Nucleotide change	Protein change	Type	References
1	c.286C > T	p.Gln96X	Nonsense	Current study
2	c.1069C > T	p. Arg357X	Nonsense	Wolach et al., 2019
3	c.1795C > T	p.Gln599Ser	Missense	Suratannon et al., 2016
4	c.1597C > T	p.Gln533X	Nonsense	Crazzolaro et al., 2015
5	c.1426C > T	p.Gln476X	Nonsense	Harris et al., 2012
6	c.238_244dup	p.Lys82ThrfsX67	Insertion	Van de Vijver et al., 2012
7	c.922G > A	p. Gly308Arg	Missense	McDowall et al., 2010
8	c.1275delT	p.Glu426ArgfsX3	Frame shift	McDowall et al., 2010
9	c.161-2A > C	p.Asn54ArgfsX142	Splice site	Robert et al., 2011
10	c.687G > A	p.Trp229X	Nonsense	Manevich-Mendelson et al., 2009; Robert et al., 2011
11	c.48G > A	p.Trp16X	Nonsense	Malinin et al., 2009
12	c.1671-2A > G	Deletion exon 14 p.Phe558TrpfsX141	Splice site	Svensson et al., 2009
13	c.1729C > T	p.Arg577X	Nonsense	Kuijpers et al., 2009
14	c.1717C > T	p.Arg573X	Nonsense	Kuijpers et al., 2009; Jurk et al., 2010
15	c.1525C > T	p.Arg509X	Nonsense	Mory et al., 2008; Kuijpers et al., 2009; Malinin et al., 2009; Svensson et al., 2009
16	c.1537C > T	p.Arg513X	Nonsense	Mory et al., 2008; Kuijpers et al., 2009; Malinin et al., 2009; Svensson et al., 2009

FERMT3, leukocyte adhesion deficiency-III; c, variation at cDNA level; p, variation at protein level.

domain, an F1 domain, an F2 with PH domain and an F3 domain that have the binding site to integrin beta subunit. The identified mutation lies within the FO domain in exon 3, and is different from the mutations that were previously identified at the N-terminal of the protein, specifically in the pleckstrin homology and *FERMT3* sub domains (Robert et al., 2011). Robert et al. (2011) reported a p.N54Rfs142 mutation at a splice site within the FO domain. In that report, *in vitro* studies suggested that this mutation was causing a decrease in the mRNA level resulting in an unstable transcript. The nonsense mutation p.Q96X that we identified in this study is also lying within the FO domain. Similarly, nonsense mutations leading to defects in protein expression were reported in patients of Turkish (Mory et al., 2008; Kuijpers et al., 2009; Svensson et al., 2009), Arab (Kuijpers et al., 2009; Malinin et al., 2009), Maltese (Svensson et al., 2009), and African American origin (McDowall et al., 2010).

To date, very few cases have been described for Leukocyte Adhesion Deficiency all over the world; most of the affected individuals (323) were diagnosed with LAD1 (AlmarzaNovoa et al., 2008), while LAD-3 seems to be more sporadic. It is possible that LAD is reported with even lower frequency, due to the failure in correctly diagnosing rare entities. LAD3 cases caused by genetic mutations in *FERMT3* were reported in Turkish and Maltese patients; a homozygous nonsense mutation (R509X) was reported in the Turkish patients, while the Maltese patient was found to be homozygous for an A-to-G substitution in exon 14 at the splice acceptor site (Svensson et al., 2009). Expression studies confirmed that both mutations were destabilizing *KINDLIN3* mRNA. However, in these studies, Western blotting showed no expression of

KINDLIN3 protein in the patients, whereas expression was normal in their parents.

A novel p. R573X nonsense mutation in *FERMT3* was reported in a Turkish patient, while p.W229X in Arabic patients. *In vitro* studies revealed that *FERMT3* protein was not present in leukocytes and platelets of all tested patients, which had, however, similar defects in neutrophil and platelet function (Kuijpers et al., 2009). In addition to its adhesion properties, *FERMT3* gene product is also involved in leukocyte migration. This was confirmed by the *in vitro* effects of the homozygous mutations (G308R and 1275delT) in the *FERMT3* gene, which were the cause of severe LAD3 in an African American girl (McDowall et al., 2010). Almost all cases of LAD III were diagnosed with innate immune defects. However, Suratannon et al. (2016) identified p.Gln599Ser mutation in *FERMT3* gene in Thai patient that presented with humoral immune defect (Suratannon et al., 2016). **Table 3** summarizes all the mutations in the *FERMT3* identified in the literature. To the best of our knowledge, this *FERMT3* variant is a novel mutation that broadens the mutation spectrums of LAD3. Thus, this finding shows that the recessive *FERMT3* mutation c.C286T (p.Q96*) likely caused LAD-3 in our studied Pakistani pedigree.

CONCLUSION

In conclusion, this study wants to stress the importance of early diagnosis. As in the majority of primary immunodeficiency diseases, the prognosis of LAD3 is extremely dependent on early age diagnosis, with timely management of bacterial infections and consideration for HSCT. In addition, this autosomal

recessive disorder has high incidence in areas with high rate of consanguineous marriages. Therefore, broadening the spectrum of known mutations underlying the phenotype of such a life-threatening disease can help offering and performing better genetic counseling and prenatal diagnosis.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations Institutional Review Board (ERC/IRB). The protocol was approved by the Institutional Review Board (ERC/IRB). Written informed consent was obtained from the parents of the subjects in accordance with the Declaration of Helsinki.

REFERENCES

- AlmarzaNovoa, E., Kasbekar, S., Thrasher, A. J., Kohn, D. B., Sevilla, J., Nguyen, T., et al. (2008). Leukocyte adhesion deficiency-I: a comprehensive review of all published cases. *J. Allergy Clin. Immunol. Pract.* 6, 1418–1420. doi: 10.1016/j.jaip.2017.12.008
- Boudreaux, M. K., Wardrop, K. J., Kiklevich, V., Felsburg, P., and Snekvik, K. A. (2010). Mutation in the canine kindlin-3 gene associated with increased bleeding risk and susceptibility to infections. *Thromb. Haemost.* 103, 475–477. doi: 10.1160/TH09-09-0571
- Crazzolar, R., Maurer, K., Schulze, H., Zieger, B., Zustin, J., and Schulz, A. S. (2015). A new mutation in the kindlin-3 gene ablates integrin-dependent leukocyte, platelet, and osteoclast function in a patient with leukocyte adhesion deficiency-III. *Pediatr. Blood Cancer* 62, 1677–1679. doi: 10.1002/psc.25537
- Harris, E. S., Smith, T. L., Springett, G. M., Weyrich, A. S., and Zimmerman, G. A. (2012). Leukocyte adhesion deficiency-I variant syndrome (LAD-IV, LAD III): molecular characterization of the defect in an index family. *Am. J. Hematol.* 87, 311–313. doi: 10.1002/ajh.22253
- Hu, P., Wu, S., Yuan, L., Lin, Q., Zheng, W., Xia, H., et al. (2017). Compound heterozygous POMT1 mutations in a Chinese family with autosomal recessive muscular dystrophydystroglycanopathy C1. *J. Cell Mol. Med.* 21, 1388–1393. doi: 10.1111/jcmm.13068
- Jurk, K., Schulz, A. S., Kehrel, B. E., Rapp, D., Schulze, H., Möbust, D., et al. (2010). Novel integrin-dependent platelet malfunction in siblings with leukocyte adhesion deficiency-III (LAD-III) caused by a point mutation in FERMT3. *Thromb. Haemost.* 103, 1053–1064. doi: 10.1160/TH09-10-0689
- Kuijpers, T. W., van de Vijver, E., Weterman, M. A. J., de Boer, M., Tool, A. T. J., van den Berg, T. K., et al. (2009). LAD-1/variant syndrome is caused by mutations in FERMT3. *Blood* 113, 4740–4746. doi: 10.1182/blood-2008-10-182154
- Ley, K., Laudanna, C., Cybulsky, M. I., and Nourshargh, S. (2007). Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat. Rev. Immunol.* 7, 678–689. doi: 10.1038/nri2156
- Malinin, N. L., Zhang, L., Choi, J., Ciocea, A., Razorenova, O., Ma, Y. Q., et al. (2009). A point mutation in KINDLIN3 ablates activation of three integrin subfamilies in humans. *Nat. Med.* 15, 313–318. doi: 10.1038/nm.1917
- Manevich-Mendelson, E., Feigelson, S. W., Pasvolosky, R., Aker, M., Grabovsky, V., Shulman, Z., et al. (2009). Loss of kindlin-3 in LAD-III eliminates LFA-1 but not VLA-4 adhesiveness developed under shear flow conditions. *Blood* 114, 2344–2353. doi: 10.1182/blood-2009-04-218636
- McDowall, A., Svensson, L., Stanley, P., Patzak, I., Chakravarty, P., Howarth, K., et al. (2010). Two mutations in the KINDLIN3 gene of a new leukocyte adhesion deficiency III patient reveal distinct effects on leukocyte function in vitro. *Blood* 115, 4834–4842. doi: 10.1182/blood-2009-08-238709
- Mory, A., Feigelson, S. W., Yarali, N., Kilic, S. S., Bayhan, G. I., Gershoni-Baruch, R., et al. (2008). Kindlin-3: a new gene involved in the pathogenesis of LAD-III. *Blood* 112:2591. doi: 10.1182/blood-2008-06-163162
- Robert, P., Canault, M., Farnarier, C., Nurden, A., Grosdidier, C., Barlogis, V., et al. (2011). A novel leukocyte adhesion deficiency III variant: kindling3 deficiency results in integrin- and nonintegrin-related defects in different steps of leukocyte adhesion. *J. Immunol.* 19, 5273–5283. doi: 10.4049/jimmunol.1003141
- Suratannon, N., Yeetong, P., Srichomthong, C., Amarinthukrowh, P., Chatchatee, P., Sosothikul, D., et al. (2016). Adaptive immune defects in a patient with leukocyte adhesion deficiency type III with a novel mutation in FERMT3. *Pediatr. Allergy Immunol.* 27, 214–217. doi: 10.1111/pai.12485
- Svensson, L., Howarth, K., McDowall, A., Patzak, I., Evans, R., Ussar, S., et al. (2009). Leukocyte adhesion deficiency-III is caused by mutations in KINDLIN3 affecting integrin activation. *Nat. Med.* 15, 306–312. doi: 10.1038/nm.1931
- Van de Vijver, E., Maddalena, A., Sanal, Ö., Holland, S. M., Uzel, G., Madkaikar, M., et al. (2012). Hematologically important mutations: leukocyte adhesion deficiency (first update). *Blood Cells Mol. Dis.* 48, 53–61. doi: 10.1016/j.bcmd.2011.10.004
- Wolach, B., Gavrieli, R., Wolach, O., Stauber, T., Abuzaitoun, O., Kuperman, A., et al. (2019). Leucocyte adhesion deficiency-A multicentre national experience. *Eur. J. Clin. Invest.* 49:e13047. doi: 10.1111/eci.13047
- Zhu, L., Cheng, J., Zhou, B., Wei, C., Yang, W., Jiang, D., et al. (2017). Diagnosis for choroideremia in a large Chinese pedigree by next generation sequencing (NGS) and noninvasive prenatal testing (NIPT). *Mol. Med. Rep.* 15, 1157–1164. doi: 10.3892/mmr.2017.6119
- Zimmerman, G. A. (2009). LAD syndromes: FERMT3 kindles the signal. *Blood* 113, 4485–4486. doi: 10.1182/blood-2009-01-198853

AUTHOR CONTRIBUTIONS

SabS contributed to the study design, data interpretation, and manuscript writing. SZ and SaiS were responsible for clinical examination and evaluation of patient and family. SA performed the laboratory work. AA contributed to the bioinformatics analysis. SM and TS were involved in study design, patients' recruitment, and supervised the study and reviewed the manuscript.

ACKNOWLEDGMENTS

We appreciate the participation of patient family who participated in this study.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shahid, Zaidi, Ahmed, Siddiqui, Abid, Malik and Shamsi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Next-Generation Sequencing Analysis Reveals Novel Pathogenic Variants in Four Chinese Siblings With Late-Infantile Neuronal Ceroid Lipofuscinosis

Xiao-Tun Ren¹, Xiao-Hui Wang¹, Chang-Hong Ding¹, Xiang Shen², Hao Zhang², Wei-Hua Zhang¹, Jiu-Wei Li¹, Chang-Hong Ren¹ and Fang Fang^{1*}

¹ Department of Neurology, National Centre for Children's Health, Beijing Children's Hospital, Capital Medical University, Beijing, China, ² Running Gene Inc., Beijing, China

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University, China

Reviewed by:

Theodora Katsila,
National Hellenic Research
Foundation, Greece
Fan Jin,
Zhejiang University, China

*Correspondence:

Fang Fang
13910150389@163.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 08 November 2018

Accepted: 08 April 2019

Published: 25 April 2019

Citation:

Ren X-T, Wang X-H, Ding C-H,
Shen X, Zhang H, Zhang W-H, Li J-W,
Ren C-H and Fang F (2019)
Next-Generation Sequencing Analysis
Reveals Novel Pathogenic Variants in
Four Chinese Siblings With
Late-Infantile Neuronal Ceroid
Lipofuscinosis. *Front. Genet.* 10:370.
doi: 10.3389/fgene.2019.00370

Neuronal Ceroid Lipofuscinoses (NCLs) are progressive degenerative diseases mainly affect brain and retina. They are characterized by accumulation of autofluorescent storage material, mitochondrial ATPase subunit C, or sphingolipid activator proteins A and D in lysosomes of most cells. Heterogenous storage material in NCLs is not completely disease-specific. Most of CLN proteins and their natural substrates are not well-characterized. Studies have suggested variants of Late-Infantile NCLs (LINCLs) include the major type CLN2 and minor types CLN5, CLN6, CLN7, and CLN8. Therefore, combination of clinical and molecular analysis has become a more effective diagnosis method. We studied 4 late-infantile NCL siblings characterized by seizures, ataxia as early symptoms, followed by progressive regression in intelligence and behavior, but mutations are located in different genes. Symptoms and progression of 4 types of LINCLs are compared. Pathology of LINCLs is also discussed. We performed Next-Generation Sequencing on these phenotypically similar families. Three novel variants c.1551+1insTGAT in TPP1, c.244G>T in CLN6, c.554-5A>G in MFSD8 were identified. Potential outcome of the mutations in structure and function of proteins are studied. In addition, we observed some common and unique clinical features of Chinese LINCL patient as compared with those of Western patients, which greatly improved our understanding of the LINCLs.

Keywords: Next-Generation Sequencing (NGS), Neuronal Ceroid Lipofuscinosis (NCL), late-infantile, CLN2, CLN5, CLN6, CLN7

INTRODUCTION

NCLs (Neuronal Ceroid Lipofuscinoses) are characterized by accumulation of lysosomal autofluorescent storage material and progressive neurodegeneration (Oishi et al., 1999; Getty and Pearce, 2011; Blom et al., 2013; Kollmann et al., 2013; Mink et al., 2013; Sands, 2013; Patino et al., 2014). The most common clinical features of NCLs include epileptic seizures, progressive regression of intelligence, loss of motor function, retinal degeneration, and premature death (Haltia and Goebel, 2013; Mink et al., 2013; Warrier et al., 2013; De Silva et al., 2015). NCL is one of the most frequent classes of childhood-onset neurodegenerative diseases with prevalence around

0.5–8 per 100,000 live births varying by the regions (Oishi et al., 1999; Getty and Pearce, 2011; Cotman et al., 2013; Haltia and Goebel, 2013; Beltran et al., 2018). So far there are 13 genes identified as candidate genes of NCLs, i.e., *CLN1/PPT1*, *CLN2/TPP1*, *CLN3*, *CLN4/DNAJC5*, *CLN5*, *CLN6*, *CLN7/MFSD8*, *CLN8*, *CLN10/CTSD*, *CLN11/GRN*, *CLN12/ATP13A2*, *CLN13/CTSF*, and *CLN14/KCTD7* (Kollmann et al., 2013; Warrier et al., 2013). Most human NCLs follow autosomal recessive model except that caused by *CLN4/DNAJC5* gene which presents autosomal dominant form (Haltia and Goebel, 2013; Kollmann et al., 2013; Warrier et al., 2013). However, the storage material of NCLs is not disease-specific and the function of most CLN protein has not been well classified. It has been difficult to diagnose merely based on clinical findings. Enzyme assay is used to help diagnose NCLs with mutations in enzyme-coding genes, *CLN1/PPT1*, *CLN2/TPP1*, *CLN10/CTSD* with deficiency of respective enzymes (Kamate et al., 2012; Mole and Williams, 2013). Another enzyme-coding genes, *CLN13/CTSF*, as a novel candidate gene identified in year 2013 is also presumably to cause impairment in enzyme cathepsin F (Mole and Williams, 2013; Schulz et al., 2013), but no supportive experiment due to few cases. Merely based on clinical presentations, NCL could not be easily distinguished from other diseases including Leber's Hereditary Optic Neuropathy (LHON), the symptoms of which include seizures, regression, ataxia, and vision impairment (Fang et al., 2017). Due to limitations of other methods, sequencing has emerged as an effective diagnosis method.

By onset ages, NCLs are classified into congenital, infantile, late-infantile, juvenile, and adult NCLs. Late-Infantile NCLs (LINCLs) include the classic *CLN2* disease and variant *CLN5*, *CLN6*, *CLN7*, and *CLN8* disease (Getty and Pearce, 2011; Mole and Williams, 2013; Schulz et al., 2013; Warrier et al., 2013; Patino et al., 2014). In this study, we present four families with *CLN2*, *CLN5*, *CLN6*, and *CLN7* disease, respectively.

CLN2 disease caused by mutations in *CLN2* gene which encodes the tripeptidyl peptidase 1, a lysosomal serine protease that removes tripeptides from N-terminus of peptides (Kollmann et al., 2013). Mitochondrial ATP-synthase subunit C, a significant component of storage material in LINCL has been demonstrated as one of the substrates of *TPP1* (Ezaki et al., 2000). Deficiency of *TPP1* activity results in accumulation of mitochondrial ATP-synthase subunit C, which may be the pathology of LINCL.

CLN5 gene encodes a soluble polypeptide which predominantly colocalizes with lysosomal-associated membrane protein-1 (LAMP1). Mutations in *CLN5* may cause retention in the ER/Golgi (Isosomppi et al., 2002; Lebrun et al., 2009; Schmiedt et al., 2010). Since *CLN5* is a highly glycosylated protein, it may play an essential role as a sensor in trafficking or integrity of lysosomes (Kollmann et al., 2013).

CLN6 gene encodes a highly conserved membrane protein that exclusively resides in endoplasmic reticulum (ER) (Heine et al., 2004; Kollmann et al., 2013). However, the exact function of *CLN6* is still unknown. Previous experiment shows that loss of *CLN6* activity may affect lysosomal degradation of Arylsulfatase A (Heine et al., 2004). This finding may indicate *CLN6* could play a role in degradation involving ER. Besides, *CLN6* was also

proved to interact with Collapsin Response Mediator Protein-2 (CRMP-2). This interaction probably affects maturation and integrity of axonal outgrowth thus contribute to neuronal dysfunction of LINCL patients (Benedict et al., 2009).

CLN7/MFSD8 encodes a lysosomal membrane protein called Major Facilitator Superfamily Domain-containing protein 8 (MFSD8). This protein is ubiquitously expressed with several splicing variants. It could transport small solutes by electrochemical gradients (Siintola et al., 2007). However, the specific substrates of *CLN7* require further investigation.

In the present study, we found three novel mutations of LINCLs. They are likely pathogenic by analyzing their functional consequences and correlation with the phenotypes. In addition, we also observed some common and unique clinical features of Chinese LINCL patient as compared with those of Western patients, which may improve understanding of the LINCLs.

CASE PRESENTATION

Eight patients including four probands were born in four healthy non-consanguineous Chinese families with normal pregnancy and perinatal history. Pedigrees of four families were presented in **Figure 1A**. All four families possess unremarkable family history. Typically, unsteady gait was observed between age 3–5 as the initial symptom. Only in Family one, seizures were observed as the first symptom at slightly earlier ages from 8 months to 3 years old. Regression in cognition and behavior were then observed in all affected children. Ataxia, seizures were also presented in all patients. Low vision was found in a few patients at this stage. At later stage, most patients lost the ability to sit, stand or walk unaided, they also lost their vision. Two patients in family one died at age seven, one patient in family three died at age 16 (for a comparison of all patients see **Figure 1B**).

Cerebellar atrophy was confirmed by MRI imaging in all four probands. Proband in family 1 also had very significant cerebral atrophy accompanied by atrophy of brain stem, while proband in family three had abnormal myelination (**Figure 1C**). Only symptomatic treatment was used for patients, whereas no obvious improvement was observed. Proteolytic activity of *TPP1* was completely lost in proband 1 (individual II:2 of family 1). The precise information of patients was collected in **Table S1**. All phenotypes of patients was standardized as HPO terms.

Proband 1 (II:2 of family 1) presented general epileptic discharges accompanied by burst-suppression. For proband 2 (II:1 of family 2), EEG background rhythm was slow, general medium to high amplitude slow waves, transient, or continuous slow spike-and-waves was observed. EEG of proband 3 (II:2 of family 3) showed frequent spike-and-wave and slow spike-and-wave discharged during the sleep period in right central electrode. For proband 4 (II:1 of family 4), massive epileptic discharges were noticed.

Since nearly all the probands have siblings who presents similar phenotypes, Whole-Exome Sequencing was performed to investigate the molecular genetic basis of the disease in these family. Sanger sequencing was performed to confirm the identified mutations.

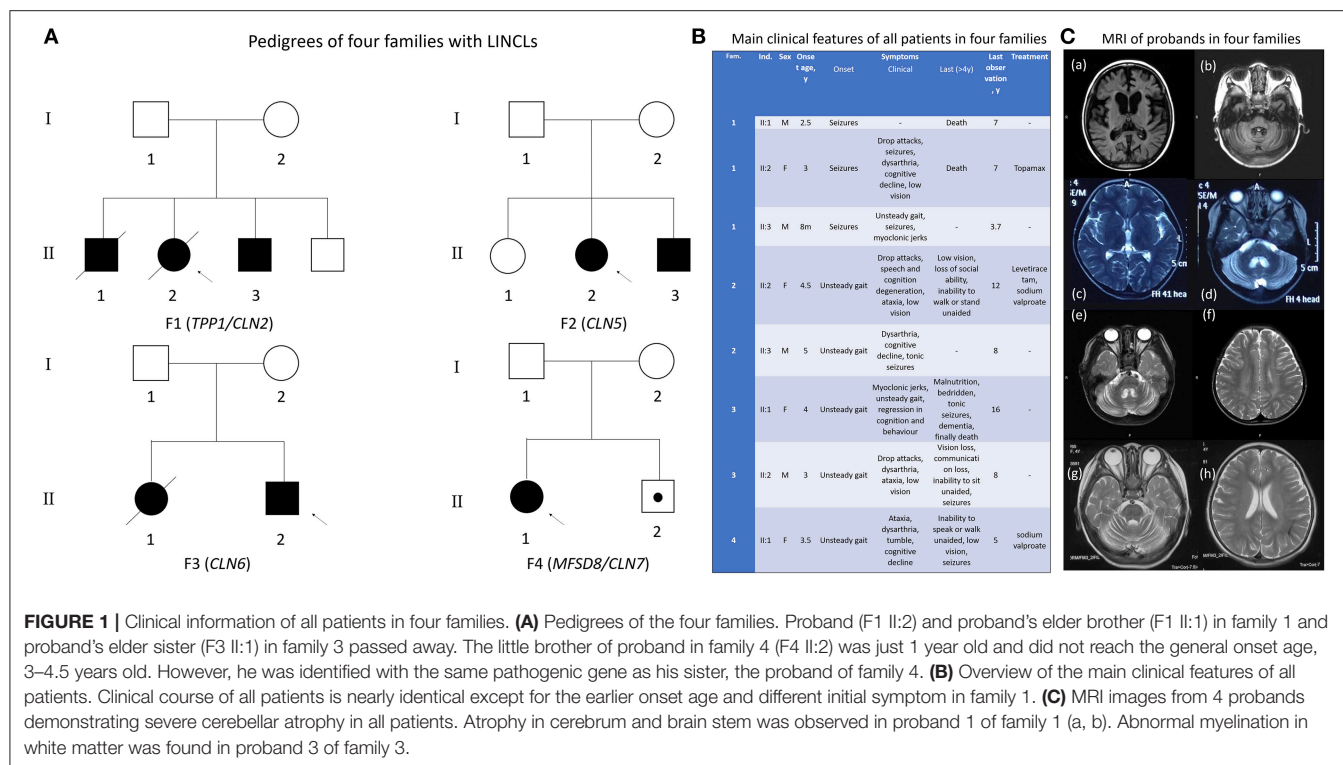


FIGURE 1 | Clinical information of all patients in four families. **(A)** Pedigrees of the four families. Proband (F1 II:2) and proband's elder brother (F1 II:1) in family 1 and proband's elder sister (F3 II:1) in family 3 passed away. The little brother of proband in family 4 (F4 II:2) was just 1 year old and did not reach the general onset age, 3–4.5 years old. However, he was identified with the same pathogenic gene as his sister, the proband of family 4. **(B)** Overview of the main clinical features of all patients. Clinical course of all patients is nearly identical except for the earlier onset age and different initial symptom in family 1. **(C)** MRI images from 4 probands demonstrating severe cerebellar atrophy in all patients. Atrophy in cerebrum and brain stem was observed in proband 1 of family 1 (a, b). Abnormal myelination in white matter was found in proband 3 of family 3.

Informed consents for genetic analyses were obtained from the children's parents. The study was approved by the ethics committee of Beijing Children's Hospital. Written informed consent was obtained from the patients' parents for the publication of this report and any accompanying images.

METHODS

Reference Sequence

All positions in the mutated genes are annotated to reference sequence, namely RefSeq NM_000391, Ensembl ENST00000299427 for TPP/CLN2; NM_006493, Ensembl ENST00000377453 for CLN5; NM_017882, Ensembl ENST00000249806 for CLN6, NM_152778, Ensembl ENST00000296468 for MFSD/CLN7 in this publication. Whole Exome Sequence data were mapped and aligned to Human Genome Build GRCh37/hg19.

Next-Generation Sequencing

Proband DNA was sequenced to discover the causal gene. DNA was isolated from peripheral blood using DNA Isolation Kit (Bioteke, AU1802). One microgram of genomic DNA was fragmented into 200–300 bp length by Covaris Acoustic System. The DNA fragments were then processed by end-repairing, A-tailing and adaptor ligation, a 4-cycle pre-capture PCR amplification, targeted sequences capture. Captured DNA fragments were eluted and amplified by 15 cycle post capture PCR. The final products were sequenced with 150-bp paired-end reads on Illumina HiSeq X platform according to the standard manual. The raw data produced on HiSeq X were filtered and aligned against the human reference genome (hg19) using the BWA Aligner (<http://bio-bwa.sourceforge.net/>).

The quality recalibration was performed using GATK Base Recalibrator (Genome Analysis ToolKit) (www.broadinstitute.org/gatk). The single-nucleotide polymorphisms (SNPs) and small insertions or deletions (indel) were called by GATK Unified Genotyper (Genome Analysis ToolKit) (www.broadinstitute.org/gatk). Variants were annotated using ANNOVAR (annovar.openbioinformatics.org/en/latest/).

Method of Mapping, Genotype, SNP Calling, and Indel Calling

Image analysis and base calling were performed using the Illumina Pipeline. BWA Aligner (<http://bio-bwa.sourceforge.net/>) was used to align clean reads to human reference genome (hg19), the parameters were set as default. The alignment result was then passed to GATK to identify the breakpoints, the parameters were set as “mismatch Fraction=0.05, lod=5, masReadsF or Realignment=30,000, maxReadsInRam=1,000,000.”

We selected variations obtained from exome sequencing with minor allele frequencies <0.05 in any of the following databases (dbSNP, Hapmap, 1000 Genomes Project). Effects of single-nucleotide variants (SNVs) were predicted by SIFT, Polyphen-2, and MutationTaster programs. All variants were interpreted according to ACMG standards and categorized to be pathogenic, likely pathogenic, variants of unknown clinical significance (VUS), likely benign, and benign. We further compared the rest of the deleterious variations in the patients with their unaffected parents and investigated the function of all identified genes according to the published reports and OMIM database.

Sanger Sequencing

The candidate causal genes discovered via WES were then confirmed by Sanger sequencing and co-segregation analyses among the family were also conducted. The primers were designed using Primer Premier 5.0 (Premier Biosoft) and PCR was carried out to amplify the fragments covering the mutated sites. The PCR products were further purified with Zymoclean PCR purification Kit and then sequenced by ABI 3730 DNA Sequencer (Applied Biosystems, Foster City, CA, United States). Sanger sequencing results were analyzed by Chromas Lite v2.01 (Technelysium Pty Ltd., Tewantin, QLD, Australia).

RESULT

Variants Identified by Whole Exome Sequencing

Next-Generation Sequencing was carried out in exome of probands. All four probands were identified with mutations in *TPP1/CLN2*, *CLN5*, *CLN6*, and *MFSD8/CLN7*, respectively (Table 1, Figure S1). Among these seven mutations, c.1551+1insTGAT of *TPP1* gene, c.244G>T of *CLN6* and c.554-5A>G of *MFSD8* gene are novel mutations that haven't been reported before. According to ACMG guidelines, mutation c.1551+1insTGAT of *TPP1* was interpreted as pathogenic since this mutation is a null variant (PVS1), with extremely low frequency (PM2), and the phenotype of the patient is specific for the disease that related to the gene (PP4). Mutation c.244G>T of *CLN6* can be classified as likely pathogenic since it is absent from controls (PM2), was detected in trans with a recently reported pathogenic variant c.892G>A (PM3) co-segregated in multiple affected family members (PP1), predicted to be deleterious by multiple lines of computational evidence (PP3) i.e., predicted to be damaging by SIFT with score 0, deleterious by PROVEAN with score -5.92, probably damaging by PolyPhen-2 with score 1.00, disease causing by MutationTaster with score >0.99, the phenotype of the patient was also specific for disease CLN6 (PP4). Intronic variant c.554-5A>G of *MFSD8* was predicted to be VUS (variants of uncertain significance) according to ACMG standard, since this variant is absent from controls (PM2), detected in trans with a pathogenic variant c.1444C>T (PM3), and the phenotype of patient was similar to CLN7 (PP4).

Sanger Sequencing

All variants identified in Next-Generation Sequencing were then confirmed in other family members by Sanger Sequencing. Both parents of probands are the carriers of one of the two mutations, respectively. Other patients in the family also carry the same mutation as the proband. All results are shown in Figure S2.

DISCUSSION

NCLs are a group of neurological diseases without typical clinical symptoms. The symptoms of NCLs could not be well-distinguished from other neurological diseases especially at early stage. Traditional enzymatic activity detection can only determine certain types of NCLs, i.e., CLN1, CLN2, and CLN10. Other types of NCLs cannot

be well-diagnosed until gene sequencing was introduced (Patino et al., 2014). Here we performed Whole-Exome Sequencing in four Chinese siblings with LINCLs, mutations in CLN genes were identified in all families including 3 novel mutations. Genetic test especially Whole-Exome Sequencing is now a suggestive tool in diagnosis of rare disease and is accepted and recommended by more clinicians now (Jin et al., 2018; Shen, 2018).

Pathogenesis of Novel Mutations

In this study, 3 novel mutations in *TPP1/CLN2*, *CLN6*, and *CLN7* were found, respectively. According to ACMG guidelines, mutation c.1551+1insTGAT in *TPP1/CLN2* was characterized as pathogenic, mutation c.244G>T (G82W) in *CLN6* was interpreted as likely pathogenic and mutation c.554-5A>G was classified as VUS.

For mutation c.1551+1insTGAT, *in silico* analysis suggests that this variant is a four-nucleotide-insertion in exonic region (Figure 2) It is probably not a splicing variant but an insertion variant which affects all the downstream sequence. Insertion of these 4 nucleotides would cause nonsense variant p.V518X. Resulted protein will be truncated after the residue D517. Residue from 518 to 563 would be missing. The highly conserved Ca-binding loop in sedolisin family (aa517-547) (Wlodawer et al., 2001, 2003) would be destroyed. Ca²⁺ is the cofactor of enzyme TPP1 and it was demonstrated necessary for the autocatalysis of the precursor TPP1 into mature form (Kuizon et al., 2010). Destruction of Ca-binding loop would disrupt the Ca²⁺ binding and sequentially hinder the autocatalysis of precursor TPP1. Another vital residue W542 which involves in tripeptidyl peptidase activity and autocatalytic activity of TPP1 (Kuizon et al., 2010) is also obliterated when the protein is truncated after D517.

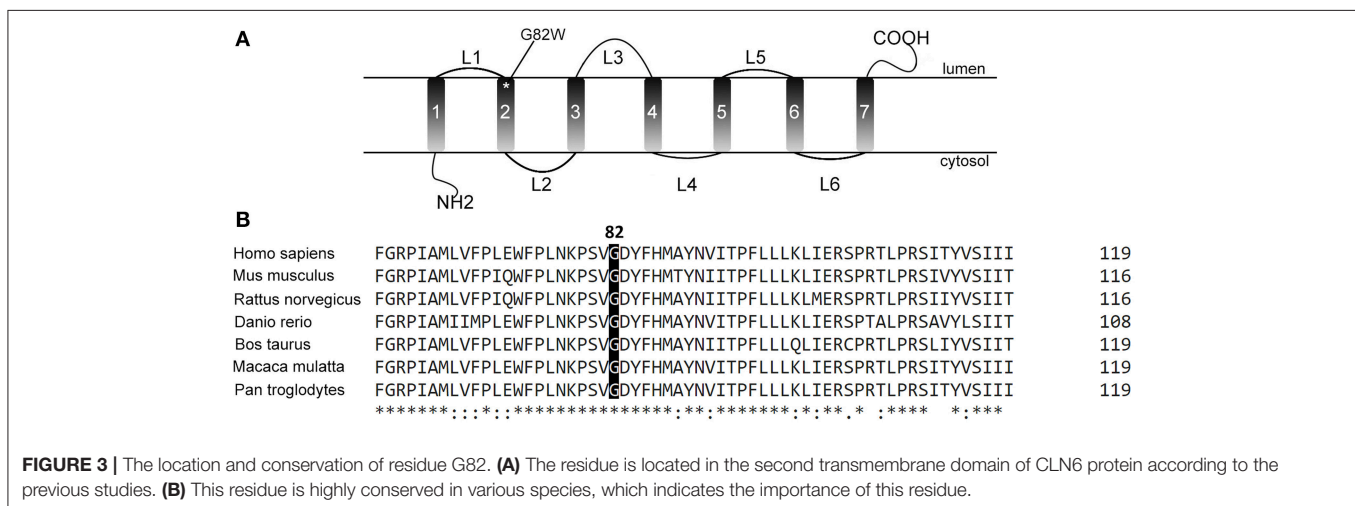
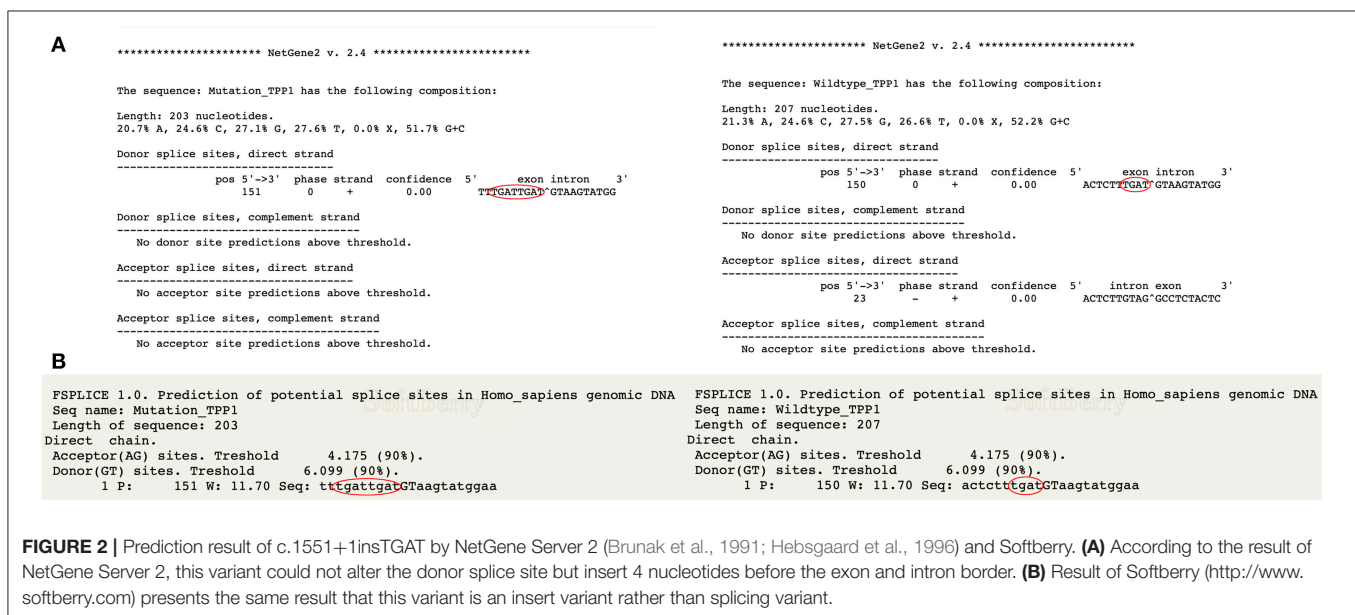
Mutation c.244G>T (G82W) in *CLN6* is a missense mutation which changed the nonpolar negative amino acid Glycine into a nonpolar neutral amino acid. This mutation was predicted as deleterious mutation by PolyPhen2, SIFT, PROVEAN, and MutationTaster. This result indicates that this alteration may harm the proper folding of protein and consequently affect the protein function. In addition, protein CLN6 is a transmembrane protein. The mutated residue is located on the second transmembrane domain (Figure 3A). Although the exact function and interaction of this residue is still unclear, alteration of this residue may change the anchor of CLN6 protein in lysosomal membrane. This G82 residue is highly conservative among various species (Figure 3B). It indicates that this residue is functionally important.

Mutation c.554-5A>G, it is not a variation in canonical splice sites thus cannot be simply classified as null variant. GT/AG mRNA processing rule is valid in almost all eukaryotes including the wild type *MFSD8* sequence (Figure 4). The mutation c.554-5A>G changes the normal intronic site "aa" into another splice acceptor recognition sequence "ag" (Figure 4), which may influence the normal splicing. This hypothesis is further strengthened when predicted by Human Splicing Finder (Figure 5).

TABLE 1 | Gene sequencing result of 4 patients with NCL.

Proband	Orientation	Inheritance	Gene	Pathogenic variants
1	homozygous	AR	TPP1	c.1551+1insTGAT aberrant splicing of Exon 12
2	compound heterozygous	AR	CLN5	c.1068_1069del (Xin et al., 2010) p.(L358AfsX4) c.1100_1103del (Kohan et al., 2008) p.(K368SfsX15)
3	compound heterozygous	AR	CLN6	c.244G>T p.(G82W) c.892G>A (Sun et al., 2018) p.(E298K)
4	compound heterozygous	AR	MFSD8	c.1444C>T (Aiello et al., 2009) p.(R482X) c.554-5A>G aberrant splicing of Exon7

Overview of the gene sequenced by next-generation sequencing in 4 probands.




```

>Mutation_MFSD8
tataataattcatctgtgaaaattttatcttagtaagTTTTTCAGACTTGTTTACATTCCTTGGAGAAAAAGGTGT
GACATGGGATGTGATTAACTGCAGATAAACATGTATACAACACCAGTTTACTTAGCGCCTTCCTG
GGAATTTTAAATATTATTCTGATCCTTGCCATACTAAGGtaattagatagaatgata
>Wildtype_MFSD8
tataataattcatctgtgaaaattttatcttagtaagTTTTTCAGACTTGTTTACATTCCTTGGAGAAAAAGGTGT
GACATGGGATGTGATTAACTGCAGATAAACATGTATACAACACCAGTTTACTTAGCGCCTTCCTG
GGAATTTTAAATATTATTCTGATCCTTGCCATACTAAGGtaattagatagaatgata

```

FIGURE 4 | Sequences mutated MFSD8 and wildtype MFSD8. Mutation c.554-5A>G altered the “aa” sequence into an intronic splice acceptor site “ag”. This might induce splice from after the mutated “ag” site and induce four nucleotides insertion (TAAG) before the real exon thus altered all the downstream sequence.

Mutant sequence Substitution at position 16 (A>G)

```

1 gaaaatttta tcttagtaag TTTTCAGAC TTGTTTACA TTCCTGGAG AAAAAGGTGT GACATGGGAT GTGATTAAAC TGCAGATAAA CATGTATACA
101 ACACCAAGTTT TACTTAGCGC CTCCTGGGA ATTTTAAATA TTATTCTGAT CCTTGCCATA CTAAGgtaat tagatagaaa tgata

```

Total sequence length: 185 nucleotides

The sequences analyzed in HSF are underlined.

Interpreted Data

This table shows only relevant results related to the mutation position and context.

The mutation occurs in the late intronic positions, the following table show results of acceptor splice sites that could be affected by the mutation

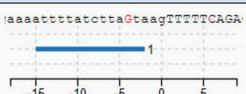
Predicted signal	Prediction algorithm	cDNA Position	Interpretation
New Acceptor Site	1 - HSF Matrices		Activation of an intronic cryptic acceptor site Potential alteration of splicing.

FIGURE 5 | Prediction result of mutation c.554-5>G in MFSD8. This mutation would active an intronic cryptic acceptor site and potentially alter the splicing according to Human Splicing Finder Desmet et al. (2009).

Phenotype Study of LINCL Patients

Although CLN2, CLN5, CLN6, and CLN7 are all LINCLs, their symptoms and onset ages are slightly different in previous reports. Comparing to CLN2, clinical course of CLN5 is milder and slower, the onset age is significantly later, age of death is also significantly delayed. Onset age of visual loss in CLN5 is also significantly later than any other types of LINCLs. Reported age of death of CLN5 is around age 15 and most patients were still alive when reports were published. CLN6 was first reported as the NCL that presented similar clinical course to CLN2. The development of CLN6 was slightly slower than CLN2. Time for ambulation loss and death varied a lot. Seizures appear at the early stage in most CLN6 patients. Development of CLN7 is more severe than CLN2 as most of the patients lost ambulation within 2 years after onset, but the age of death varied from 6.5 to 18 years old. According to the previous reports, clinical information of LINCLs was summarized in **Table 2** (Santavuori et al., 1982, 1991; Eva et al., 1988; Taratuto et al., 1995; Gao et al., 2002; Steinfeld et al., 2002; Sharp et al., 2003; Topcu et al., 2004; Siintola et al., 2007; Cismondi et al., 2008; Kohan et al., 2008; Aiello et al., 2009; Al-Muhaizea et al., 2009; Cannelli et al., 2009; Kousi et al., 2009; Stogmann et al., 2009; Xin et al., 2010; Perez-Poyato et al., 2012;

Guerreiro et al., 2013; Patino et al., 2014; Canafoglia et al., 2015; Sato et al., 2016).

Table 3 above is the clinical information of Chinese CLN2, CLN5, CLN6, and CLN7 patients mentioned in this study. The initial symptoms in this study are all involved in muscle system and motor function. The onset ages of visual loss are all slightly later than most of the reported cases. The time for becoming bedridden and death is in the range of other reported cases.

Motor function disruption then progress to intellectual function (language and cognition disorders) are predominant in the clinical course of the Chinese patients in current study. Visual system is last affected. The disease course of patients in our study is particularly consistent. Another independent report of a Chinese CLN6 patient with homozygous mutation c.892G>A (p.E298K) presented a patient with uncoordinated movements and seizures at 1.5 years, then slow response and developmental milestones were observed. Visual loss was not observed when the boy was 5 years old at the last observation (Sun et al., 2018). Other studies on CLN2 and CLN5 also presented that visual decline is never the first symptom in Chinese LINCL patients (Chang et al., 2012; Ge et al., 2018). Normal vision was found in all three Chinese CLN5 patients in the study of Ge and

TABLE 2 | Clinical information of reported cases of CLN2, CLN5, CLN6, and CLN7 before.

	CLN2	CLN5	CLN6	CLN7
Onset age	1–6 (3.1)	4.5–8	1.5–8	2–7
Initial symptoms	Motor dysfunction, speech impairment	Visual failure, psychomotor retardation, clumsiness	Mental retardation, seizures	Seizures, ataxia, motor development delay
Onset age of visual loss	4	7	4	4
Time for becoming bedridden	5	13	5~second decade	Within 2 years after onset
Time for death	End of first decade	Few died at 15	9-mid 20 s	6.5–18

TABLE 3 | Clinical information of Chinese CLN2, CLN5, CLN6, and CLN7 patients in this study.

	CLN2	CLN5	CLN6	CLN7
Onset age	8 m–3 y	4.5–5 y	3–4 y	3.5
Initial symptoms	Seizures	Unsteady giant	Unsteady giant	Unsteady giant
Onset age of visual loss	5 y	12 y	8 y	5 y
Time for becoming bedridden	6 y	12 y	15 y	5 y
Time for death	7 y	–	16 y	–

his colleagues (Ge et al., 2018). These studies and ours suggest that the Chinese LINCL patients may have a consistent clinical course that is slightly different from the western LINCL patients. Expanded study on Chinese patients may shed more light on the observed difference.

Follow Up and Treatments

There is no cure in NCLs, and the treatments are limited to palliative care (Getty and Pearce, 2011). In this study, we used Topiramate to treat the CLN2 patient, Levetiracetam and sodium valproate to treat CLN5 patient, sodium valproate to treat CLN7 patient. These treatments did not generate desired outcome. We then reviewed the development of novel treatment, enzyme replacement therapy, stem cell transplantation and gene therapy.

Enzyme replacement therapy has been reported to be a more effective and safe way to treat with strong improvement observed in murine and canine models (Katz et al., 2014; Lu et al., 2015). A phase half clinical trial of the intracerebroventricular enzyme in CLN2 patients proved the safety of enzyme replacement treatment. Significant improvement of motor-language function was reported after treatment (<https://www.clinicaltrials.gov/ct2/show/results/NCT01907087>).

Stem cell transplantation is also considered to treat NCLs. However, hematopoietic stem cell transplantation did not perform ideally, only transient effect was observed in a few experiments (Lonnqvist et al., 2001; Yuza et al., 2005). Transplantation of neural stem cell performed better in murine model (Tamaki et al., 2009). Whereas in human, it did not change the neurological function or attenuates seizures (Selden et al., 2013).

Gene therapy was also tested. Human enzyme gene such as PPT1 or TPP1 was integrated into Adeno-Associated Virus 2 (AAV2) of AAV2/5 vector and intracranial injected into murine models or canine CLN2 models. Recombinant AAV-PPT1 successfully lowered the accumulation of autofluorescent

material, increased brain mass, slowed neurodegeneration and protected behavioral functions (Griffey et al., 2004, 2006; Roberts et al., 2012; Katz et al., 2015). Safety trial for gene therapy in human is still underway (<https://www.clinicaltrials.gov/ct2/show/NCT00151216>).

Individual II:2 of family 4 still has not reached the onset age. Disease symptoms are not presented. These novel therapies may give positive effect on this patient as all therapies performed better at the presymptomatic stage. It certainly requires genetic test to identify causal mutation carrier in order to perform any treatment before any symptoms observed.

CONCLUSION

This study described four Chinese LINCL siblings who were diagnosed by WES. The patients of these four families had similar disease courses started from motor regression or seizures to cognition regression and visual loss but carried mutations in different genes i.e. CLN2, CLN5, CLN6, and CLN7. The clinical features of LINCLs in these four Chinese siblings were not significantly different from those of Western patients. However, all Chinese LINCL patients in this study presented similar clinical course despite the affected genes. We assumed it as an ethnic specific clinical course according to our observation. Expanded sample size will be helpful to investigation of phenotype-genotype correlation. Besides, a platform for better communication, data and diagnostic experience sharing between Chinese and international clinicians is also required for further investigation (Jia and Shi, 2017).

Moreover, three mutations that detected in this study are novel mutations, and two of them occurred in intronic regions. These findings expanded the variant diversity of LINCLs.

ETHICS STATEMENT

This study was carried out is approved by Capital Medical University Beijing Children's Hospital Ethics Committee. The protocol was approved by the Capital Medical University Beijing Children's Hospital Ethics Committee. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

CONSENT FOR PUBLICATION

The patient's parents gave written informed consent to studies and publication of clinical information, images and sequencing data.

AUTHOR CONTRIBUTIONS

X-TR and X-HW designed the study. X-TR, X-HW, C-HD, W-HZ, J-WL, C-HR, and FF collected the clinical information of all patients. X-TR, X-HW, and C-HD collected the follow-up and prognosis information of all patients. XS and HZ performed the Next-Generation and Sanger Sequencing. X-TR, X-HW, XS, and HZ wrote the manuscript. X-TR, XS, HZ, and C-HD revised the manuscript. All authors listed have made a substantial, direct and intellectual contribution to the work and approved it for publication.

ACKNOWLEDGMENTS

We are grateful to all of the family members for their participation in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00370/full#supplementary-material>

Figure S1 | Next-Generation Sequencing results of 4 probands. (a) Proband 1 (Individual II:2 of family 1) is homozygous of mutation c.1551+1insTGAT. (b,c)

Proband 2 (Individual II:2 of family 2) carries 2 mutations in CLN5 gene, c.1068_1069del (b) and c.1100_1103del (c). (d,e) Proband 3 (Individual II:2 of family 3) carries 2 mutations in CLN6 gene, namely c.244G>T (d) and c.892G>A (e). (f,g) Proband 4 (Individual II:1 of family 4) carries 2 mutations in MFSD8 gene, c.1444C>T (f) and c.554-5A>G (g).

Figure S2 | Sanger Sequencing results of 4 probands. (A) Individual II:2 and II:3 of family 1 are the homozygous of mutation c.1551+1insTGAT. Their parents are all carriers of this mutation and don't have the disease or any symptoms, which is corresponding to the inheritance pattern. (B,C) Individual II:2 and II:3 are the compound heterozygous of mutations c.1068_1069del and c.1100_1103del and meanwhile suffered from LINCL. Their father (I:1) and sister (II:1) are the carriers of mutation c.1100_1103del, their mother is the carrier of mutation c.1068_1069del. Their parent and sister are all healthy with no regression which also corresponds to the inheritance pattern. (D,E) Individual II:1 and II:2 of family 3 are the compound heterozygous of mutation c.244G>T and c.892G>A, and these mutations are inherited from their mother and father. Their mother is the carrier of mutation c.244G>T and their father carries mutation c.892G>A. The two carrier parents are health with no disease symptom, while 2 heterozygous children are the patients of CLN6. (F,G) Individual II:1 and II:2 are compound heterozygous of mutations c.1444C>T and c.554-5A>G and their parents are the carriers of these mutations, mother with c.1444C>T and father with c.554-5A>G. Two parents don't have disease, proband individual II:1 of family 4 is the patient of NCL, which is true to the inheritance pattern. However, another heterozygous, individual II:2 of family 4, is asymptomatic. This is because individual is too young and doesn't reach the onset age of CLN7.

Table S1 | The precise information of patients.

REFERENCES

- Aiello, C., Terracciano, A., Simonati, A., Discepoli, G., Cannelli, N., Claps, D., et al. (2009). Mutations in MFSD8/CLN7 are a frequent cause of variant late infantile neuronal ceroid lipofuscinosis. *Hum. Mutat.* 30, E530–E540. doi: 10.1002/humu.20975
- Al-Muhaizea, M. A., Al-Hassnan, Z. N., and Chedrawi, A. (2009). Variant late infantile neuronal ceroid lipofuscinosis (CLN6 gene) in Saudi Arabia. *Pediatr. Neurol.* 41, 74–76. doi: 10.1016/j.pediatrneurol.2009.01.012
- Beltran, L., Valenzuela, G. R., Loos, M., Vargas, R., Lizama, R., Spinsanti, P., et al. (2018). Late-onset childhood neuronal ceroid lipofuscinosis: early clinical and electroencephalographic markers. *Epilepsy Res.* 144, 49–52. doi: 10.1016/j.eplepsyres.2018.05.005
- Benedict, J. W., Getty, A. L., Wishart, T. M., Gillingwater, T. H., and Pearce, D. A. (2009). Protein product of CLN6 gene responsible for variant late-onset infantile neuronal ceroid lipofuscinosis interacts with CRMP-2. *J. Neurosci. Res.* 87, 2157–2166. doi: 10.1002/jnr.22032
- Blom, T., Schmiedt, M. L., Wong, A. M., Kytala, A., Soronen, J., Jauhiainen, M., et al. (2013). Exacerbated neuronal ceroid lipofuscinosis phenotype in Cln1/5 double-knockout mice. *Dis. Model. Mech.* 6, 342–357. doi: 10.1242/dmm.010140
- Brunak, S., Engelbrecht, J., and Knudsen, S. (1991). Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J. Mol. Biol.* 220, 49–65.
- Canafoglia, L., Gilioli, I., Invernizzi, F., Sofia, V., Fugnanesi, V., Morbin, M., et al. (2015). Electroclinical spectrum of the neuronal ceroid lipofuscinoses associated with CLN6 mutations. *Neurology* 85, 316–324. doi: 10.1212/WNL.0000000000001784
- Cannelli, N., Garavaglia, B., Simonati, A., Aiello, C., Barzaghi, C., Pezzini, F., et al. (2009). Variant late infantile ceroid lipofuscinoses associated with novel mutations in CLN6. *Biochem. Biophys. Res. Commun.* 379, 892–897. doi: 10.1016/j.bbrc.2008.12.159
- Chang, X., Huang, Y., Meng, H., Jiang, Y., Wu, Y., Xiong, H., et al. (2012). Clinical study in Chinese patients with late-infantile form neuronal ceroid lipofuscinoses. *Brain Dev.* 34, 739–745. doi: 10.1016/j.braindev.2011.12.005
- Cismondì, I. A., Kohan, R., Ghio, A., Ramirez, A. M., and Halac, I. N. (2008). Gene symbol: CLN6. Disease: neuronal ceroid lipofuscinosis, late Infantile. *Hum. Genet.* 124:324.
- Cotman, S. L., Karaa, A., Staropoli, J. F., and Sims, K. B. (2013). Neuronal ceroid lipofuscinosis: impact of recent genetic advances and expansion of the clinicopathologic spectrum. *Curr. Neurol. Neurosci. Rep.* 13:366. doi: 10.1007/s11910-013-0366-z
- De Silva, B., Adams, J., and Lee, S. Y. (2015). Proteolytic processing of the neuronal ceroid lipofuscinosis related lysosomal protein CLN5. *Exp. Cell Res.* 338, 45–53. doi: 10.1016/j.yexcr.2015.08.021
- Desmet, F. O., Hamroun, D., Lalande, M., Collod-Beroud, G., Claustres, M., and Beroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res.* 37:e67. doi: 10.1093/nar/gkp215
- Eva, A., Jacob, J. C., Frederick, A., Stirling, C., Leonhard, W., Berkovic, S. F., et al. (1988). The Newfoundland aggregate of neuronal ceroid-lipofuscinosis. *Am. J. Med. Genet.* 31, 111–116. doi: 10.1002/ajmg.1320310615
- Ezaki, J., Takeda-Ezaki, M., and Kominami, E. (2000). Tripeptidyl peptidase I, the late infantile neuronal ceroid lipofuscinosis gene product, initiates the lysosomal degradation of subunit c of ATP synthase. *J. Biochem.* 128, 509–516. doi: 10.1093/oxfordjournals.jbchem.a022781
- Fang, F., Liu, Z., Fang, H., Wu, J., Shen, D., Sun, S., et al. (2017). The clinical and genetic characteristics in children with mitochondrial disease in China. *Sci. China Life Sci.* 60, 746–757. doi: 10.1007/s11427-017-9080-y
- Gao, H., Boustany, R. M., Espinola, J. A., Cotman, S. L., Srinidhi, L., Antonellis, K. A., et al. (2002). Mutations in a novel CLN6-encoded transmembrane protein cause variant neuronal ceroid lipofuscinosis in man and mouse. *Am. J. Hum. Genet.* 70, 324–335. doi: 10.1086/338190
- Ge, L., Li, H. Y., Hai, Y., Min, L., Xing, L., Min, J., et al. (2018). Novel mutations in CLN5 of Chinese patients with neuronal ceroid Lipofuscinosis. *J. Child Neurol.* 33, 837–850. doi: 10.1177/0883073818789024
- Getty, A. L., and Pearce, D. A. (2011). Interactions of the proteins of neuronal ceroid lipofuscinosis: clues to function. *Cell. Mol. Life Sci.* 68, 453–474. doi: 10.1007/s00018-010-0468-6
- Griffey, M., Bible, E., Vogler, C., Levy, B., Gupta, P., Cooper, J., et al. (2004). Adeno-associated virus 2-mediated gene therapy decreases autofluorescent storage material and increases brain mass in a murine model of infantile neuronal ceroid lipofuscinosis. *Neurobiol. Dis.* 16, 360–369. doi: 10.1016/j.nbd.2004.03.005
- Griffey, M. A., Wozniak, D., Wong, M., Bible, E., Johnson, K., Rothman, S. M., et al. (2006). CNS-directed AAV2-mediated gene therapy ameliorates functional

- deficits in a murine model of infantile neuronal ceroid lipofuscinosis. *Mol. Ther.* 13, 538–547. doi: 10.1016/j.ymthe.2005.11.008
- Guerreiro, R., Bras, J. T., Vieira, M., Warriar, V., Agrawal, S., Stewart, H., et al. (2013). CLN6 disease caused by the same mutation originating in Pakistan has varying pathology. *Eur. J. Paediatr. Neurol.* 17, 657–660. doi: 10.1016/j.ejpn.2013.04.011
- Haltia, M., and Goebel, H. H. (2013). The neuronal ceroid-lipofuscinoses: a historical introduction. *Biochim. Biophys. Acta* 1832, 1795–1800. doi: 10.1016/j.bbdis.2012.08.012
- Hebsgaard, S. M., Korning, P. G., Tolstrup, N., Engelbrecht, J., Rouze, P., and Brunak, S. (1996). Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. *Nucleic Acids Res.* 24, 3439–3452.
- Heine, C., Koch, B., Storch, S., Kohlschütter, A., Palmer, D. N., and Bräulke, T. (2004). Defective endoplasmic reticulum-resident membrane protein CLN6 affects lysosomal degradation of endocytosed arylsulfatase A. *J. Biol. Chem.* 279, 22347–22352. doi: 10.1074/jbc.M400643200
- Isosomppi, J., Vesa, J., Jalanko, A., and Peltonen, L. (2002). Lysosomal localization of the neuronal ceroid lipofuscinosis CLN5 protein. *Hum. Mol. Genet.* 11, 885–891. doi: 10.1093/hmg/11.8.885
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Jin, Y., Zhang, L., Ning, B., Hong, H., Xiao, W., Tong, W., et al. (2018). Application of genome analysis strategies in the clinical testing for pediatric diseases. *Pediatr. Investig.* 2, 72–81. doi: 10.1002/ped4.12044
- Kamate, M., Prashanth, G. P., and Hattiholi, V. (2012). Clinico-investigative profile of infantile and late-infantile neuronal ceroid lipofuscinoses. *Neurol. India* 60, 316–320. doi: 10.4103/0028-3886.98524
- Katz, M. L., Coates, J. R., Sibigroth, C. M., Taylor, J. D., Carpentier, M., Young, W. M., et al. (2014). Enzyme replacement therapy attenuates disease progression in a canine model of late-infantile neuronal ceroid lipofuscinosis (CLN2 disease). *J. Neurosci. Res.* 92, 1591–1598. doi: 10.1002/jnr.23423
- Katz, M. L., Tecedor, L., Chen, Y., Williamson, B. G., Lysenko, E., Wininger, F. A., et al. (2015). AAV gene transfer delays disease onset in a TPP1-deficient canine model of the late infantile form of Batten disease. *Sci. Transl. Med.* 7:313ra180. doi: 10.1126/scitranslmed.aac6191
- Kohan, R., Cannelli, N., Aiello, C., Santorelli, F., Cismondi, A., Mila, M., et al. (2008). Novel human pathological mutations. *Hum. Genet.* 123, 537–555. doi: 10.1007/s00439-008-0503-y
- Kollmann, K., Uusi-Rauva, K., Scifo, E., Tyynela, J., Jalanko, A., and Bräulke, T. (2013). Cell biology and function of neuronal ceroid lipofuscinosis-related proteins. *Biochim. Biophys. Acta* 1832, 1866–1881. doi: 10.1016/j.bbdis.2013.01.019
- Kousi, M., Siintola, E., Dvorakova, L., Vaskova, H., Turnbull, J., Topcu, M., et al. (2009). Mutations in CLN7/MFSD8 are a common cause of variant late-infantile neuronal ceroid lipofuscinosis. *Brain* 132, 810–819. doi: 10.1093/brain/awn366
- Kuizon, S., Dimaiuta, K., Walus, M., Jenkins, E. C. Jr., Kuizon, M., Kida, E., et al. (2010). A critical tryptophan and Ca²⁺ in activation and catalysis of TPPI, the enzyme deficient in classic late-infantile neuronal ceroid lipofuscinosis. *PLoS ONE* 5:e11929. doi: 10.1371/journal.pone.0011929
- Lebrun, A. H., Storch, S., Ruschendorf, F., Schmiedt, M. L., Kyttälä, A., Mole, S. E., et al. (2009). Retention of lysosomal protein CLN5 in the endoplasmic reticulum causes neuronal ceroid lipofuscinosis in Asian sibship. *Hum. Mutat.* 30, E651–E661. doi: 10.1002/humu.21010
- Lonnqvist, T., Vanhanen, S. L., Vetterranta, K., Autti, T., Rapola, J., Santavuori, P., et al. (2001). Hematopoietic stem cell transplantation in infantile neuronal ceroid lipofuscinosis. *Neurology* 57, 1411–1416. doi: 10.1212/WNL.57.8.1411
- Lu, J. Y., Nelvagal, H. R., Wang, L., Birnbaum, S. G., Cooper, J. D., and Hofmann, S. L. (2015). Intrathecal enzyme replacement therapy improves motor function and survival in a preclinical mouse model of infantile neuronal ceroid lipofuscinosis. *Mol. Genet. Metab.* 116, 98–105. doi: 10.1016/j.ymgme.2015.05.005
- Mink, J. W., Augustine, E. F., Adams, H. R., Marshall, F. J., and Kwon, J. M. (2013). Classification and natural history of the neuronal ceroid lipofuscinoses. *J. Child Neurol.* 28, 1101–1105. doi: 10.1177/0883073813494268
- Mole, S. E., and Williams, R. E. (2013). “Neuronal ceroid-lipofuscinoses,” in *GeneReviews*®, eds M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens, and A. Amemiya, A (Seattle, WA: University of Washington).
- Oishi, K., Ida, H., Kurosawa, K., and Eto, Y. (1999). Clinical and molecular analysis of Japanese patients with neuronal ceroid lipofuscinosis. *Mol. Genet. Metab.* 66, 344–348. doi: 10.1006/mgme.1999.2835
- Patino, L. C., Battu, R., Ortega-Recalde, O., Nallathambi, J., Anandula, V. R., Renukaradhya, U., et al. (2014). Exome sequencing is an efficient tool for variant late-infantile neuronal ceroid lipofuscinosis molecular diagnosis. *PLoS ONE* 9:e109576. doi: 10.1371/journal.pone.0109576
- Perez-Poyato, M. S., Mila-Recasens, M., Ferrer-Abizanda, I., Cusi-Sanchez, V., Vazquez-Lopez, M., Camino-Leon, R., et al. (2012). [Neuronal ceroid lipofuscinosis: diagnostic algorithm and clinical description of the Finnish (CLN5) and Turkish (CLN7) variants late infantile]. *Rev. Neurol.* 54, 544–550. doi: 10.33588/rn.5409.2011657
- Roberts, M. S., Macauley, S. L., Wong, A. M., Yilmaz, D., Hohm, S., Cooper, J. D., et al. (2012). Combination small molecule PPT1 mimetic and CNS-directed gene therapy as a treatment for infantile neuronal ceroid lipofuscinosis. *J. Inher. Metab. Dis.* 35, 847–857. doi: 10.1007/s10545-011-9446-x
- Sands, M. S. (2013). Considerations for the treatment of infantile neuronal ceroid lipofuscinosis (infantile Batten disease). *J. Child Neurol.* 28, 1151–1158. doi: 10.1177/0883073813495960
- Santavuori, P., Rapola, J., Nuutila, A., Raininko, R., Lappi, M., Launes, J., et al. (1991). The spectrum of Jansky-Bielschowsky disease. *Neuropediatrics* 22, 92–96. doi: 10.1055/s-2008-1071423
- Santavuori, P., Rapola, J., Sainio, K., and Raitta, C. (1982). A variant of Jansky-Bielschowsky disease. *Neuropediatrics* 13, 135–141. doi: 10.1055/s-2008-1059612
- Sato, R., Inui, T., Endo, W., Okubo, Y., Takezawa, Y., Anzai, M., et al. (2016). First Japanese variant of late infantile neuronal ceroid lipofuscinosis caused by novel CLN6 mutations. *Brain Dev.* 38, 852–856. doi: 10.1016/j.braindev.2016.04.007
- Schmiedt, M. L., Bessa, C., Heine, C., Ribeiro, M. G., Jalanko, A., and Kyttälä, A., (2010). The neuronal ceroid lipofuscinosis protein CLN5: new insights into cellular maturation, transport, and consequences of mutations. *Hum. Mutat.* 31, 356–365. doi: 10.1002/humu.21195
- Schulz, A., Kohlschütter, A., Mink, J., Simonati, A., and Williams, R. (2013). NCL diseases—clinical perspectives. *Biochim. Biophys. Acta* 1832, 1801–1806. doi: 10.1016/j.bbdis.2013.04.008
- Selden, N. R., Al-Uzri, A., Huhn, S. L., Koch, T. K., Sikora, D. M., Nguyen-Driver, M. D., et al. (2013). Central nervous system stem cell transplantation for children with neuronal ceroid lipofuscinosis. *J. Neurosurg. Pediatr.* 11, 643–652. doi: 10.3171/2013.3.PEDS12397
- Sharp, J. D., Wheeler, R. B., Parker, K. A., Gardiner, R. M., Williams, R. E., and Mole, S. E. (2003). Spectrum of CLN6 mutations in variant late infantile neuronal ceroid lipofuscinosis. *Hum. Mutat.* 22, 35–42. doi: 10.1002/humu.10227
- Shen, Y. (2018). Next-generation sequencing based molecular testing is an equalizer for diagnostic service of rare genetic disorders in China. *Pediatr. Investig.* 2, 96–97. doi: 10.1002/ped4.12036
- Siintola, E., Topcu, M., Aula, N., Lohi, H., Minassian, B. A., Paterson, A. D., et al. (2007). The novel neuronal ceroid lipofuscinosis gene MFSD8 encodes a putative lysosomal transporter. *Am. J. Hum. Genet.* 81, 136–146. doi: 10.1086/518902
- Steinfeld, R., Heim, P., Von Gregory, H., Meyer, K., Ullrich, K., Goebel, H. H., et al. (2002). Late infantile neuronal ceroid lipofuscinosis: quantitative description of the clinical course in patients with CLN2 mutations. *Am. J. Med. Genet.* 112, 347–354. doi: 10.1002/ajmg.10660
- Stogmann, E., El Tawil, S., Wagenstaller, J., Gaber, A., Edris, S., Abdelhady, A., et al. (2009). A novel mutation in the MFSD8 gene in late infantile neuronal ceroid lipofuscinosis. *Neurogenetics* 10, 73–77. doi: 10.1007/s10048-008-0153-1
- Sun, G., Yao, F., Tian, Z., Ma, T., and Yang, Z. (2018). A first CLN6 variant case of late infantile neuronal ceroid lipofuscinosis caused by a homozygous mutation in a boy from China: a case report. *BMC Med. Genet.* 19:177. doi: 10.1186/s12881-018-0690-x
- Tamaki, S. J., Jacobs, Y., Dohse, M., Capela, A., Cooper, J. D., Reitsma, M., et al. (2009). Neuroprotection of host cells by human central nervous system stem cells in a mouse model of infantile neuronal ceroid lipofuscinosis. *Cell Stem Cell* 5, 310–319. doi: 10.1016/j.stem.2009.05.022

- Taratuto, A. L., Saccoliti, M., Sevlever, G., Ruggieri, V., Arroyo, H., Herrero, M., et al. (1995). Childhood neuronal ceroid-lipofuscinoses in Argentina. *Am. J. Med. Genet.* 57, 144–149. doi: 10.1002/ajmg.1320570207
- Topcu, M., Tan, H., Yalnizoglu, D., Usubutun, A., Saatci, I., Aynaci, M., et al. (2004). Evaluation of 36 patients from Turkey with neuronal ceroid lipofuscinosis: clinical, neurophysiological, neuroradiological, and histopathologic studies. *Turk. J. Pediatr.* 46, 1–10.
- Warrier, V., Vieira, M., and Mole, S. E. (2013). Genetic basis and phenotypic correlations of the neuronal ceroid lipofuscinoses. *Biochim. Biophys. Acta* 1832, 1827–1830. doi: 10.1016/j.bbdis.2013.03.017
- Wlodawer, A., Li, M., Dauter, Z., Gustchina, A., Uchida, K., Oyama, H., et al. (2001). Carboxyl proteinase from *Pseudomonas* defines a novel family of subtilisin-like enzymes. *Nat. Struct. Biol.* 8, 442–446. doi: 10.1038/87610
- Wlodawer, A., Li, M., Gustchina, A., Oyama, H., Dunn, B. M., and Oda, K. (2003). Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases. *Acta Biochim. Pol.* 50, 81–102.
- Xin, W., Mullen, T. E., Kiely, R., Min, J., Feng, X., Cao, Y., et al. (2010). CLN5 mutations are frequent in juvenile and late-onset non-Finnish patients with NCL. *Neurology* 74, 565–571. doi: 10.1212/WNL.0b013e3181cff70d
- Yuza, Y., Yokoi, K., Sakurai, K., Ariga, M., Yanagisawa, T., Ohashi, T., et al. (2005). Allogenic bone marrow transplantation for late-infantile neuronal ceroid lipofuscinosis. *Pediatr. Int.* 47, 681–683. doi: 10.1111/j.1442-200x.2005.02126.x

Conflict of Interest Statement: XS and HZ were employed by company Running Gene Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Ren, Wang, Ding, Shen, Zhang, Zhang, Li, Ren and Fang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics

Ruili Huang^{1*}, Ivan Grishagin², Yuhong Wang¹, Tongan Zhao¹, Jon Greene², John C. Obenauer², Deborah Ngan¹, Dac-Trung Nguyen¹, Rajarshi Guha¹, Ajit Jadhav¹, Noel Southall¹, Anton Simeonov¹ and Christopher P. Austin¹

¹ Division of Pre-Clinical Innovation, National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, United States, ² Rancho BioSciences, San Diego, CA, United States

OPEN ACCESS

Edited by:

Weida Tong,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Yun Qian,
Shanghai Sixth People's Hospital,
China
Arun Samidurai,
Virginia Commonwealth University,
United States

*Correspondence:

Ruili Huang
huangru@mail.nih.gov

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 16 November 2018

Accepted: 08 April 2019

Published: 26 April 2019

Citation:

Huang R, Grishagin I, Wang Y, Zhao T, Greene J, Obenauer JC, Ngan D, Nguyen D-T, Guha R, Jadhav A, Southall N, Simeonov A and Austin CP (2019) The NCATS BioPlanet – An Integrated Platform for Exploring the Universe of Cellular Signaling Pathways for Toxicology, Systems Biology, and Chemical Genomics. *Front. Pharmacol.* 10:445. doi: 10.3389/fphar.2019.00445

Chemical genomics aims to comprehensively define, and ultimately predict, the effects of small molecule compounds on biological systems. Chemical activity profiling approaches must consider chemical effects on all pathways operative in mammalian cells. To enable a strategic and maximally efficient chemical profiling of pathway space, we have created the NCATS BioPlanet, a comprehensive integrated pathway resource that incorporates the universe of 1,658 human pathways sourced from publicly available, manually curated sources, which have been subjected to thorough redundancy and consistency cross-evaluation. BioPlanet supports interactive browsing, retrieval, and analysis of pathways, exploration of pathway connections, and pathway search by gene targets, category, and availability of corresponding bioactivity assay, as well as visualization of pathways on a 3-dimensional globe, in which the distance between any two pathways is proportional to their degree of gene component overlap. Using this resource, we propose a strategy to identify a minimal set of 362 biological assays that can interrogate the universe of human pathways. The NCATS BioPlanet is a public resource, which will be continually expanded and updated, for systems biology, toxicology, and chemical genomics, available at <http://tripod.nih.gov/bioplanet/>.

Keywords: BioPlanet, pathway, systems biology, chemical genomics, *in vitro* assay

INTRODUCTION

For most of its history, the field of toxicology has focused predominantly on whole-organism studies, with observable histological, behavioral, or developmental endpoints, or “apical endpoints,” being cataloged as occurring after exposure to chemicals. While whole-organism studies have served as the backbone of scientific and regulatory imperatives to protect human health, they suffer from lack of mechanistic insights, high cost, low throughput, and uncertain applicability to human risk assessment. However, unlike systems pharmacology and drug development, toxicology assessment has changed relatively little in the last 50 years (Kavlock et al., 2009; Hamburg, 2011)

due, in part, to the regulatory context in which most toxicological assessment takes place, and the human bias that (only) “seeing is believing.”

An example of a recently initiated effort to explore *in vitro* approaches to toxicology, the United States Tox21 program (National Research Council [NRC], 2007) was constituted in 2007 to utilize high-throughput *in vitro* testing and computational methods to transition toxicology into a predictive, mechanistic science (Collins et al., 2008; Kavlock et al., 2009; Tice et al., 2013). A collection of approximately 10,000 drugs and environmental chemicals (Attene-Ramos et al., 2013b) has been tested at 15 concentrations using a robotic platform (Inglese et al., 2006) in a wide variety of assays (Huang et al., 2016) with the initial focus on stress-response (Attene-Ramos et al., 2013a; Nishihara et al., 2015) and nuclear hormone receptor pathways (Hsu et al., 2014; Huang et al., 2014). However, given the protean nature of toxicological endpoints, and the lack of understanding of the molecular mechanism(s) that lead to most of these endpoints, characterization of the chemicals’ effects in a much broader set of assays will be required. Ideally, a set of assays could be selected or designed to measure targets that encompass all pathways that are relevant to toxicity. However, what constitutes a “toxicity pathway” is not clearly defined. A recent report (National Research Council [NRC], 2007) states that “toxicity pathways” are “cellular response pathways that, when sufficiently perturbed in an intact animal, are expected to result in adverse health effects.” This definition could potentially refer to all biological pathways, as our current understanding of the biological system is not sufficient for us to pinpoint the specific subset of pathways that fit this description. Molecular pathways are defined not only by their importance in normal physiology, but also by the disease or adverse events caused by their dysfunction. Since toxicological endpoints may potentially be caused by dysfunction of any pathway operative in human cells, mechanistic understanding and predictive signatures for all endpoints may ultimately require profiling of the Tox21 and/or other chemicals in a suite of assays that encompass all human pathways, representing a highly implausible scenario.

As a first step to enabling this goal, we aimed to develop a complete and non-redundant catalog of all human pathways, and construct an informatics platform to represent and browse the pathways, their healthy and disease state annotations, and targets within and relationships among them at varying levels of detail. Such a platform would enable the rational construction of a minimal set of assays that could be used to query all of pathway space experimentally, given that many pathways overlap and together form a network subsuming all cellular functions. Toward this goal, this platform can serve as a starting point for the systematic design of experiments to better understand how biological systems function. When linked with bioactivity data, the pathway data can be used to examine and predict the network effects of chemicals and other perturbations. Such a public resource would not only be critical to fulfilling the goals of *in vitro* toxicology efforts, but also provide fundamental values to the biomedical research community as a whole.

Existing pathway databases tend to focus on particular areas of biology, e.g., metabolism vs. signaling, and a comprehensive and uniform resource that covers all known pathways and

their annotations does not exist (Galperin and Cochrane, 2011; Galperin and Fernandez-Suarez, 2012). Moreover, information in many databases are computationally generated, e.g., HumanCyc¹, and not derived from direct experimental evidence, which is generally deemed more reliable. Other efforts that attempt to integrate individual resources, e.g., Pathway Commons (Cerami et al., 2011) simply combine data from various databases without further curation or validation of the information collected to remove redundancy or improve data quality. Different types of data are often mixed together with no distinctions made between, e.g., pathways and protein–protein interactions, experimental results and computational predictions, and no additional annotations are provided. Commercial pathway resources and tools are claimed to be more comprehensive (e.g., Ingenuity, GeneGo) (Thomas and Bonchev, 2010) yet the access by the research community to these products is hampered by the high cost. Our aim is to develop an open-source solution to enable researchers worldwide to access the tools and the data without encumbrance.

We report here the construction, features, and utilization of a comprehensive integrated and non-redundant pathway resource, the NCATS BioPlanet (**Figure 1**). The resource hosts information only from public sources that have been herein further manually curated to ensure the quality of the data. Along with our pathway warehouse, the NCATS BioPlanet software platform allows easy browsing and visualization of the universe of pathways, and exploration of associations among them. Additionally, we curated the set of annotated pathways in terms of the biological space covered and the current availability of assays, either commercial or academic, to probe each subspace. After eliminating redundancy across the pathway databases used to create the BioPlanet, we found that human cells incorporate 1,658 pathways. Starting with these pathways, we utilized a condensation approach to construct a minimal set of assays to cover all of pathway space. This minimal set of pathways will serve as the starting point to prioritize pathways for testing in a wide variety of systems biology efforts, and provides a reduced-complexity set for the systems pharmacology community. The NCATS BioPlanet will be continually updated and is publicly accessible at <http://tripod.nih.gov/bioplanet/>.

DATA, METHODS, AND RESULTS

Source Databases

Annotations for pathways and gene–gene or protein–protein interactions were obtained from a number of publically available databases, in which pathway annotations are also manually generated based on experimental observations to ensure the quality of our data sources. The locations and contents of these databases are listed in **Table 1**. Annotations of human disease genes were downloaded from the Online Mendelian Inheritance in Man (OMIM) database (McKusick, 1998). Gene target information for assays was extracted from PubChem bioassay descriptions (PubChem, 2010).

¹<http://humancyc.org/>

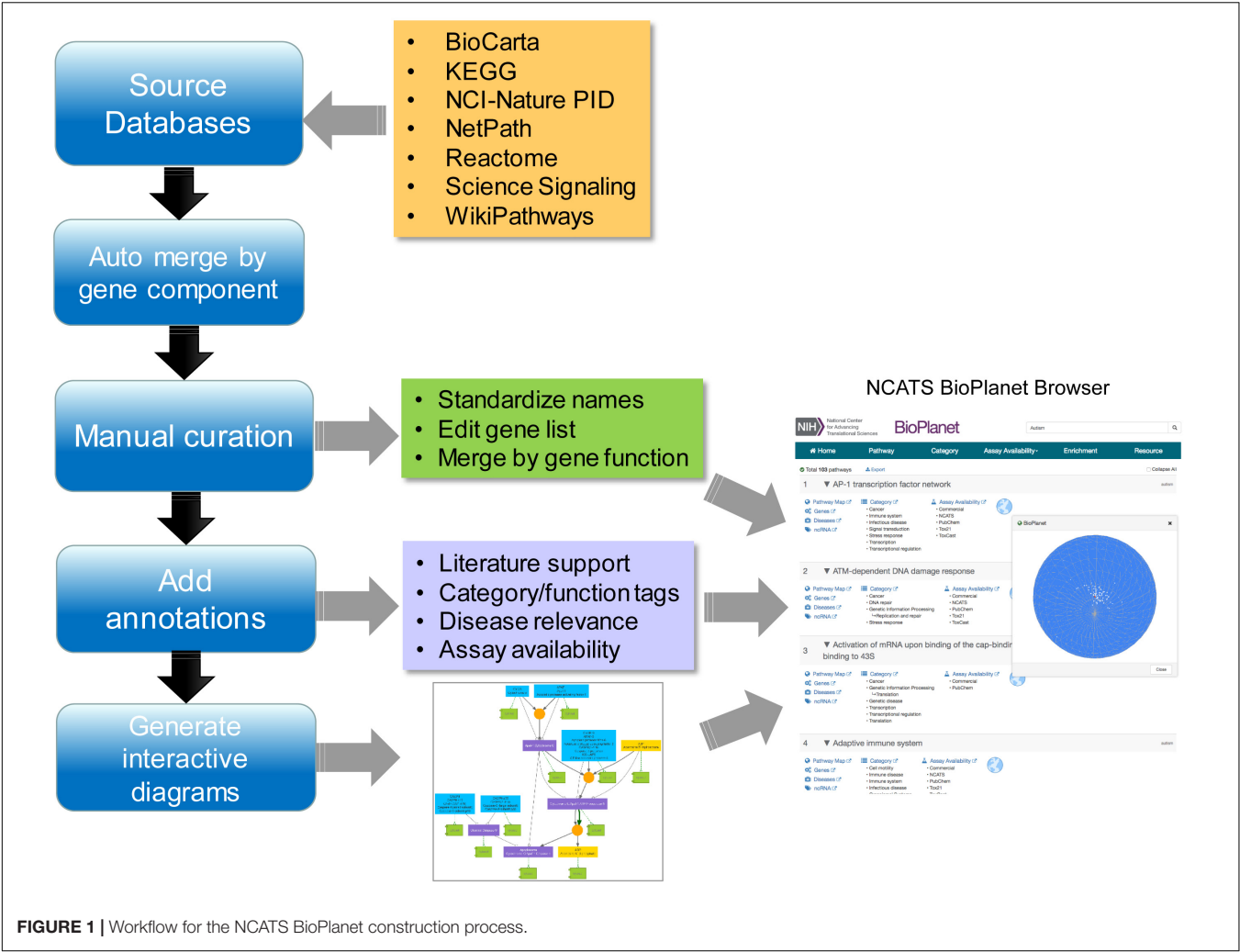


TABLE 1 | Pathway data sources.

Database	Number of human pathways	Number of genes	URL
KEGG – Kyoto Encyclopedia of Genes and Genomes	214	5520	http://www.genome.jp/kegg/pathway.html
BioCarta*	314	1494	https://cgap.nci.nih.gov/Pathways/BioCarta_Pathways
Reactome - a curated knowledgebase of biological pathways	1283	6125	http://www.reactome.org/
WikiPathways	204	4064	http://www.wikipathways.org/
NCI-Nature – Pathway Interaction Database*	722	3725	https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NCI_PID/
Science Signaling*	58	1234	http://stke.sciencemag.org/about/help/cm
NetPath	35	2877	http://www.netpath.org/

*Original database site is no longer supported. The URL provided here points to some data hosted at an alternative site.

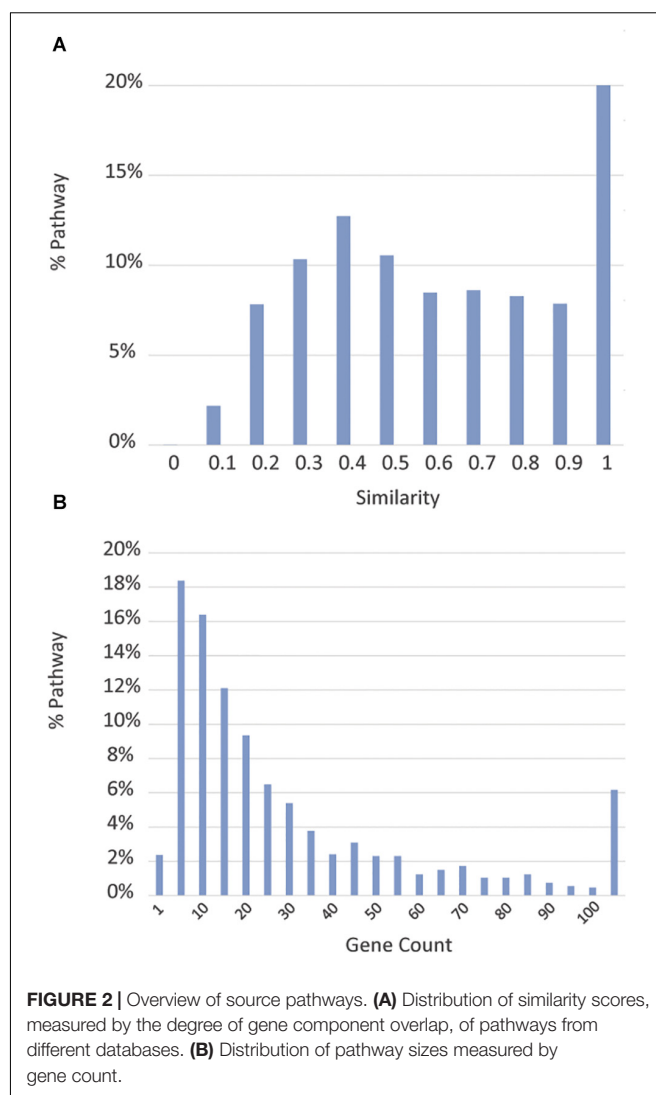
The present study focused on pathways annotating human genes. Different pathway sources focused on different aspects of the human biological system. KEGG is a large pathway database annotating over 5,500 human genes with a heavy focus on metabolism (KEGG , 2010). Over 50% of the KEGG pathways are metabolic pathways with the second largest pathway category, human diseases, making up only 14% of all KEGG pathways. The Science Signaling database (support ended in 2015) (Science Signaling, 2010) as its name indicates, is a collection of cell signaling pathways. Its pathway maps were generated based on information provided by scientists with expertise in a given field, deemed “pathway authorities,” thus

assuring the quality of the data. A result of collaborative efforts between the National Cancer Institute (NCI) and the Nature Publishing Group, the NCI-Nature Pathway Interaction Database (PID) (now retired) (NCI-Nature, 2010) is another source of curated human signaling pathways. Reactome is an open-source, curated pathway database that covers a variety of human biology including cell signaling, metabolism, human diseases and other fundamental biological processes, with some emphasis on signaling and metabolic pathways, comprising 18% and 17% of the pathways, respectively (Reactome, 2010). BioCarta pathway collection (no longer supported) operated as an open-source, community-fed forum with annotations collected on over ten different biological functions and processes, but with cell signaling as the primary category encompassing 32% of all BioCarta pathways (BioCarta, 2010). Similar to BioCarta, WikiPathways adopts the open source approach, as well, which takes input from the scientific community for the curation of biological pathways (WikiPathways, 2010). WikiPathways annotates over 4,000 human genes encompassing a range of pathway categories, including signaling (~30%) and metabolic (~10%) pathways.

Removing Redundancy

As expected, we found substantial overlaps among the pathway databases. To assess the extent of redundancy, we calculated a similarity score, defined as the ratio of genes shared between two pathways over the total number of unique genes contained in the two pathways, between each pathway and the pathway with which it has the highest gene component overlap. **Figure 2A** shows the distribution of these similarity scores. Approximately 23% of the pathways have at least one complete duplicate with identical gene components, and 31% of the pathways have at least one close match, with which they share over 90% of genes, in a different data source. Moreover, many pathways have only a few genes annotated. As shown in **Figure 2B**, about 20% of the pathways have ≤ 5 genes and 2.4% of the pathways only have one gene. Annotation of these pathways thus appears to be incomplete. For ease of downstream analysis, we chose to merge pathways with no significant difference in their gene components and exclude pathways with less than three genes whenever appropriate to minimize redundancy (see below for the procedure details). Utilizing these criteria, we found that there are 1,658 distinct pathways, encompassing 9,818 human genes, which constitute approximately 40% of all human genes. The number of pathway genes and the details of their relationships can be reasonably expected to change as functions of more genes are discovered and their interactions elucidated. Therefore, the content of the BioPlanet will be curated and updated periodically to reflect the updates from our data sources and to incorporate information from any new data sources that might emerge. As this project is constantly evolving, mistakes and incompleteness are inevitable and we have set up a mechanism for the scientific community to send us feedback and corrections to improve the quality of the BioPlanet as a public resource.

Two pathways were merged into one by merging their gene components when one of four criteria was met: (1) the overlap, defined as the number of genes shared by the two pathways



divided by the total number of unique genes in the two pathways, was $>90\%$; (2) the two pathways differ by only one gene; (3) one pathway has <3 genes and all of these genes are contained in the other pathway; (4) the two pathways have $>50\%$ overlap in their gene components and $p < 0.05$ (Fisher's exact test). The merging procedure was repeated until no two pathways met any of these criteria. After merging, the pathway gene lists were manually curated to correct mis-assigned genes and further remove redundancy (**Figure 1**; see below for detailed curation procedure) resulting in a final list of 1,658 distinct pathways.

Extensive Manual Curation

After initial merging, BioPlanet included 1,774 pathways that contained 10,040 unique genes, 9,928 of which were assigned to *Homo sapiens*. The pathway names were standardized and corrected for consistent capitalization, biological clarity, usage of Greek letters, and hyphenated terms. The genes in the pathways were also edited to remove withdrawn identifiers and replace obsolete ones. Non-human genes were removed or replaced with

corresponding human genes. However, even after the removal of non-human genes, a few pathways from mice and other species remained. Yet, despite the pathway names, all genes in these pathways were human. Therefore, to ameliorate this discrepancy, these pathways were renamed to remove the animal inference. Extremely small pathways that contained only one or two genes were merged with larger pathways, and some pathways with similar names and functions were merged. Eighty-nine sets of pathways had very similar names but different gene lists. For example, “Alzheimer’s disease” has 168 genes, and a separate pathway called “Alzheimer’s disease” has 82 genes, but some genes from the latter set are not among the genes in the former one. To resolve similarly named pathway sets like these, each set was manually examined to establish whether their gene lists had sufficiently similar functions to substantiate merging of these pathways, or whether the functions were sufficiently different, and the pathways should have been kept separate under distinctly different names. To make these decisions, the genes unique to each pathway were uploaded to the DAVID annotation resource². Using DAVID’s Functional Annotation Clustering tool, the top annotation cluster characterizing the gene list was used to determine the collective function of these genes. Gene Ontology Biological Process terms, KEGG pathways, and BioCarta pathways were preferred when available. Based on these results, pathway sets with gene lists that had sufficiently similar functions were merged. The pathways with gene lists that had distinct functions were preserved as separate pathways and renamed to distinguish them better. In total, 714 of the 1,774 pathway names (40%) were edited. Some pathways were removed or merged with other ones during the process, reducing the total number of pathways to 1,658. The number of unique genes represented in the pathways was reduced from 10,040 to 9,818.

Literature supporting the pathways and interactions were first added computationally. For the 303 pathways with no literature association found through the automated approach, references were sourced manually. GeneRif³ was used to link genes to literature references (PubMed IDs). Pathway names and gene lists were used to search PubMed to find pathway-literature linkages. PubMed IDs shared between genes and pathways were then identified to establish the gene-pathway association. An average of 50 abstracts from each method were spot-checked to ensure the method was producing the correct results. A total of 234,347 unique references were found for all 1,658 pathways, with each pathway having at least one reference. Further curation of the gene–gene interactions within each pathway is currently underway. Publications supporting the interactions selected by the pathway authors are retrieved from the source files and added to the BioPlanet pathways.

Pathway Tagging

Keyword tags were used to group functionally related pathways into categories. The GO Slim biological processes⁴, a small set of high-level functions characterizing an organism, were used

to generate the list of pathway tags. Some GO Slim terms were rejected for being too long (“anatomical structure formation involved in morphogenesis”), only applying to one pathway (“ribosome biogenesis”), or describing processes that do not exist in humans (“photosynthesis”). Disease-related tags were added based on the top-level disease categories at Disease Ontology⁵. Tags used by the source databases to group pathways were also collected for inclusion. Redundant tags were removed or merged with existing tags. A total of 51 tags were eventually selected, and grouped into seven categories: Major Systems, Cell Cycle, Genetic Information Processing, Metabolism, Development, Signaling, and Disease. The tags in each category are listed in **Table 2**.

GO annotations for human genes were used to tag many of the pathways automatically. The tag keywords were first matched manually to GO terms in the top 4 levels of the GO hierarchy. For each GO term, up to three tags were assigned. Most level-4 terms were not manually tagged unless they also occurred in a higher level. These terms were then associated with genes using the GO annotations, and the gene lists for each pathway were used to determine whether enough genes with one tag were present to assign that tag to the whole pathway. Specifically, we required that (1) at least 10% of the genes in the pathway have the same tag and (2) at least four genes have the same tag. The automated GO term method assigned at least one tag to 84% of the pathways. However, the GO term method missed some obvious tags suggested by the pathway titles. For example, “HIV-induced T cell apoptosis” would be expected to get tags for “Infectious disease” (“HIV” in the name), “Immune response” (“T cell”), and “Cell death” (“apoptosis”). For this reason, a second component was added to the automatic tagging algorithm. A list of keywords was created that would be expected to match each tag, and the occurrence of these keywords in the pathway title would assign the corresponding tags. For example, the tag “Nucleic acid metabolism” would be assigned if the pathway title contained words like “Nucleobase,” “Nucleotide,” “Nucleoside,” “Purine,” or “Pyrimidine.” This keyword method assigned at least one tag to 58% of the pathways. The combination of the two methods yielded 92% pathways with at least one tag. To measure how well the automated tagging process worked, 10 pathways were selected for manual review. The results showed that the automated process produced a high false positive (63%) and low false negative rate (30%). For this reason, we decided to manually review all of the tagged pathways, removing tags that seemed irrelevant and adding tags that were missed.

A manual workflow was then applied to add missing tags, remove false positive tags, and add disease tags to pathways. For each pathway, one or more summaries of the pathway were found from online scientific sources like PubMed, Entrez Gene (Maglott et al., 2005) and BioCarta, and the decisions to add or remove tags were based on these summaries. The Comparative Toxicogenomics Database (CTD)⁶ was used to find relevant disease associations. The list of gene IDs from the pathway was entered into CTD’s gene set analyzer and disease Venn diagram. The first method shows a list of diseases associated

²<http://david.abcc.ncifcrf.gov>

³<http://www.ncbi.nlm.nih.gov/gene/about-generif>

⁴<http://geneontology.org/docs/go-subset-guide/>

⁵<http://disease-ontology.org/>

⁶<http://ctdbase.org/>

with the input gene set, ranked by *p*-value, and the second method shows the overlap between the input gene set and the disease gene set. Rather than relying on a *p*-value threshold or minimum number of genes, high-ranking diseases in the list were accepted if they were consistent with the pathway description. This prevented the problem we observed in some of the automated tag assignments, when, for example, a subset of pathway genes may be involved in an Infectious Disease but the corresponding pathway is not primarily associated with any such Infectious Disease. After manual curation, all pathways have at least one tag assigned. The median number of tags per pathway is 5, while the maximum number is 15.

Assay Availability for Pathway Interrogation

Since one rationale for creating the BioPlanet is to enable the experimental assessment of chemical modulation of a wide range of human pathways, we next explored the current availability of extant bioassays to probe the 1,658 distinct human pathways. We examined bioassays from four sources, which cover 2,685 gene targets in total (in the order of decreasing priority): (1) assays from the Tox21 program that have been run at NCATS, (2) other NCATS assays, (3) other bioassays in PubChem, and (4) assays from commercial vendors not yet employed by Tox21, NCATS, or PubChem assay providers. Phenotypic assays with no specific gene targets were excluded from the analysis. **Figure 3A** shows the coverage of the 1,658 human pathways by assays from these four sources. If a pathway was covered by assays from multiple sources, only the source with the highest priority was counted. For example, if an assay was available from both Tox21 and PubChem, the pathway would be counted as covered by Tox21 in **Figure 3A** (see **Supplementary Figure S1** for the coverage of

the BioPlanet pathways by each individual source). All available assay sources for each pathway can be found in the BioPlanet database and browser. Here, to get an initial estimate, we have not made a distinction between assays that measure a specific gene target in a pathway, and pathway assays, i.e., assays that measure signaling throughout that pathway. We found that 88% of the pathways have at least one gene target with an assay available from one of the four assay sources, and 12% of the pathways do not have a bioassay available from these sources (**Supplementary Figure S2**). Of the four sources, the Tox21 assays cover 63% of the pathways; when combined with the other NCATS assays, these two sources cover 70% of the 1,658 pathways. Assays from other PubChem assay providers cover 12% of the pathways and we found other commercial assays for another 6% of the pathways. Recent developments in the field of precision medicine and RNA based therapeutics have highlighted the role of non-coding RNA (ncRNA) (Cech and Steitz, 2014) such as lncRNA (Volders et al., 2019), miRNA (Chou et al., 2018), and circRNA (Glazar et al., 2014) in healthy and disease conditions. When annotated by the availability of non-coding RNAs, we found that >99% of the BioPlanet pathways are regulated by at least one ncRNA (**Supplementary Figure S1**).

Next, we examined the assay availability for disease-related and non-related pathways (**Figures 3C,D**). Of the 1,658 human pathways, 97% contain at least one gene that is implicated in a genetic disease according to OMIM (**Figure 3B**). As of July 10, 2017, OMIM annotates 15,649 genes, including 6,013 phenotypes (usually diseases) that have been attributed to cognate genes⁷. Genes that cause genetic diseases have been identified in 97% of annotated pathways to date. Disease-related pathways

⁷<http://omim.org/statistics/entry>

TABLE 2 | Pathway tags.

Major systems	Metabolism	Signaling
Circulatory system	Nucleic acid metabolism	Cell signaling
Digestive system	Carbohydrate metabolism	G-protein coupled receptor
Endocrine system	Protein metabolism	Nuclear receptor
Excretory system	Lipid metabolism	Transcriptional regulation
Immune system	Vitamin and cofactor metabolism	Stress response
Musculoskeletal system	Small molecule metabolism	Environmental adaptation
Nervous system	Xenobiotic metabolism	Chronology
Sensory system	Energy metabolism	Transport
Genetic information processing	Protein folding, sorting, and degradation	
DNA replication	Protein modification	Disease
DNA repair		Cancer
Transcription		Cardiovascular disease
RNA processing	Cell cycle	Genetic disease
Translation	Cell cycle	Immune disease
	Cell growth	Infectious disease
	Cell death	Neurological disease
Development	Cell division	Physical disorder
Development	Cell proliferation	Endocrine and metabolic disease
Adhesion	Reproduction	Sepsis
Cell differentiation		Substance dependence
Cell motility		

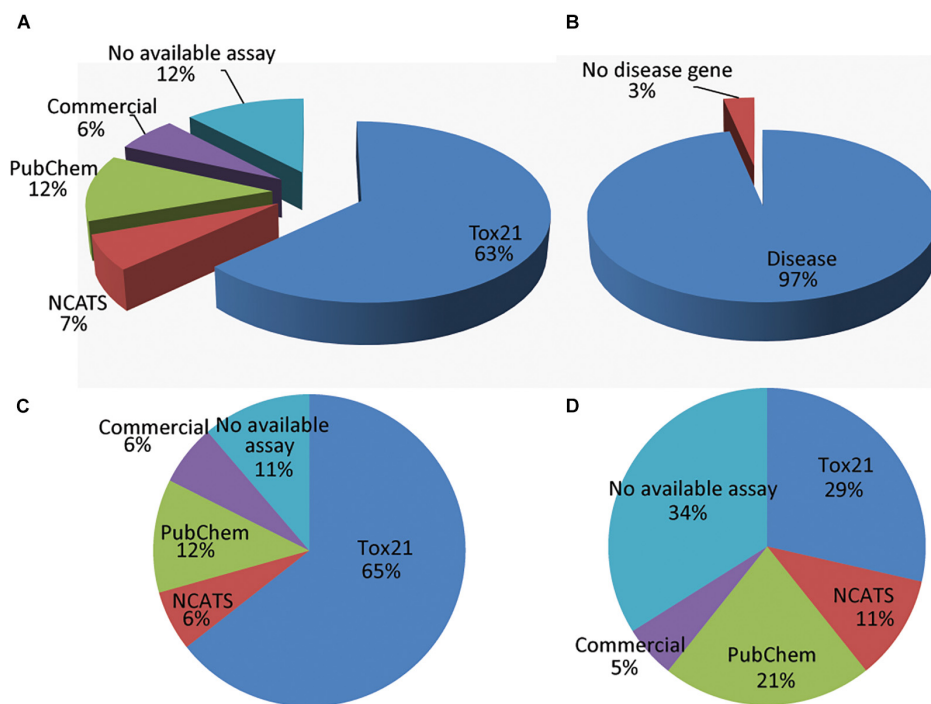


FIGURE 3 | Disease pathways and assay availability from four sources (in the order of decreasing priority): (1) Tox21, (2) NCATS, (3) PubChem, and (4) commercial vendors. If a pathway was covered by assays from multiple sources, only the source with the highest priority was counted. **(A)** Assay availability for all pathways. **(B)** Fraction of pathways that are disease pathways. **(C)** Assay availability for disease pathways. **(D)** Assay availability for non-disease pathways.

have significantly better assay coverage (89% have at least one bioassay) than pathways that do not contain any disease-related genes (66% have bioassays). Compared to the other PubChem assays, Tox21 and NCATS showed relatively better coverage of disease-related pathways (71% assays are from Tox21 or NCATS) than other pathways (only 40% assays are from Tox21 or NCATS). **Figure 4** shows the assay availability for different pathway categories. Cell signaling is by far the best-covered pathway category with 99% of the 488 pathways having a bioassay available. In contrast, metabolism, the second largest pathway category, has only 76% of the 351 pathways having a bioassay from the four assay sources. The human disease pathway category shown in **Figure 4** was not defined by having an OMIM gene, but from the annotations obtained from the pathway data sources. Nevertheless, 89% of these 103 pathways identified as human disease pathways have an available bioassay. In fact, 97% of metabolic pathways contain OMIM genes, which is almost the same as the percentage of signaling pathways containing OMIM genes (98%). This suggests that the apparent lack of interest in developing assays to probe metabolic pathways is unwarranted if the drive behind the wide interest in studying signaling pathways is their well acknowledged role in disease processes.

Probing the Pathway Universe With Minimum Number of Assays

The ultimate goal of systems pharmacology, of which the Tox21 program is an exemplar, is to characterize the activity of a

broad range of chemicals across the full spectrum of 1,658 human pathways. However, since performing 1,658 separate assays is experimentally unfeasible, and given that pathways are overlapping in their component genes and functions, and together constitute an interconnected network web, we reasoned that it should be possible to account for all of pathway space with a reduced number of assays that could cover multiple pathways. We thus sought to define a minimal set of gene targets that could be experimentally assayed and cover all of pathway space with some degree of overlap and redundancy to assure complete coverage.

We identified a minimum set of 362 genes that cover the entire list of 1,658 pathways (**Supplementary Table S1a**). Specifically, genes were first sorted by the number and size of pathways in which they participate, such that genes that appear in more pathways and smaller pathways were ranked higher. An iterative algorithm was then applied to go through the gene list collecting the highest ranked genes while keeping track of the pathways covered by the genes collected. The algorithm stopped when all pathways were covered and the 362 genes collected form the maximum coverage list. As most of these genes participate in multiple pathways, it is not surprising to find that this set of genes is significantly enriched (82 out of 362, $p < 1.0 \times 10^{-4}$) with genes that have been reported to be essential for the viability of human cells (Blomen et al., 2015; Fraser, 2015). When availability of assays in Tox21, NCATS, PubChem, or commercial sources was taken into account, that is, higher priority was assigned to genes with assays available in one of these three sources, a

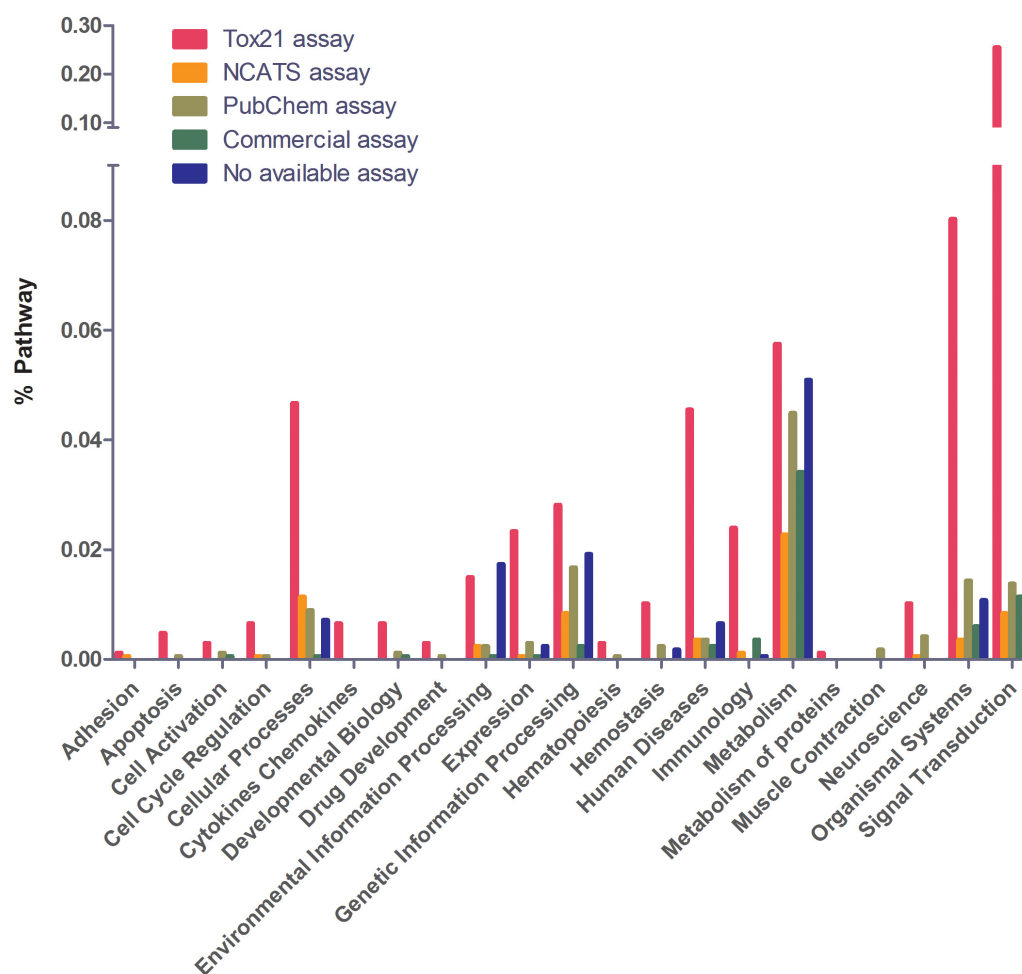


FIGURE 4 | Assay availability for different pathway categories. Cell signaling is the best covered category. Metabolism is the least covered category. Tox21 has the best coverage in human disease pathways.

minimum set of 411 genes was identified to cover all pathways (**Supplementary Table S1b**). More genes are required in this case because not all the genes that can cover the largest number of pathways have assays available, thus additional genes are needed to cover the same pathways.

The underlying premise of testing compounds in a reduced number of assays as a proxy for all biological pathway activity space is that it is possible to identify “indicator pathways” based on genes that regulate/participate in several pathways, such that activity in this indicator assay would allow inference that the compound would be active in other pathways that share this gene product. In this case, screening multiple pathway assays sharing the same gene target(s) would be redundant and thus unnecessary in a global assessment of compound activity on biological space. This premise predicts a positive correlation between the degree of compound activity overlap and the extent of gene sharing of two pathway assays. To test this prediction, we evaluated data generated from screening of the pilot phase Tox21 collection of 2,870 compounds against a set of 25 pathway assays (**Supplementary Table S2**). The

degrees of gene sharing and activity overlap were calculated for each pathway assay pair. Briefly, the degree of gene sharing between two pathways was defined as the ratio of genes shared by the two pathways over the total number of unique genes in the two pathways. The compound activity overlap between two assays was defined as the ratio of compounds active in both assays over the number of compounds active in either assay. A significant positive correlation was found ($r = 0.41$, $p < 1.0 \times 10^{-20}$), and the correlation improved to 0.57 when the degree of gene sharing between two pathway assays was $>20\%$. Though this correlation is statistically significant and supports the notion that achievement of a compound's comprehensive pathway activity footprint via testing in the full 1,658 pathways will be feasible, the extent to which pathway activity may be confidently inferred from activity in other “indicator” assays is unclear and will require experimental testing. One of the major goals of the Tox21 program and other systems biology initiatives is to generate and make public just these kinds of diverse pathway data and predictive algorithms, and experimentally test their

utility. As data are generated, they will be linked to the BioPlanet for straightforward browsing and correlation testing by others and ourselves.

The BioPlanet Pathway Browser

We report here what we believe to be the most comprehensive non-redundant enumeration to date of pathways extant in human cells, and the connections between them. To facilitate the browsing, visualization, and analysis of the pathway universe, we have constructed a unified database and a web-based software platform, the NCATS BioPlanet⁸, that is publicly available (**Figure 5**). From the main page, users may browse pathways by name, category, or assay availability (**Figure 5A**). The BioPlanet web browser also supports free text search enhanced by the availability of autocomplete suggestions as shown in **Figure 5B**. Users may search the BioPlanet by keywords, such as those that appear in a gene or pathway name, or a disease, or gene identifiers such as Entrez gene IDs. Batch search is also supported, that allows a user to paste in multiple gene IDs or keywords and retrieve their records via a single query. Search is performed on each individual search term as well as combinations of terms, and each pathway returned is labeled by the searched term(s) used to retrieve that pathway (**Figure 5C**). In the search results view, each pathway is labeled with functional category tags, disease relevance, and assay availability (**Figure 5C**). References to the original data sources are also provided. Each search result is a card that contains links to all pathway details: Pathway Map, Genes, ncRNAs, Diseases, Categories, and Assays.

In particular, Pathway Map is the most detailed graphical representation of a pathway demonstrating all known interactions between genes, proteins, nucleic acids, and small molecules in that pathway (**Figure 5D**). Importantly, these maps show the entirety of the pathway data stored in BioPAX or SBML formats obtained from public sources (*vide supra*), and curated, and thus provide the highest amount of detail known to date, without compromising the visual clarity. Moreover, this pathway diagram is searchable and interactive, where a click on each component will show a tooltip with known literature references and identifiers.

The browser also provides the mapping of pathways on a 3-dimensional globe, in which the distance between any two pathways on the globe surface is proportional to their degree of their gene component overlap (**Figure 5E**). This allows users to conduct a pathway similarity analysis at a glance, and demonstrates the interaction between different biological processes.

A gene enrichment analysis tool is also provided where the user can input a list of genes and determine which BioPlanet pathways are enriched in said list (**Figure 5F**).

DISCUSSION

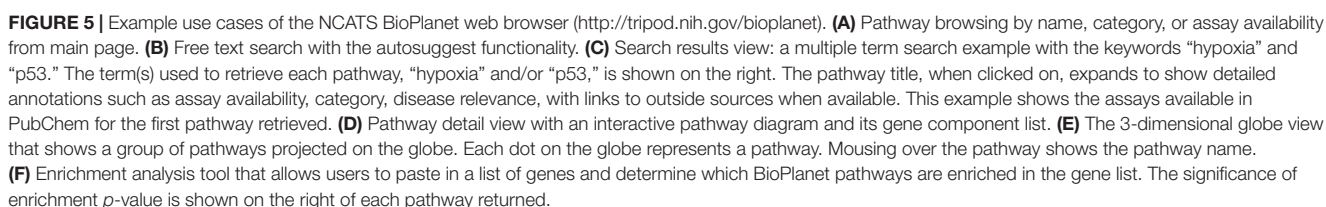
The Human Genome Project ushered in a continuing era of comprehensive enumeration of all biological system components.

Building on human and model organism reference genome sequences, comprehensive identification or production of genes (Collins et al., 1998, 2003), cDNAs (Strausberg et al., 1999; Gerhard et al., 2004; Temple et al., 2009), SNPs and haplotypes (International HapMap Consortium, 2003; Thorisson et al., 2005), structural and functional elements of genomes (ENCODE Project Consortium, 2004; Birney et al., 2007; Celniker et al., 2009), knockout mice (Austin et al., 2004), transcriptomes (Katayama et al., 2005), and microRNAs (He and Hannon, 2004; Bentwich et al., 2005) have been accomplished. Excellent efforts at enumeration of molecular, metabolic, and signaling pathways have been undertaken by multiple groups, but to date there has not been a synthesis of these efforts into a single collection of all pathways operant in human cells. The BioPlanet is the first attempt at creating such a resource, aiming to be comprehensive, non-redundant, relational, and easy to navigate.

Furthermore, the BioPlanet pathways are extensively annotated in terms of functional categories, disease relevance, assay availability and lncRNA regulation, which seems to be insufficient or lacking in various pathway databases. Using disease pathways as an example, most data sources we examined do not have explicit indications on which pathways have been associated with diseases. BioCarta and Reactome sort their pathways into several different categories but a general category for disease pathways is not available. KEGG is the only database with a “human diseases” category, but the genes listed in these human disease pathways only account for 27% of the OMIM disease genes. This shows that many pathways that might have disease relevance have not been explicitly annotated as such in previous pathway databases. Since one of our aims in creating the NCATS BioPlanet database was to enumerate a complete and non-redundant listing of all human disease-related pathways, we included the prevalence of disease genes as a principal feature in annotating all pathways in the BioPlanet. In addition, we manually examined and assigned a category to each pathway that did not have a category annotation in its source database. Furthermore, the complete and non-redundant feature of the BioPlanet would enable users to not only get a complete and concise interpretation of their experimental results from, e.g., genomic or proteomic screens, but also design an optimal set of targets or *in vitro* assays to comprehensively interrogate the biological space as detailed later below. This would not be possible with any other existing databases.

It is important to emphasize that BioPlanet, like other cataloging efforts before it, is an attempt to represent complex and often state-dependent systems in a uniform way and as such is subject to oversimplification. In addition, since the BioPlanet is built on a foundation of current understanding of pathways and their interconnections, there are undoubtedly errors in it, both representational and biological. We, therefore, view the BioPlanet as a work in progress, and designate the version currently available on our website (see text footnote 8) as BioPlanet 1.0 in recognition of its evolving nature. Like the data that went into creating the current version of BioPlanet, which was derived from the community of scientists worldwide, we view the ongoing curation and growth of the BioPlanet as a community “wiki” type effort, and therefore actively encourage

⁸<http://tripod.nih.gov/bioplanet>



comments, corrections, contributions, and suggestions for additional features via the BioPlanet page at <http://tripod.nih.gov/bioplanet>. All contributions will be acknowledged and attributed on this page.

We hope that the research community will find the BioPlanet useful, both for systems biology analyses as well as hypothesis generation. However, beyond its utility as a catalog, we hope that the BioPlanet will facilitate perturbation studies using small molecules, RNAi, gene knockouts, and other forms of biological modulation. The ultimate test of any network map is its ability to predict effects when a node in the network is perturbed. We will be adding capabilities for linking data from small molecule and siRNA screens performed at our Center to the next version of the BioPlanet, and we look forward to linking data obtained by other researchers as well. The current version of BioPlanet contains only human pathways, therefore, as a future endeavor, pathways for other species will be added both for their own importance in biological research and in comparison to their human counterparts, since human-animal pathway differences are likely drivers of non-concordance of chemical effects on humans and animals.

In the nearest term, the BioPlanet will find utility in the selection of *in vitro* assays to strengthen predictive toxicology methods (National Research Council [NRC], 2007; Kavlock et al., 2009). An underlying premise of *in vitro* toxicology approaches is that any pathway which plays an important role in human physiology could, if sufficiently perturbed, yield pathophysiology, i.e., toxicity. As we have demonstrated recently, an optimally designed panel of *in vitro* assays with targets diverse enough to sufficiently cover the biological response space could achieve good performance in predicting *in vivo* human toxicity, such as adverse drug effects (Huang et al., 2016, 2018). The BioPlanet would be an ideal guiding tool in designing such an assay panel. By analogy to genome-wide association studies (GWAS), we might refer to the present *in vitro* toxicology approaches, such as Tox21, as “pathway-ome-wide activity study,” or PWAS. Like GWAS studies, in which querying of all polymorphisms in the genomes of thousands of participants has been considered impractical, PWAS of all 1,658 pathways across thousands of chemicals is similarly difficult: by way of example, a 15-point concentration-response quantitative high throughput (qHTS) screen of the Tox21 “10K” set requires at least one week for each assay even with the ultrahigh-throughput robotic platform being utilized. GWAS studies were rendered practical by the comprehensive cataloging of SNPs and the discovery of the SNPs that are inherited together in haplotype blocks, thus allowing the imputation of SNPs not directly tested via the presence of a reduced numbers of “tag SNPs.” While there are 1,658 total pathways currently characterized, our analysis suggests that assaying only 362 will allow the imputation of activity in the remaining pathways. Importantly, this reduced number, while consistent with current data, will require ongoing data production to test and refine this very concept and the actual number of independent assays required to adequately query all of pathway activity space, with such data being provided in PubChem and other public-facing portals on a continuing basis.

While BioPlanet was initially conceived as a tool to guide systems toxicology efforts, it has implications and applications across the spectrum of systems biology, systems pharmacology, and disease pathophysiology. We look forward to continuing to collaborate with the research community to further develop and populate the BioPlanet, and thus achieve its potential as a resource for discovery.

AUTHOR CONTRIBUTIONS

RH coordinated the project, sourced and compiled pathway lists to construct the BioPlanet database, helped with data curation, helped to build the BioPlanet database and browser, performed statistical analysis of all data, and wrote the manuscript. TZ and YW built the BioPlanet database and browser. IG, JG, and JO curated all data and wrote the manuscript. IG generated all pathway diagrams. DN helped with assay vender searching and data curation. D-TN helped to extract data from PubChem and build the BioPlanet browser. RG helped with data curation and the BioPlanet browser. AJ coordinated the data curation. NS helped to coordinate data curation and construction of the BioPlanet browser. AS directed the project and wrote the manuscript. CA conceived and directed the project and wrote the manuscript. All authors reviewed the manuscript.

FUNDING

This work was supported by the Intramural Research Program of the National Toxicology Program (Interagency Agreement #Y2-ES-7020-01), National Institute of Environmental Health Sciences, the United States Environmental Protection Agency (Interagency Agreement #Y3-HG-7026-03), and the National Center for Advancing Translational Sciences, National Institutes of Health. The views expressed in this article are those of the authors and do not necessarily reflect the statements, opinions, views, conclusions, or policies of the National Center for Advancing Translational Sciences, National Institutes of Health, or the United States Government. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ACKNOWLEDGMENTS

We would like to thank Dr. Mikyung Lee for assistance in generating pathway diagrams and Drs. David Gerhold and Matthew Hall for helpful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00445/full#supplementary-material>

REFERENCES

- Attene-Ramos, M. S., Huang, R., Sakamuru, S., Witt, K. L., Beeson, G. C., Shou, L., et al. (2013a). Systematic study of mitochondrial toxicity of environmental chemicals using quantitative high throughput screening. *Chem. Res. Toxicol.* 26, 1323–1332. doi: 10.1021/tx4001754
- Attene-Ramos, M. S., Miller, N., Huang, R., Michael, S., Itkin, M., Kavlock, R. J., et al. (2013b). The Tox21 robotic platform for the assessment of environmental chemicals - from vision to reality. *Drug Discov. Today* 18, 716–723. doi: 10.1016/j.drudis.2013.05.015
- Austin, C. P., Battey, J. F., Bradley, A., Bucan, M., Capecchi, M., Collins, F. S., et al. (2004). The knockout mouse project. *Nat. Genet.* 36, 921–924. doi: 10.1038/ng0904-921
- Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* 37, 766–770. doi: 10.1038/ng1590
- BioCarta (2010). *BioCarta - Charting Pathways of Life*. Available at: <http://www.biocarta.com/genes/index.asp> (accessed February 12, 2009).
- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.
- Blomen, V. A., Majek, P., Jae, L. T., Bigenzahn, J. W., Nieuwenhuis, J., and Staring, J. (2015). Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096. doi: 10.1126/science.aac7557
- Cech, T. R., and Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157, 77–94. doi: 10.1016/j.cell.2014.03.008
- Celniker, S. E., Dillon, L. A., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., et al. (2009). Unlocking the secrets of the genome. *Nature* 459, 927–930.
- Cerami, E. G., Gross, B. E., Demir, E., Rodchenkov, I., Babur, O., Anwar, N., et al. (2011). Pathway commons, a web resource for biological pathway data. *Nucleic Acids Res.* 39, D685–D690. doi: 10.1093/nar/gkq1039
- Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., and Liao, K. W. (2018). miRTarBase update: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46, D296–D302. doi: 10.1093/nar/gkx1067
- Collins, F. S., Gray, G. M., and Bucher, J. R. (2008). Toxicology. Transforming environmental health protection. *Science* 319, 906–907. doi: 10.1126/science.1154619
- Collins, F. S., Morgan, M., and Patrinos, A. (2003). The human genome project: lessons from large-scale biology. *Science* 300, 286–290. doi: 10.1126/science.1084564
- Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. (1998). New goals for the U.S. human genome project: 1998–2003. *Science* 282, 682–689. doi: 10.1126/science.282.5389.682
- ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA elements) project. *Science* 306, 636–640. doi: 10.1126/science.1105136
- International HapMap Consortium (2003). The international HapMap project. *Nature* 426, 789–796. doi: 10.1038/nature02168
- Fraser, A. (2015). Essential human genes. *Cell Syst.* 1, 381–382. doi: 10.1016/j.cels.2015.12.007
- Galperin, M. Y., and Cochrane, G. R. (2011). The 2011 nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Res.* 39, D1–D6. doi: 10.1093/nar/gkq1243
- Galperin, M. Y., and Fernandez-Suarez, X. M. (2012). The 2012 nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Res.* 40, D1–D8. doi: 10.1093/nar/gkr1196
- Gerhard, D. S., Wagner, L., Feingold, E. A., Shenmen, C. M., Grouse, L. H., and Schuler, G. (2004). The status, quality, and expansion of the NIH full-length cDNA project: the mammalian gene collection (MGC). *Genome Res.* 14, 2121–2127. doi: 10.1101/gr.2596504
- Glazar, P., Papavasileiou, P., and Rajewsky, N. (2014). circBase: a database for circular RNAs. *RNA* 20, 1666–1670. doi: 10.1261/rna.043687.113
- Hamburg, M. A. (2011). Advancing regulatory science. *Science* 331:987. doi: 10.1126/science.1204432
- He, L., and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531. doi: 10.1038/nrg1379
- Hsu, C. W., Zhao, J., Huang, R., Hsieh, J. H., Hamm, J., Chang, X., et al. (2014). Quantitative high-throughput profiling of environmental chemicals and drugs that modulate farnesoid X receptor. *Sci. Rep.* 4:6437. doi: 10.1038/srep06437
- Huang, R., Sakamuru, S., Martin, M. T., Reif, D. M., Judson, R. S., and Houck, K. A. (2014). Profiling of the Tox21 10K compound library for agonists and antagonists of the estrogen receptor alpha signaling pathway. *Sci. Rep.* 4:5664. doi: 10.1038/srep05664
- Huang, R., Xia, M., Sakamuru, S., Zhao, J., Lynch, C., Zhao, T., et al. (2018). Expanding biological space coverage enhances the prediction of drug adverse effects in human using in vitro activity profiles. *Sci. Rep.* 8:3783. doi: 10.1038/s41598-018-22046-w
- Huang, R., Xia, M., Sakamuru, S., Zhao, J., Shahane, S. A., Attene-Ramos, M., et al. (2016). Modelling the Tox21 10 K chemical profiles for in vivo toxicity prediction and mechanism characterization. *Nat. Commun.* 7:10425. doi: 10.1038/ncomms10425
- Inglese, J., Auld, D. S., Jadhav, A., Johnson, R. L., Simeonov, A., Yasgar, A., et al. (2006). Quantitative high-throughput screening: a titration-based approach that efficiently identifies biological activities in large chemical libraries. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11473–11478. doi: 10.1073/pnas.0604348103
- Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., et al. (2005). Antisense transcription in the mammalian transcriptome. *Science* 309, 1564–1566. doi: 10.1126/science.1112009
- Kavlock, R. J., Austin, C. P., and Tice, R. R. (2009). Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal.* 29, 485–487; discussion 492–497.
- KEGG (2010). *Kyoto Encyclopedia of Genes and Genomes Pathway Database*. Available at: <http://www.genome.jp/kegg/pathway.html> (accessed July 22, 2015).
- Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2005). Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.* 33, D54–D58.
- McKusick, V. A. (1998). *Mendelian Inheritance in Man*, 12 Edn. Baltimore, MD: Johns Hopkins University Press.
- National Research Council [NRC] (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: The National Academies Press.
- NCI-Nature (2010). *NCI-Nature Pathway Interaction Database*. Available at: <http://pid.nci.nih.gov/> (accessed September 8, 2015).
- Nishihara, K., Huang, R., Zhao, J., Shahane, S. A., Witt, K. L., Smith-Roe, S. L., et al. (2015). Identification of genotoxic compounds using isogenic DNA repair deficient DT40 cell lines on a quantitative high throughput screening platform. *Mutagenesis* 31, 69–81. doi: 10.1093/mutage/gev055
- PubChem (2010). *PubChem Bioassay*. Available at: <http://www.ncbi.nlm.nih.gov/pcassay> (accessed June 29, 2017).
- Reactome (2010). *Reactome - A Curated Knowledgebase of Biological Pathways*. Available at: <http://www.reactome.org/> (accessed September 14, 2015).
- Science Signaling (2010). *Database of Cell Signaling*. Available at: <http://stke.sciencemag.org/cm/> (accessed September 18, 2015).
- Strausberg, R. L., Feingold, E. A., Klausner, R. D., and Collins, F. S. (1999). The mammalian gene collection. *Science* 286, 455–457. doi: 10.1126/science.286.5439.455
- Temple, G., Gerhard, D. S., Rasooly, R., Feingold, E. A., Good, P. J., and Robinson, C. (2009). The completion of the mammalian gene collection (MGC). *Genome Res.* 19, 2324–2333. doi: 10.1101/gr.095976.109
- Thomas, S., and Bonchev, D. (2010). A survey of current software for network analysis in molecular biology. *Hum. Genomics* 4, 353–360.

- Thorisson, G. A., Smith, A. V., Krishnan, L., and Stein, L. D. (2005). The international HapMap project web site. *Genome Res.* 15, 1592–1593. doi: 10.1101/gr.4413105
- Tice, R. R., Austin, C. P., Kavlock, R. J., and Bucher, J. R. (2013). Improving the human hazard characterization of chemicals: a Tox21 update. *Environ. Health Perspect.* 121, 756–765. doi: 10.1289/ehp.1205784
- Volders, P. J., Anckaert, J., Verheggen, K., Nuytens, J., Martens, L., Mestdagh, P., et al. (2019). LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* 47, D135–D139. doi: 10.1093/nar/gky1031
- WikiPathways (2010). *WikiPathways - Pathways for the People*. Available at: <http://www.wikipathways.org/index.php/WikiPathways> (accessed September 10, 2015).

Conflict of Interest Statement: IG, JG, and JO were employed by company Rancho BioSciences.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Huang, Grishagin, Wang, Zhao, Greene, Obenauer, Ngan, Nguyen, Guha, Jadhav, Southall, Simeonov and Austin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A New *SLC10A7* Homozygous Missense Mutation Responsible for a Milder Phenotype of Skeletal Dysplasia With Amelogenesis Imperfecta

Virginie Laugel-Haushalter^{1*}, Séverine Bär², Elise Schaefer^{1,3}, Corinne Stoetzel¹, Véronique Geoffroy¹, Yves Alembik³, Naji Kharouf^{4,5}, Mathilde Huckert⁴, Pauline Hamm⁴, Joseph Hemmerlé^{4,5}, Marie-Cécile Manière^{4,6}, Sylvie Friant², Hélène Dollfus^{1,3,7} and Agnès Bloch-Zupan^{4,6,8,9*}

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Muhammad Jawad Hassan,
National University of Medical
Sciences (NUMS), Pakistan
Maria Paola Lombardi,
University of Amsterdam,
Netherlands

*Correspondence:

Virginie Laugel-Haushalter
virginie.laugel@gmail.com
Agnès Bloch-Zupan
agnes.bloch-zupan@unistra.fr

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 07 February 2019

Accepted: 07 May 2019

Published: 28 May 2019

Citation:

Laugel-Haushalter V, Bär S, Schaefer E, Stoetzel C, Geoffroy V, Alembik Y, Kharouf N, Huckert M, Hamm P, Hemmerlé J, Manière M-C, Friant S, Dollfus H and Bloch-Zupan A (2019) A New *SLC10A7* Homozygous Missense Mutation Responsible for a Milder Phenotype of Skeletal Dysplasia With Amelogenesis Imperfecta. *Front. Genet.* 10:504. doi: 10.3389/fgene.2019.00504

¹Laboratoire de Génétique Médicale, UMR_S INSERM U1112, Faculté de Médecine, FMTS, Institut Génétique Médicale d'Alsace (IGMA), Université de Strasbourg, Strasbourg, France, ²Laboratoire de Génétique Moléculaire, Génomique, Microbiologie (GMGM), UMR7156, Centre National de Recherche Scientifique (CNRS), Université de Strasbourg, Strasbourg, France, ³Service de Génétique Médicale, Hôpitaux Universitaires de Strasbourg, IGMA, Strasbourg, France, ⁴Faculté de Chirurgie Dentaire, Université de Strasbourg, Strasbourg, France, ⁵Laboratoire de Biomatériaux et Bioingénierie, Inserm UMR_S 1121, Strasbourg, France, ⁶Pôle de Médecine et Chirurgie Bucco-dentaires, Hôpital Civil, Centre de référence des maladies rares orales et dentaires, O-Rares, Filière Santé Maladies rares TETE COU, European Reference Network ERN CRANIO, Hôpitaux Universitaires de Strasbourg (HUS), Strasbourg, France, ⁷Centre de Référence pour les affections rares en génétique ophtalmologique, CARGO, Filière SENSEGENE, Hôpitaux Universitaires de Strasbourg, Strasbourg, France, ⁸Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), INSERM U1258, CNRS-UMR7104, Université de Strasbourg, Illkirch-Graffenstaden, France, ⁹Eastman Dental Institute, University College London, London, United Kingdom

Amelogenesis imperfecta (AI) is a heterogeneous group of rare inherited diseases presenting with enamel defects. More than 30 genes have been reported to be involved in syndromic or non-syndromic AI and new genes are continuously discovered (Smith et al., 2017). Whole-exome sequencing was performed in a consanguineous family. The affected daughter presented with intra-uterine and postnatal growth retardation, skeletal dysplasia, macrocephaly, blue sclerae, and hypoplastic AI. We identified a homozygous missense mutation in exon 11 of *SLC10A7* (NM_001300842.2: c.908C>T; p.Pro303Leu) segregating with the disease phenotype. We found that *Slc10a7* transcripts were expressed in the epithelium of the developing mouse tooth, bones undergoing ossification, and in vertebrae. Our results revealed that *SLC10A7* is overexpressed in patient fibroblasts. Patient cells display altered intracellular calcium localization suggesting that *SLC10A7* regulates calcium trafficking. Mutations in this gene were previously reported to cause a similar syndromic phenotype, but with more severe skeletal defects (Ashikov et al., 2018; Dubail et al., 2018). Therefore, phenotypes resulting from a mutation in *SLC10A7* can vary in severity. However, AI is the key feature indicative of *SLC10A7* mutations in patients with skeletal dysplasia. Identifying this important phenotype will improve clinical diagnosis and patient management.

Keywords: skeletal dysplasia, amelogenesis imperfecta, NGS (next generation sequencing), human, rare diseases

INTRODUCTION

Skeletal dysplasias are a heterogeneous group of diseases affecting bone and cartilage formation resulting in short stature. These diseases are associated with shorter long bones and abnormal shape and/or size of the skeleton, spine, and head and, eventually other anomalies (including neurological, cardiac, and respiratory defects). Distinguishing individual pathologies in this wide group of diseases is improving thanks to advances in genomic technologies and genetic analyses (Alanay and Lachman, 2011).

Enamel and bone formation share a common mineralization process involving the precipitation of inorganic hydroxyapatite nanocrystals within organic matrices to form biological structures (Abou Neel et al., 2016). When fundamental mineralization processes are disrupted, skeletal dysplasia-associated syndromes can include enamel alterations.

Amelogenesis imperfecta (AI) is a heterogeneous group of rare inherited diseases presenting with anomalies in dental enamel formation. More than 30 genes are known to be involved in AI, and new genes are continuously being discovered (Smith et al., 2017).

Here, we report the case of a patient with a mutation in *SLC10A7*, a potential calcium transporter, presenting with skeletal dysplasia associated with AI. We identified the first *SLC10A7* missense mutation occurring at the end of the protein (p.Pro303Leu) leading to milder skeletal anomalies compared to previously described patients with mutations in this gene (Ashikov et al., 2018; Dubail et al., 2018).

MATERIALS AND METHODS

Patient

The patient was examined at the Centre de Référence des Maladies Rares orales et dentaires (O-Rares), Strasbourg, France and at the Centre de Compétence des Maladies Osseuses Constitutionnelles (OSCAR) in Strasbourg. The oral phenotype was documented using the D[4]/phenodent registry protocol¹. This clinical study is registered at <https://clinicaltrials.gov>: NCT01746121/NCT02397824 and with the French Ministry of Higher Education and Research Bioethics Commission as a biological collection “Oral Manifestations of Rare Diseases” DC-2012-1677/DC-2012-1002 and was acknowledged by the person protection committee. The parents gave written informed consents for the genetic analyses performed on the salivary samples both for them and their children in accordance with the Declaration of Helsinki. They also gave written consent for the publication of this case report and the images of their daughter presented in Figure 1.

Electron Microscopy

Teeth were washed, stored in 70% ethanol at 4°C, embedded in Epon 812 (Euromedex, France), sectioned, and polished.

Sections were then etched with 20% (w/w) citric acid, dehydrated, and analyzed by using an optical numeric microscope (KEYENCE, Japan) and a Quanta 250 FEG scanning electron microscope (SEM) (FEI Company, the Netherlands). Specimens also underwent chemical analysis (see **Supplementary Methods**).

Whole-Exome Sequencing

Whole-exome sequencing (WES) was performed on the affected patient (II.4) and her parents (I.1 and I.2) by Integragen (Evry, France, 2014). Exons were captured using SureSelect Human All Exon Kits (Agilent, France) with the company's probe library (Agilent Human All Exon v5 + UTR 75 Mb Kit) and sequenced with an Illumina HISEQ2000 (Illumina, USA) as paired-end 75 bp reads, resulting in an average coverage of 80X. The sequence reads were then aligned to the reference sequence of the human genome (GRCh37) (see **Supplementary Methods**).

Bioinformatics Analysis

Annotation and ranking of SNV/indel were performed by VaRank (Geoffroy et al., 2015) in combination with the Alamut Batch software (Interactive Biosoftware, France) (see **Supplementary Methods**).

Sample Collection and Sanger Sequencing and Segregation

Saliva samples from the affected daughter, her unaffected parents, and siblings were collected (I.1, I.2, II.4, II.1, II.2, and II.3). The amplification of the region of interest (see **Supplementary Table S1** for primers sequences) was performed on genomic DNA template followed by a bidirectional Sanger sequencing (see **Supplementary Methods**).

Multiple Protein Sequences Alignment

The *SLC10A7* human last transmembrane domain (TM10) was aligned with the *SLC10A7* sequence of other species using Uniprot website². The data were then imported and visualized with Jalview³ and colored according to the “ClustalX” coloring scheme.

In situ Hybridization

Mouse embryos embedding and sectioning, probe synthesis and *in situ* hybridization were performed as previously described in Laugel-Haushalter et al., 2012. DIG-labeled antisense riboprobe was transcribed *in vitro* with SP6 RNA polymerase (see **Supplementary Figure S1** for template sequence). The experiments were realized in accordance with the European Community Council Directive (86/609/EEC).

Western Blot Analysis

Patient and control primary fibroblasts were grown in DMEM (2 mM glutamine, 10% FCS), collected, and lysed in Ripa buffer containing protease cocktail inhibitor (Roche 06538282001).

¹www.phenodent.org

²<http://www.uniprot.org>

³<http://www.jalview.org/>

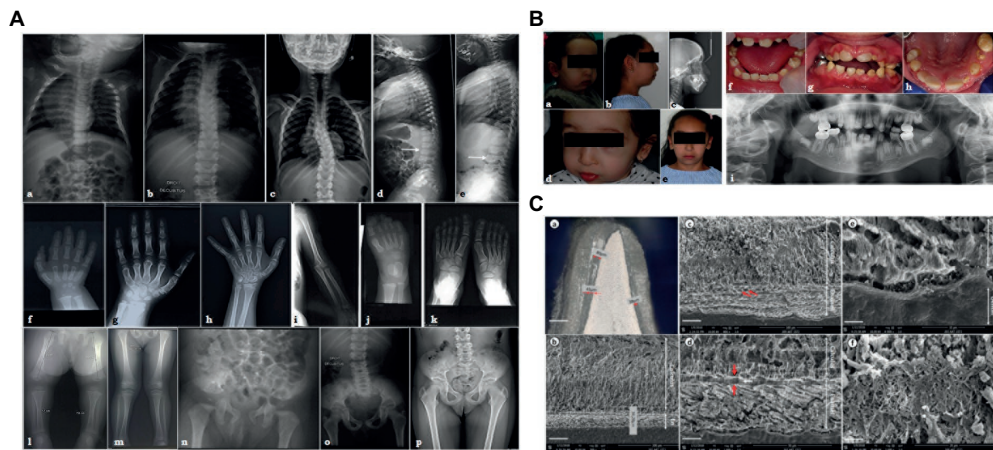


FIGURE 1 | (A) X-ray images of the skeletal anomalies at 3 months old (a,d,f,i,j,l,n), 4 years old (b,e,g,i,k,m,o) and 9 years old (c,h,p). Radiographs of the vertebral column showed the progressive scoliosis (a–c) and the ballooning of the vertebrae (d,e). Radiographs of hands and feet (f–h,j,k) displayed a progressive carpal and tarsal ossification. The patient also presented with short long bones (i,l,m), a genu valgus (l), and a horizontal acetabulum (n–p). **(B)** Clinical and radiological images of the facial and dental anomalies at 2 years old (a,d,f) and 9 years old (b,c,e,g–i). AI is visible in the clinical intraoral photographs affecting both the primary (f) and permanent dentition (g,h). The panoramic radiograph showed no radio-opacity contrast between dentine and enamel (i). **(C)** Optical numeric microscope and scanning electronic microscopy (SEM) of primary teeth. Numeric optical view of a sagittal section of a primary tooth. Dentine appears brighter than enamel. Lower central primary incisor (a). SEM micrograph of the thin enamel layer covering the vestibular side of the tooth. A thick layer of calcified calculus is present (b). Higher magnification of the thin enamel layer. Pseudo-prismatic structures are evident (arrows) (c). Micrograph of the thin enamel showing an outer aprismatic layer (between red arrows) exhibiting incremental lines (white arrow) (d). Enamel-dental junction showing typical scallops and unusual non-mineralized collagen fibers at the interface (e). Higher magnification of the calculus material capping the tooth. Many entangled calcified filamentous bacteria are observed (f).

Proteins obtained were used for Western blot, where SLC10A7 was detected with a specific monoclonal antibody (Novus NBP1-59875), followed by secondary HRP-coupled antibody hybridization (NA934V, GE Healthcare). Detection by ECL using the ChemiDoc™ Touch (BioRad) imaging system was performed. Quantification was performed using the ImageLab software (BioRad). SLC10A7 was quantified relative to the total amount of protein per lane revealed by TCE (T54801 Sigma) UV-labeling of the tryptophan residues, as shown on the stain-free loading control.

Calcium Localization

The control and patient fibroblasts were grown to 75% confluence in six-well plates with sterile coverslips. Cells were then incubated with 4 μ M Fluo4-AM (Thermo Fisher Scientific, #F14201) in DMEM medium (2 mM glutamate, 10% FBS) for 15 or 30 min at RT, quickly rinsed with fresh medium without Fluo4, and rinsed a second time for 15 or 30 min at RT. Cells were mounted directly in PBS and observed on a Zeiss Axio Observer D1 fluorescent microscope using a 40X objective.

RESULTS

Patient Phenotype

The investigated patient, a girl, is the consanguineous fourth child of an Algerian couple with no reported personal or family medical history. She presented with intra-uterine growth retardation and short femurs detected during the third trimester of pregnancy. The child was born at term with a birth height of 42 cm (<<third percentile), a birth weight of 2,890 g (10th percentile), and a

head circumference of 32.5 cm (10th percentile). Rhizomelia and brachydactyly of hands and feet were noticed at birth.

Her growth was regular at -3 SD for height, -2.5 SD for weight, and -1 SD for head circumference. Clinical examination revealed joint laxity without dislocations, articular limitations or pain, and a progressive scoliosis. By age 7, the child measured 106 cm (-3 SD), weighted 18 kg (-2.5 SD), and had a head circumference of 51 cm (0SD).

Radiographs were undertaken at 3 months (Figures 1Aa,Ad, Af,Aj,Al,An), at 4 years (Figures 1Ab,Ae,Ag,Ai,Ak,Am,Ao), and at 9 years (Figures 1Ac,Ah,Ap). Advanced carpal ossification, brachymetacarpia (Figures 1Af–Ah.), and short long bones (Figures 1Ai,Al,Am) were noticed at birth. Mesomelia grew more evident with age (Figures 1Af–Ah). Tarsal bone ossification was also advanced (Figures 1Aj,Ak). Vertebral bodies were initially considered as normal, although the latest radiographs showed abnormal expansions or ballooning (Figures 1Ad,Ae). Spinal hyperlordosis and scoliosis developed over time (Figures 1Aa–Ac). Horizontal acetabulum, large femoral necks, large iliac wings, and large clavicles were also noticed (Figures 1An–Ap). Neither metaphyseal nor epiphyseal anomalies were observed (Figures 1Al,Am). Different diagnoses of constitutional bones diseases such as achondroplasia and hypochondroplasia suggested Silver-Russell syndrome and diastrophic dwarfism, but molecular analyses ruled out these syndromes. Array-CGH was also normal (data not shown). The child had normal psychomotor development, although some learning difficulties were recently noticed.

The patient also presented with facial dysmorphisms, including microretrognathia, short neck, short nose, flat face, and blue sclerae (Figures 1Ba,Bb,Bd,Be). She had a narrow pharyngeal

tract, a hypodivergent profile with protruding incisors, and biproalveoly to compensate (**Figure 1Bc**) a lingual dysfunction and lip inocclusion.

The child had smaller teeth, incisor infraclusion, and severe hypoplastic/hypomineralized amelogenesis imperfecta on both the primary and permanent dentitions with colored and softer enamel (**Figure 1B**). On radiographs, there was no radio-opacity contrast between hypomineralized enamel and dentine (**Figure 1Bi**).

Initial ENT and ophthalmologic examinations were normal. Mild hypermetropia and astigmatism were then noticed in the first years of life.

Enamel Shows Quantitative and Qualitative Defects

Optical microscopy assessments of sagittal sections of a primary naturally exfoliated tooth revealed severe enamel hypoplasia (**Figure 1Ca**). The maximum thickness of enamel was observed on the vestibular side of the tooth and reached 80 μm (around 700 μm in a similar control tooth), and the entire tooth was capped by dental calculus. **Figure 1Cb** shows the large thickness of the calculus capping compared to the narrow enamel layer. In SEM, the thin enamel exhibited wide pseudo-rod patterns (**Figure 1Cc**). Moreover, electron microscopy disclosed a very thin outer layer of aprismatic enamel with incremental lines parallel to the surface (**Figure 1Cd**). At the scalloped enamel-dentinal junction, a non-mineralized interphase of collagen fibrils was evident (**Figure 1Ce**). At this higher magnification, the individual enamel crystals could be observed. Microscopy observations of the calculus capping clearly showed calcified bacterial structures (**Figure 1Cf**). When analyzed by energy dispersive X-ray spectroscopy (EDX), the calculus material had a Ca/P ratio of 1.56 ± 0.04 ($n = 12$) (**Supplementary Figure S2A**). Comparatively, EDX measurements were, respectively, 1.74 ± 0.068 ($n = 15$) and 1.64 ± 0.062 ($n = 15$) for enamel and dentine tissues (**Supplementary Figure S2B**). Although both spectra looked similar, the carbon content was higher in dental calculus, i.e., $C = 15.4 \pm 0.85$ (wt.%) and $C = 5.13 \pm 0.93$ (wt.%) for enamel (**Supplementary Figure S2**).

Mutation in *SLC10A7* Associated With Syndromic Amelogenesis Imperfecta

Whole-exome sequencing was performed on the index case and her parents. By using stringent criteria, identifying variants consistent with autosomal recessive disease inheritance (**Supplementary Table S2**), by manual curation (**Supplementary Table S3**), and by Sanger sequencing (**Supplementary Figure S3**, **Figure 2A**), we validated the homozygous mutation in exon 11 of *SLC10A7*, a gene involved in a novel type of skeletal dysplasia associated with short stature, AI, and scoliosis (SSASKS, #OMIM618363) (Ashikov et al., 2018; Dubail et al., 2018). The mutation leading to an amino acid change from proline to leucine in exon 11 at position 303 (NM_001300842.2:c.908C>T, p.Pro303Leu) was absent from databases (1,000 genomes, GnomAD).

The mutation affected the last transmembrane domain (TM10) of the *SLC10A7* protein. The alignment of this domain sequence with sequences of other species showed that the sequence was largely conserved between species, with the mutated proline (Pro303) being highly conserved (**Figure 2B**). The mutation was also predicted to be disease causing by SIFT (Vaser et al., 2016) and deleterious by Polyphen2 (Polymorphism Phenotyping v2) (PPH2) (Adzhubei et al., 2010). Collectively, these findings suggest an important function for this amino acid.

Expression Pattern of *SLC10A7* in Developing Mouse Bone and Tooth

The expression of *Slc10a7* at E14.5 mouse tooth cap stage had been reported in our previous transcriptomic study showing that the gene was expressed at similar levels in both molars and incisors (**Supplementary Table S4**; Laugel-Haushalter et al., 2012).

To gain insight into *Slc10a7* gene expression during bone and tooth development in mice, we performed an *in situ* hybridization analysis on mouse fetuses. *Slc10a7* positive signals were observed in the epithelial compartment of E14.5 cap stage teeth (**Figures 3Aa,Ab**). At E16.5, the transcripts were mostly localized in the inner dental epithelium and in the epithelial loop of the bell stage teeth. A discrete expression was also detected in the outer dental epithelium. At E18.5 labeling was observed in the inner dental epithelium of incisors and in ameloblasts and odontoblasts of molars. *Slc10a7* expression was investigated in developing bones, and mRNA transcripts were detected in bones undergoing ossification and vertebrae. We detected an expression not only in vertebrae at E16.5 (**Figures 3Ac,Ad**) and E18.5 (**Figures 3Ak,AI**) but also in the E16.5 humerus (**Figure 3Ag**) and femur (**Figure 3Ah**).

SLC10A7 Is Overexpressed in Patient Fibroblasts

The *SLC10A7* protein was identified by western blot analysis in protein extracts from patient and unaffected control primary skin fibroblasts (**Figure 3B**). Quantifications were done in three independent experiments using total protein extract (as revealed by the stain-free labeling) as loading control. The level of *SLC10A7* protein was approximately two times higher in the patient's cells compared to the unaffected control individual cells (**Figure 3B**). The results shown here were further validated using two different *SLC10A7* commercial antibodies (data not shown).

Abnormal Distribution of Cellular Calcium in Patient Fibroblasts

As the *SLC10A7* yeast *Saccharomyces cerevisiae* homologue (termed RCH1) is involved in the regulation of a calcium transporter (Zhao et al., 2016), this prompted us to investigate the distribution of calcium in patient's and control fibroblasts. Calcium labeling with a Fluo4 probe showed no difference in localization after 15 min of staining (**Figure 3C**). However, after 30 min, while the probe was mainly localized within the cytoplasm in the control cells, it was mostly retained within the nucleus in the patient's fibroblasts. This suggests that

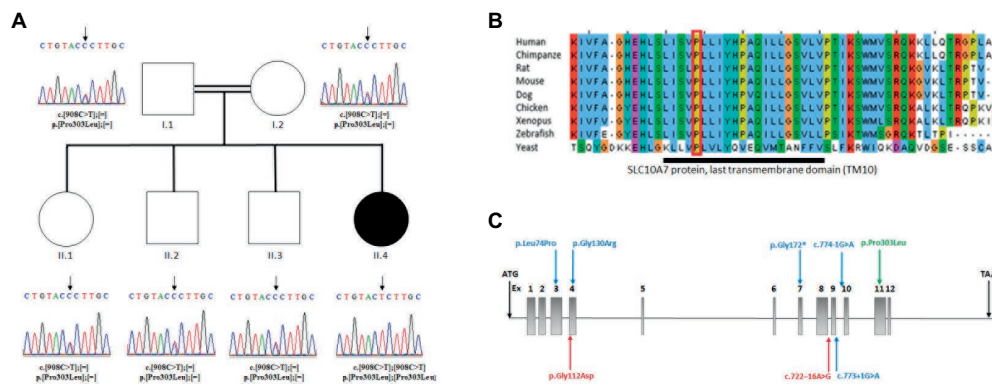


FIGURE 2 | (A) Analysis of the *SLC10A7* mutation: Pedigree of the AI patient and DNA sequencing chromatograms of the whole family. Parents and other children were heterozygous and the patient was homozygous for the mutation. An arrow points to the mutation. **(B)** Multiple sequence alignment of *SLC10A7* last transmembrane domain (TM10). The largely conserved sequence of the proteins last transmembrane domain is represented by the dark bar. The amino-acid affected by the missense mutation in the patient (red square) is conserved from human to yeast. **(C)** Human *SLC10A7* mutations in the literature. The *SLC10A7* gene contains 12 exons. The mutations described in Dubail et al. (2018) are represented by blue arrows and those reported in Ashikov et al., 2018 by red arrows. Our mutation is the only mutation located at the end of the gene (exon 11) and is represented by a green arrow.

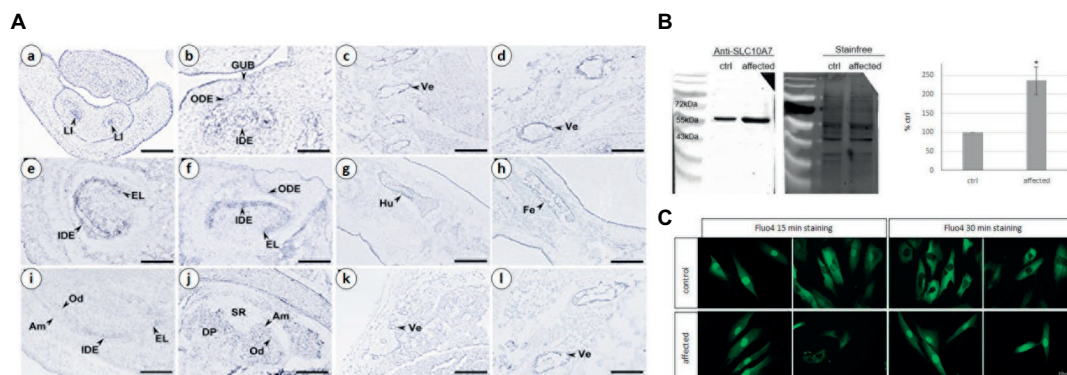


FIGURE 3 | (A) Analysis of mouse *Slc10a7* transcripts distribution by *in situ* hybridization. Sections illustrate *Slc10a7* expression in the developing bone, vertebrae, molars, and incisors. Developmental stages and section planes are: E14.5 frontal (a,b), E16.5 sagittal (c-h), E18.5 sagittal (i-l) sections. Am, ameloblasts; DP, dental papilla; EL, epithelial loop; GUB, gubernaculum; IDE, inner dental epithelium; LI, lower incisor; Od, odontoblasts; ODE, outer dental epithelium; SR, stellate reticulum; Ve, vertebrae. Scale bars: 10 μ m (b); 40 μ m (a,c,g,h); 30 μ m (d,i); 80 μ m (e,f); 150 μ m (k,l). **(B)** *SLC10A7* expression in control and patient skin fibroblasts as detected by western blot. The stain-free membrane, displaying the total amount of protein loaded, was used for quantification. The western blot shown corresponds to one significant experiment ($n = 3$). The data plotted were the mean of three independent experiments. Statistical analysis was done with the *t*-test and the *p* was determined $*p < 0.05$. **(C)** Calcium localization in control and patient fibroblasts. The cells were stained for 15 min or 30 min with 4 μ M of Fluo4, then washed for 15 or 30 min, respectively and observed by fluorescent microscopy. Images were taken with a 400 \times magnification, scale bar 10 μ m.

SLC10A7 could be involved in calcium transport and that the mutation of the patient resulting in an increased level of protein interferes with this function (Figure 3C).

DISCUSSION

Human Mutations

The study by (Dubail et al., 2018) described six patients with homozygous mutations in the *SLC10A7* gene presenting with SSASKS (#OMIM618363). Two patients had splice mutations in intron 9 c.774-G>A, c.773+1G>A; two had missense mutations c.221T>C and c.388G>A; and two showed the same stop mutation c.514C>T. Ashikov et al., 2018 reported two siblings patients

with a *SLC10A7* compound heterozygous mutation c.335G>A and c.722-16A>G and two patients with the same clinical presentation with no *SLC10A7* cDNA-identified mutations. In this paper, we report a homozygous missense mutation in exon 11, the only missense mutation reported to date at the end of this protein. The location near the end of the protein could potentially explain the milder phenotype in our patient (Figure 2C).

The patients described by (Dubail et al., 2018) displayed a more severe phenotype with multiple joint dislocations and had clinical features resembling Desbuquois syndrome (OMIM#251450, #615777 caused by mutations in *CANT1* and *XYLT1* genes), a chondrodysplasia with defects in GAG biosynthesis. The only distinguishable feature between those patients and Desbuquois-like patients was the AI phenotype (Table 1).

In the study by Ashikov et al., 2018, patients were believed similar to Desbuquois syndrome, sharing some intellectual disability traits not observed in our patient. However, Ashikov et al., 2018 patients also presented AI, a feature not reported so far in Desbuquois syndrome (Table 1). Comparing our patient to Desbuquois dysplasia patients, growth retardation was less severe (−3SD compare to −4 to −10SD). The patient also did not present with multiple dislocations or characteristic features like accessory ossification centers distal to the second metacarpal, bifid distal phalanx, or delta phalanx of the thumb. However, our patient had a similar facial appearance—round face, microretrognathism, blue sclerae, prominent eyes, and a short neck. Like other reported patients, our proband presented with AI. Since classical Desbuquois syndrome patients do not present this phenotype, AI may be a key symptom to distinguish *SLC10A7*-related disorders from other chondrodysplasias.

Mutant Mice and *SLC10A7* Expression Pattern

Slc10a7^{−/−} null mice exhibited a dysmorphic face, moderate skeletal dysplasia, ligamentous laxity, a reduced bone mass (Dubail et al., 2018) and they were smaller with shorter limbs (Brommage et al., 2014). In the mouse incisors, the most external layer, the aprismatic enamel layer, was missing, and numerous areas of hypoplasia were observed in the external prismatic enamel layer (Brommage et al., 2014; Dubail et al., 2018). As they mimic many clinical features of the patient phenotype, they provide a good model to study the physio-pathological mechanisms of the syndrome. It is interesting to note that no joint dislocations were observed, suggesting that *Slc10a7*^{−/−} mice

are mimicking the milder phenotype observed in our patient and in previous studies (Brommage et al., 2014). Here, we also showed that the expression pattern of *Slc10a7* in mice was consistent with the organs affected in all the described patients. Indeed, the gene was expressed in vertebrae, bones undergoing ossification and in the epithelial part of the tooth. Those findings are supporting the fact that even if the phenotype of SSAKS patients can vary in severity and include extra clinical features, short stature, scoliosis, and AI are recurrent features to identify patients with mutation in *SLC10A7*.

Role of *SLC10A7* in Calcium Transport

Here, we show that the level of *SLC10A7* protein is two times higher in affected patient skin fibroblasts compared to controls. Moreover, incubation of patient cells with a fluorescent calcium probe (Fluo4) led to its accumulation into the nucleus (cytoplasmic localization observed in control cells). This suggests that *SLC10A7* may play a role in calcium homeostasis, as reported for the *SLC10A7* yeast homologue Rch1 (Jiang et al., 2012). Indeed, in yeast, Rch1 is expressed in response to increased calcium levels and acts as a negative regulator of calcium uptake by binding to a calcium transporter at the plasma membrane (Jiang et al., 2012). Based on these yeast results, *SLC10A7* may also be involved in calcium homeostasis, and the patient missense mutations could alter this function, potentially leading to calcium accumulation in the nucleus of patient fibroblasts and higher expression of *SLC10A7* to compensate for this defect.

Previous studies concluded that defects in calcium uptake led to defects in GAG synthesis explaining *SLC10A7* phenotypes. This mechanism could also produce skeletal defects.

TABLE 1 | Review of the literature of *SLC10A7* patients and clinical comparison with Desbuquois syndrome.

Clinical features	Reported patients (Dubail et al., 2018)	Reported patients (Ashikov et al., 2018)	Our patient	Desbuquois syndrome
Intra-uterine growth retardation	6/6		+	+
Postnatal growth retardation	6/6	5/5	+ (−3SD)	+ (−4SD to −10SD)
Micrognathia	6/6	2/5	+	+
Congenital multiple dislocations	6/6	—	—	+
Amelogenesis imperfecta	6/6	5/5	+	—
Advanced carpal ossification	6/6		+	+
Scoliosis	6/6	5/5	+	+
Blue sclerae			+	+
Prominent eyes		+	+	+
Flat face			+	+
Short neck		+	+	+
Presence of hand anomalies, namely, accessory ossification center distal to the second metacarpal, bifid distal phalanx, or delta phalanx of the thumb	—		—	+ (type 1)
Brachydactyly		+	+	+
Short long bones with “Swedish key” appearance of the proximal femur	4/6			+
Organ malformations			—	+/—
Mental retardation		4/5	—	+/—

The patients from Dubail et al. (2018) presented with a phenotype close to the Desbuquois syndrome phenotype. Patient from (Ashikov et al., 2018) and our proband did not present joint dislocation. All the patients were affected by a skeletal dysplasia associated with AI.

Dental defects, though, could be due to a small misregulation in calcium homeostasis, sufficient to induce enamel defects. Moreover, as discussed in Dubail et al., 2018, some mutations in calcium channels and transporters/exchangers (*STIM1*, *ORAI1*, *SLC24A4*, *SLC24A5*, *WDR72*) (El-Sayed et al., 2009; Feske, 2010; Duan, 2014; Wang et al., 2014) lead to AI without any associated GAG biosynthesis defects.

Amelogenesis Imperfecta as a Key Clinical Feature in Delineation of Skeletal Dysplasia

Skeletal dysplasias are a large, diverse group of diseases in which recent progress in genetics has allowed identification of many causative genes. An accurate clinical examination can also provide key diagnostic clues. Here, we report the case of a patient with a skeletal dysplasia associated with AI. Skeletal dysplasia with associated *SLC10A7* mutations varies in severity and clinical features (Ashikov et al., 2018; Dubail et al., 2018) but AI is always present. Indeed, mutation in *SLC10A7* was recently associated to a novel type of skeletal dysplasia with short stature, AI, and scoliosis (#OMIM611459). To date, combined skeletal dysplasia/AI-associated syndromes have been observed in non-lethal Raine syndrome with mutations in *FAM20C* (OMIM#259775) (Elalaoui et al., 2016), brachyolmia with AI caused by mutations in *LTBP3* (Huckert et al., 2015), tricho-dento-osseous syndrome caused by mutations in *DLX3* (OMIM#190320) (Niemenen et al., 2011), and congenital disorder of glycosylation, type IIk caused by mutations in *TMEM165* (OMIM #614727).

As the association of the two clinical features seems to be rare, this suggests that AI is the key factor pointing to a mutation in *SLC10A7* in patients presenting with skeletal dysplasia and scoliosis. Moreover, comparing our patient to those in the literature (Ashikov et al., 2018; Dubail et al., 2018), we noticed that the skeletal phenotype was milder, suggesting that some patients could be difficult to diagnose. Thus, the finding of AI can help improving patient diagnosis.

ETHICS STATEMENT

The oral phenotype was documented using the D[4]/phenodent registry protocol, a Diagnosing Dental Defects Database (see www.phenodent.org, for assessment form), which is approved by CNIL (French National commission for informatics and liberty, number 908416). This clinical study is registered at <https://clinicaltrials.gov>: NCT01746121 and NCT02397824 and with the MESR (French Ministry of Higher Education and Research) Bioethics Commission as a biological collection “Oro-dental Manifestations of Rare Diseases” DC-2012-1677 within DC-2012-1002 and was acknowledged by the CPP (person protection committee) Est IV December 11th 2012. The patient and the non-affected family members gave written informed consents in accordance with the Declaration of Helsinki, both for the D[4]/phenodent registry and for genetic analyses performed on the salivary samples included in the biological collection.

AUTHOR CONTRIBUTIONS

ES, MH, PH, YA, and M-CM collected the salivary samples and detailed the patients' phenotype. VL-H, CS, and MH identified the molecular basis of the disease through NGS assays. VL-H, SB, ES, CS, VG, NK, JH, M-CM, SF, HD, and AB-Z analyzed the data and wrote the manuscript. AB-Z designed the study and was involved from conception, fund seeking to drafting, and critical review of the manuscript. All authors therefore contributed to the conception, design, data acquisition, analysis, and interpretation, drafted and critically revised the manuscript. All authors gave final approval and agreed to be accountable for all aspects of the work.

FUNDING

This work was financed by and contributed to the actions of the project No. 1.7 “RARENET: a trinational network for education, research and management of complex and rare disorders in the Upper Rhine” co-financed by the European Regional Development Fund (ERDF) of the European Union in the framework of the INTERREG V Upper Rhine program as well as to the ERN (European reference network) CRANIO initiative. AB-Z is a 2015 USIAS Fellow, Institute of Advanced Studies (Institut d'Etudes Avancées) de l'Université de Strasbourg, France. This work was also supported by grants from the French Ministry of Health (National Program for Clinical Research, PHRC 2008 N°4266 Amelogenesis imperfecta), the University Hospital of Strasbourg (HUS, API, 2009–2012, “Development of the oral cavity: from gene to clinical phenotype in Human,” and the grant ANR-10-LABX-0030-INRT, a French State fund managed by the Agence Nationale de la Recherche under the frame program Investissements d'Avenir labeled ANR-10-IDEX-0002-02. This work was also funded by INSERM).

ACKNOWLEDGMENTS

The authors are grateful to the family for participation and invaluable contributions. The computing resources for this work were provided by the BICS and BISTRO bioinformatics platforms in Strasbourg. They are grateful to Mrs. Marzena Kawczynski and Mr. Sébastien Troester for continuous support and help with patient data management. They are grateful to Dr. Karen Niederreither as well as Pr. Ophir Klein for critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/article/10.3389/fgene.2019.00504/full#supplementary-material>

REFERENCES

- Abou Neel, E. A., Aljabo, A., Strange, A., Ibrahim, S., Coathup, M., Young, A. M., et al. (2016). Demineralization–remineralization dynamics in teeth and bone. *Int. J. Nanomedicine* 11, 4743–4763. doi: 10.2147/IJN.S107624
- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Alanay, Y., and Lachman, R. S. (2011). A review of the principles of radiological assessment of skeletal dysplasias. *J. Clin. Res. Pediatr. Endocrinol.* 3, 163–178. doi: 10.4274/jcrpe.463
- Ashikov, A., Abu Bakar, N., Wen, X.-Y., Niemeijer, M., Rodrigues Pinto Osorio, G., Brand-Arzamendi, K., et al. (2018). Integrating glycomics and genomics uncovers *SLC10A7* as essential factor for bone mineralization by regulating post-Golgi protein transport and glycosylation. *Hum. Mol. Genet.* 27, 3029–3045. doi: 10.1093/hmg/ddy213
- Brommage, R., Liu, J., Hansen, G. M., Kirkpatrick, L. L., Potter, D. G., Sands, A. T., et al. (2014). High-throughput screening of mouse gene knockouts identifies established and novel skeletal phenotypes. *Bone Res.* 2:14034. doi: 10.1038/boneres.2014.34
- Duan, X. (2014). Ion channels, channelopathies, and tooth formation. *J. Dent. Res.* 93, 117–125. doi: 10.1177/0022034513507066
- Dubail, J., Huber, C., Chantepie, S., Sonntag, S., Tüysüz, B., Mihci, E., et al. (2018). *SLC10A7* mutations cause a skeletal dysplasia with amelogenesis imperfecta mediated by GAG biosynthesis defects. *Nat. Commun.* 9:3087. doi: 10.1038/s41467-018-05191-8
- Elalaoui, S. C., Al-Sheqaih, N., Ratbi, I., Urquhart, J. E., O'Sullivan, J., Bhaskar, S., et al. (2016). Non lethal Raine syndrome and differential diagnosis. *Eur. J. Med. Genet.* 59, 577–583. doi: 10.1016/j.ejmg.2016.09.018
- El-Sayed, W., Parry, D. A., Shore, R. C., Ahmed, M., Jafri, H., Rashid, Y., et al. (2009). Mutations in the beta propeller *WDR72* cause autosomal-recessive hypomaturational amelogenesis imperfecta. *Am. J. Hum. Genet.* 85, 699–705. doi: 10.1016/j.ajhg.2009.09.014
- Feske, S. (2010). CRAC channelopathies. *Pflugers Arch.* 460, 417–435. doi: 10.1007/s00424-009-0777-5
- Geoffroy, V., Pizot, C., Redin, C., Piton, A., Vasli, N., Stoetzel, C., et al. (2015). VaRank: a simple and powerful tool for ranking genetic variants. *PeerJ* 3:e796. doi: 10.7717/peerj.796
- Huckert, M., Stoetzel, C., Morkmued, S., Laugel-Haushalter, V., Geoffroy, V., Muller, J., et al. (2015). Mutations in the latent TGF-beta binding protein 3 (LTBP3) gene cause brachyolmia with amelogenesis imperfecta. *Hum. Mol. Genet.* 24, 3038–3049. doi: 10.1093/hmg/ddv053
- Jiang, L., Alber, J., Wang, J., Du, W., Yang, X., Li, X., et al. (2012). The *Candida albicans* plasma membrane protein Rch1p, a member of the vertebrate *SLC10* carrier family, is a novel regulator of cytosolic Ca²⁺ homeostasis. *Biochem. J.* 444, 497–502. doi: 10.1042/BJ20112166
- Laugel-Haushalter, V., Langer, A., Marrie, J., Fraulob, V., Schuhbaur, B., Koch-Phillips, M., et al. (2012). From the transcription of genes involved in ectodermal dysplasias to the understanding of associated dental anomalies. *Molecular Syndromology* 3, 158–168. doi: 10.1159/000342833
- Nieminen, P., Lukinmaa, P.-L., Alapulli, H., Methuen, M., Suojärvi, T., Kivirikko, S., et al. (2011). *DLX3* homeodomain mutations cause tricho-dento-osseous syndrome with novel phenotypes. *Cells Tissues Organs* 194, 49–59. doi: 10.1159/000322561
- Smith, C. E. L., Poulter, J. A., Antanaviciute, A., Kirkham, J., Brookes, S. J., Inglehearn, C. F., et al. (2017). Amelogenesis imperfecta; genes, proteins, and pathways. *Front. Physiol.* 8:435. doi: 10.3389/fphys.2017.00435
- Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., and Ng, P. C. (2016). SIFT missense predictions for genomes. *Nat. Protoc.* 11, 1–9. doi: 10.1038/nprot.2015.123
- Wang, S., Choi, M., Richardson, A. S., Reid, B. M., Seymen, F., Yildirim, M., et al. (2014). *STIM1* and *SLC24A4* are critical for enamel maturation. *J. Dent. Res.* 93, 94S–100S. doi: 10.1177/0022034514527971
- Zhao, Y., Yan, H., Happeck, R., Peiter-Volk, T., Xu, H., Zhang, Y., et al. (2016). The plasma membrane protein Rch1 is a negative regulator of cytosolic calcium homeostasis and positively regulated by the calcium/calmodulin signaling pathway in budding yeast. *Eur. J. Cell Biol.* 95, 164–174. doi: 10.1016/j.ejcb.2016.01.001

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Laugel-Haushalter, Bär, Schaefer, Stoetzel, Geoffroy, Alembik, Kharouf, Huckert, Hamm, Hemmerlé, Manière, Friant, Dollfus and Bloch-Zupan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Molecular Genetics Analysis of 70 Chinese Families With Muscular Dystrophy Using Multiplex Ligation-Dependent Probe Amplification and Next-Generation Sequencing

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University,
China

Reviewed by:

Qing Lyu,
University of Rochester,
United States
Yaqiong Jin,
Capital Medical University,
China

*Correspondence:

Zhongtao Gai
gaizhongtao@sina.com
Yi Liu
y_liu99@sina.com

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 16 October 2018

Accepted: 24 June 2019

Published: 25 July 2019

Citation:

Wang D, Gao M, Zhang K, Jin R,
Lv Y, Liu Y, Ma J, Wan Y, Gai Z and
Liu Y (2019) Molecular Genetics
Analysis of 70 Chinese Families
With Muscular Dystrophy Using
Multiplex Ligation-Dependent
Probe Amplification and Next-
Generation Sequencing.
Front. Pharmacol. 10:814.
doi: 10.3389/fphar.2019.00814

**Dong Wang^{1†}, Min Gao^{1†}, Kaihui Zhang¹, Ruifeng Jin², Yuqiang Lv¹, Yong Liu², Jian Ma¹,
Ya Wan¹, Zhongtao Gai^{1*} and Yi Liu^{1*}**

¹ Pediatric Research Institute, Qilu Children's Hospital, Shandong University, Ji'nan, China, ² Neurology Department, Qilu Children's Hospital, Shandong University, Ji'nan, China

Background: Muscular dystrophy (MD) includes multiple types, of which dystrophinopathies caused by *dystrophin* (*DMD*) mutations are the most common types in children. An accurate identification of the causative mutation at the genomic level is critical for genetic counseling of the family, and analysis of genotype–phenotype correlations, as well as a reference for the development of gene therapy.

Methods: Totally, 70 Chinese families with suspected MD probands were enrolled in the study. The multiplex ligation-dependent probe amplification (MLPA) was first performed to screen large deletions/duplications of *DMD* exons in the patients, and then, next-generation sequencing (NGS) was carried out to detect small mutations in the MLPA-negative patients.

Results: Totally, 62 mutations of *DMD* were found in 62 probands with DMD/BMD, and two compound heterozygous mutations in *LAMA2* were identified in two probands with MDC1A (a type of congenital MD), indicating that the diagnostic yield was 91.4% by MLPA plus NGS for MD diagnosis in this cohort. Out of the mutations, 51 large mutations encompassing 47 (75.8%) deletions and four duplications (6.5%) were identified by MLPA; 11 small mutations including six (9.7%) nonsense, two (3.2%) small deletions, two splice-site mutations (3.2%), and one small insertion (1.6%) were found by NGS. Large mutations were found most frequently in the hotspot region between exons 45 and 55 (70.6%). Out of the 11 patients harboring point mutations in *DMD*, 8 were novel mutations. Additionally, one novel mutation in *LAMA2* was identified. All the novel mutations were analyzed and predicted as pathogenic according to American College of Medical Genetics and Genomics (ACMG) guideline. Finally, 34 DMD, 4 BMD, 24 BMD/DMD, and 2 MDC1A were diagnosed in the cohort.

Conclusion: Our data indicated that the MLPA plus NGS can be a comprehensive and effective tool for precision diagnosis and potential treatment of MD and is particularly

necessary for the patients at very young age with only two clinical indicators (persistent hyperCKemia and typical myopathy performance on electromyogram) but no definite clinical manifestations.

Keywords: muscular dystrophy, multiplex ligation-dependent probe amplification, next-generation sequencing, *dystrophin* (*DMD*), merosin-deficient congenital muscular dystrophy type 1A, *LAMA2*

INTRODUCTION

Muscular dystrophies (MD), an inherited group of degenerative skeletal muscle disorders, are characterized by progressive/congenital weakness and breakdown of skeletal muscles encompassing great clinical and genetic heterogeneity, and even death because of cardiomyopathy and respiratory failure (Mercuri and Muntoni, 2013; Falsaperla et al., 2016; Carter et al., 2018). Currently, MD are clinically classified to six categories with various degree of severity, including Duchenne MD (DMD) and Becker MD (BMD), limb-girdle MD (LGMD), distal MD, congenital MD (CMD), facio-scapulo-humeral MD, and myotonic MD (Falsaperla et al., 2016). DMD/BMD that are caused by mutations in X chromosome-linked *dystrophin* (*DMD*) are the most common forms in childhood with an estimated incidence of 8.3 or 7.3 per 100,000 male (Wein et al., 2015; Carter et al., 2018). DMD, a severe phenotype clinically, is characterized by a progressive loss of muscle function with onset at age 2 to 5 years, lost ambulation before age 13 years and death at approximately 20 years of age, while BMD shows a mild form with patients being loss of ambulation after 16 years of age (Mercuri and Muntoni, 2013; Wein et al., 2015; Yiu and Kornberg, 2015). Other forms of MD are mostly autosomal recessive with rare prevalence in childhood but vary in region (Mercuri and Muntoni, 2013; Wein et al., 2015; Zimowski et al., 2017; Luce et al., 2018; Wang et al., 2019).

DMD is the largest gene in human, spanning 2.4 Mb of genomic region on Xp21, and containing 79 exons and 78 lengthy introns, which produces a 14.6 kb mRNA transcript (Ahn and Kunkel, 1993). To date, many mutations in *DMD* have been described, approximately 70% of which are large deletions/duplications (≥ 1 exon), while the remaining 25% to 30% are caused by small mutations (< 1 exon), encompassing point mutations or small deletions/insertions (Yiu and Kornberg, 2015). The differences of clinical manifestations between BMD and DMD result from different mutation types in *DMD* gene. Patients will suffer from DMD when the mutation leads to a frameshift (out of frame) or generates a premature stop codon; non-functional dystrophin protein is produced. On the contrary, patients will present the phenotype of BMD when the mutation maintains in reading frame (in-frame mutation); a partially functional dystrophin protein is produced (Mohammed et al., 2018).

Recently, several promising mutation-specific molecular therapies have been developed, including exon skipping to restore the reading frame and increase expression of the compensatory dystrophin, and read-through therapy of a nonsense codon to produce full-length dystrophin which is

applicable to harbor nonsense mutations for patients with DMD. Therefore, it is essential to make an early and accurate diagnosis for the patients with suspected MD to provide information on eligibility of mutation-specific treatments as well as optimal care and family planning. Multiplex ligation-dependent probe amplification (MLPA), a simple and rapid screening tool, has been developed and used to test large deletions and duplications of all 79 exons in *DMD* gene in different populations (Gatta et al., 2005; Hegde et al., 2008; Suh et al., 2017). As the small mutations are easily missed by MLPA (Stuppia et al., 2012), next-generation sequencing (NGS) has been applied for MLPA-negative patients and indicated high efficiency and cost-effectiveness but varied prevalence and mutation types in different populations of locations (Niba et al., 2014; Wang et al., 2014; Singh et al., 2018).

In this study, 70 patients clinically suspected MD and their families from Shandong province of China were investigated; MLPA was firstly used to detect large deletions and duplications in *DMD* gene in probands; then, NGS was applied to find small mutations in the MLPA-negative patients. The mutation patterns and hot spot locations were analyzed in order to establish genotype-phenotype correlations, and a reference for the development of gene therapy.

MATERIALS AND METHODS

Patients and Samples

A total of 70 unrelated hospitalized children (67 boys and three girls, mean age 3.47 ± 2.97 years; see **Table 1** for details) with a clinically suspected diagnosis of MD and their healthy parents were enrolled in this study in Qilu Children's Hospital of Shandong University (QCHSU) from July 2015 to December 2017, including 36 clinically suspected DMD, six suspected BMD, and 28 UMD (uncertain MD) without typical clinical phenotype due to relatively young age.

All participants were from Han Chinese population in Shandong Province, north of China. All probands were examined and diagnosed by experienced neuromuscular specialists at Neurology Department of QCHSU. Clinical diagnosis was based on clinical features including: 1) significantly increased serum creatine kinase (CK) level; 2) myopathic abnormalities, but normal peripheral nerve conduction velocity on EMG; 3) a positive family history with MD; 4) muscular weakness; 5) difficulty in walking; 6) Gowers sign; 7) calf muscle pseudohypertrophy; 8) difficulty climbing stairs; and 9) waddling gait and so on (Mercuri and Muntoni, 2013; Carter et al., 2018).

TABLE 1 | The clinical and laboratory features of the 70 clinically suspected muscular dystrophy (MD) probands.

Patient ID	Sex	Age (years)	Clinical manifestation	CK value (U/L)	EMG: myopathic abnormalities	Clinical diagnosis	Genotypes
							Gene: mutation
P2	Male	7.9	Difficulty walking, Gowers sign, proximal muscle weakness, calf muscle pseudohypertrophy	16,470	NA	DMD	<i>DMD</i> : DelEX2
P3	Male	7.3	Difficulty walking, Gowers sign, difficulty climbing stairs, proximal muscle weakness	49,328	NA	DMD	<i>DMD</i> : DelEX3-13
P4	Male	1.8	Gowers sign	26,963	NA	DMD	<i>DMD</i> : DelEX3-29
P5	Female	2.4	Progressive muscle weakness	5,521	+	DMD-like	<i>DMD</i> : DelEX8-12
P6	Male	8.0	Difficulty walking, muscular hypotonia, waddling gait, falls, difficulty climbing stairs, Gowers sign	13,684	NA	DMD	<i>DMD</i> : DelEX8-23
P7	Male	2.7	Muscular hypotonia, calf muscle pseudohypertrophy	46,191	+	DMD	<i>DMD</i> : DelEX10-44
P8	Male	2.9	Progressive muscle weakness	6,093	+	DMD	<i>DMD</i> : DelEX13
P13	Male	7.3	Difficulty walking, waddling gait, falls, difficulty climbing stairs, Gowers sign	22,001	+	DMD	<i>DMD</i> : DelEX45
P18	Male	7.6	Progressive muscle weakness, falls, waddling gait	16,427	NA	DMD	<i>DMD</i> : DelEX45-54
P19	Male	1.8	Progressive muscle weakness	16,530	+	DMD	<i>DMD</i> : DelEX45-54
P23	Male	6.2	Difficulty walking, difficulty climbing stairs, calf muscle pseudohypertrophy, progressive muscle weakness	12,270	+	DMD	<i>DMD</i> : DelEX46-55
P25	Male	7.0	Waddling gait, difficulty climbing stairs, Gowers sign	12,537	NA	DMD	<i>DMD</i> : DelEX48-50
P27	Male	9.0	Difficulty walking, difficulty climbing stairs, Gowers sign, falls, waddling gait	17,713	NA	DMD	<i>DMD</i> : DelEX49-50
P30	Female	8.0	Progressive muscle weakness, calf muscle hypertrophy	13,323	NA	DMD	<i>DMD</i> : DelEX50
P31	Male	1.1	Difficulty walking	10,491	NA	DMD	<i>DMD</i> : DelEX50-52
P32	Male	3.0	Progressive muscle weakness, Gowers sign	25,758	NA	DMD	<i>DMD</i> : DelEX50-52
P35	Male	9.0	Progressive muscle weakness, difficulty climbing stairs, Gowers sign	17,771	+	DMD	<i>DMD</i> : DelEX51
P36	Male	7.0	Calf muscle pseudohypertrophy, Gowers sign	14,659	NA	DMD	<i>DMD</i> : DelEX51
P37	Male	3.2	Calf muscle pseudohypertrophy	25,596	NA	DMD	<i>DMD</i> : DelEX51-55
P38	Male	7.5	Difficulty walking, difficulty climbing stairs, Gowers sign, progressive muscle weakness, Calf muscle pseudohypertrophy	20,051	NA	DMD	<i>DMD</i> : DelEX51-55
P41	Male	4.0	Progressive muscle weakness	9,765	+	DMD	<i>DMD</i> : DelEX53-55
P43	Male	1.8	Gowers sign	10,001	+	DMD	<i>DMD</i> : DelEX55-77
P44	Male	2.0	Delayed gross motor development	7,307	+	DMD	<i>DMD</i> : DelEX56-79
P46	Male	6.7	Progressive muscle weakness, waddling gait, Calf muscle pseudohypertrophy, Gowers sign	10,578	+	DMD	<i>DMD</i> : DelEX62
P48	Male	7.4	Muscle weakness, motor delay	16,320	NA	DMD	<i>DMD</i> : DupEX3-7
P49	Male	5.6	Motor delay, muscle weakness	19,726	+	DMD	<i>DMD</i> : DupEX5-7
P50	Male	2.7	Progressive muscle weakness, difficulty climbing stairs, Gowers sign, calf muscle pseudohypertrophy	31,680	NA	DMD	<i>DMD</i> : DupEX19-21
P51	Male	7.0	Difficulty walking, Gowers sign, calf muscle pseudohypertrophy	5,243	+	DMD	<i>DMD</i> : DupEX45-54
P52	Male	2.1	Difficulty walking, calf muscle pseudohypertrophy	23,479	+	DMD	<i>DMD</i> : c.2436C > T
P53	Male	6.4	Calf muscle hypertrophy, progressive muscle weakness	15,401	NA	DMD	<i>DMD</i> : c.7264dupG
P54	Male	4.0	Progressive muscle weakness, calf muscle pseudohypertrophy	2,6917	+	DMD	<i>DMD</i> : c.1231A > T
P57	Male	3.1	Calf muscle pseudohypertrophy	8,372	+	DMD	<i>DMD</i> : c.5167G > T
P58	Male	2.3	Delayed gross motor development	35,951	+	DMD	<i>DMD</i> : 10187delC
P61	Male	7.0	Difficulty walking, Gowers sign, calf muscle pseudohypertrophy	15,795	NA	DMD	<i>DMD</i> : c.2571delC
P68	Male	5.4	Gowers sign	20,611	NA	DMD	Unknown
P69	Male	1.7	Calf muscle pseudohypertrophy	4,938	NA	DMD	Unknown
P14	Male	3.7	Calf muscle pseudohypertrophy	3,087	–	BMD	<i>DMD</i> : DelEX45-47
P16	Male	9.9	Difficulty climbing stairs, calf muscle pseudohypertrophy	5,641	NA	BMD	<i>DMD</i> : DelEX45-47
P17	Male	5.3	Calf muscle pseudohypertrophy	12,744	+	BMD	<i>DMD</i> : DelEX45-48
P56	Male	3.8	Calf muscle pseudohypertrophy	5,149	NA	BMD	<i>DMD</i> : c.1812 + 1C > T
P65	Male	5.6	Calf muscle pseudohypertrophy	13,783	+	BMD	Unknown
P70	Male	11.0	No	2,095	NA	BMD	Unknown
P1	Male	0.2	No	14,982	+	UMD	<i>DMD</i> : DelEX1-79
P9	Male	1.7	No	4,315	NA	UMD	<i>DMD</i> : DelEX26-34
P10	Male	0.4	No	20,053	NA	UMD	<i>DMD</i> : DelEX44
P11	Male	0.8	No	5,325	NA	UMD	<i>DMD</i> : DelEX44-55

(Continued)

TABLE 1 | Continued

Patient ID	Sex	Age (years)	Clinical manifestation	CK value (U/L)	EMG: myopathic abnormalities	Clinical diagnosis	Genotypes
							Gene: mutation
P12	Male	0.6	No	35,340	NA	UMD	DMD: DelEX45
P15	Male	1.0	No	4,007	NA	UMD	DMD: DelEX45-47
P20	Male	0.6	No	16,936	+	UMD	DMD: DelEX45-54
P21	Male	0.1	No	53,053	+	UMD	DMD: DelEX46-48
P22	Male	0.9	Calf muscle pseudohypertrophy	7,692	+	UMD	DMD: DelEX46-52
P24	Male	0.1	No	10,935	+	UMD	DMD: DelEX46-55
P26	Male	1.7	No	16,753	+	UMD	DMD: DelEX49-50
P28	Male	0.7	No	5,927	+	UMD	DMD: DelEX49-52
P29	Male	0.1	No	6,360	NA	UMD	DMD: DelEX49-52
P33	Male	1.2	No	16,428	+	UMD	DMD: DelEX51
P34	Male	0.7	No	17,001	+	UMD	DMD: DelEX51
P39	Male	1.0	No	8,178	+	UMD	DMD: DelEX52
P40	Male	0.6	No	9,606	+	UMD	DMD: DelEX53
P42	Male	0.8	No	4,561	+	UMD	DMD: DelEX55
P45	Male	0.7	No	10,444	+	UMD	DMD: DelEX61
P47	Male	0.1	No	21,064	+	UMD	DMD: DelEX69-71
P55	Male	0.9	Calf muscle pseudohypertrophy	3,676	+	UMD	DMD: c.580C > T
P59	Male	1.2	No	10,456	NA	UMD	DMD: c.7660 + 1C > G
P60	Male	1.0	No	20,926	NA	UMD	DMD: c.7792C > T
P62	Male	1.0	Calf muscle pseudohypertrophy	21,234	NA	UMD	DMD: c.6283C > T
P63	Female	0.6	Muscle weakness, muscular hypotonia, motor delay, dysphagia, weak cry	2,594	+	UMD	LAMA2: c.2049_2050delAG LAMA2:DelEX 4
P64	Male	0.4	Muscle weakness, muscular hypotonia, motor delay, dysphagia, weak cry	2,391	+	UMD	LAMA2:c.1672C > T/ LAMA2:DelEX 4
P66	Male	1.7	No	2,913	NA	UMD	Unknown
P67	Male	3.0	No	4,093	NA	UMD	Unknown

All the variants in DMD gene and LAMA2 gene shown in **Table 1 and Supplementary Table S1** are described using the NM_004006.2 transcript and the NM_000426.3 transcript reference sequence, respectively.

Del, deletion; Dup, duplication; EX, Exon; DMD, Duchenne's Muscular Dystrophy; BMD, Becker Muscular Dystrophy; UMD, unclear phenotype because of young age. NA, not available; "+": Present; "-": Absent.

The inclusion criteria for the probands enrolling in this study were in accord with clinical features of (1) or (1) and (2) or (1) and (2) plus one or several other features described above.

We excluded patients had no clinical feature (1). The clinical diagnosis of DMD is based on: 1) progressive symmetric muscle weakness (proximal > distal) often with calf hypertrophy, 2) symptoms present before age 5 years, and 3) lost ambulation before age 13 years. The clinical diagnosis of BMD is based on: 1) progressive symmetric muscle weakness (proximal > distal) often with calf hypertrophy, 2) activity-induced cramping (present in some individuals), 3) flexion contractures of the elbows (if present, late in the course), 4) lost ambulation after age 16 years, and 5) preservation of neck flexor muscle strength (differentiates BMD from DMD) (Darras et al., 1993). Some patients, who could not be classified in this way, were recognized as DMD when an onset of weakness occurred by the age of 5, or considered as BMD when they had very mild or nearly normal motor dysfunction after 5 years old (Marden et al., 2005).

In addition, we used "UMD" to describe the uncertain MD types of the patients who cannot be diagnosed due to very young age, only manifesting persistent hyperCKemia and typical myopathy manifestation on EMG but normal peripheral nerve conduction velocity.

Blood samples obtained from the subjects were collected in EDTA vacutainer, and genomic DNA was extracted by using QIAamp DNA Blood Mini Kit (Qiagen, Shanghai, China) following the manufacturer's instructions. MLPA was used as the first-line molecular detection, and NGS was applied to detect small mutations in MLPA-negative samples (The work flow chart of the study was shown in **Figure 1**).

MLPA Assay

The MLPA reaction was carried out to screen all exons of *DMD* using SALSA MLPA probe sets P034 and P035 (MRC-Holland, Amsterdam, the Netherlands) according to the manufacturer's instructions. Amplified products were separated using a 3500 XL Genetic Analyzer (ABI, Carlsbad, CA, USA), and the data were analyzed by Coffalyser Software (MRC-Holland, Amsterdam, the Netherlands).

Next-Generation Sequencing

Totally, 1,500 ng genomic DNA was fragmented to an average size of 300 bp; then, the fragmented genomic DNA was used to prepare sequencing libraries and 8 bp barcoded sequencing adapters were ligated to the DNA fragments before final hybridization with

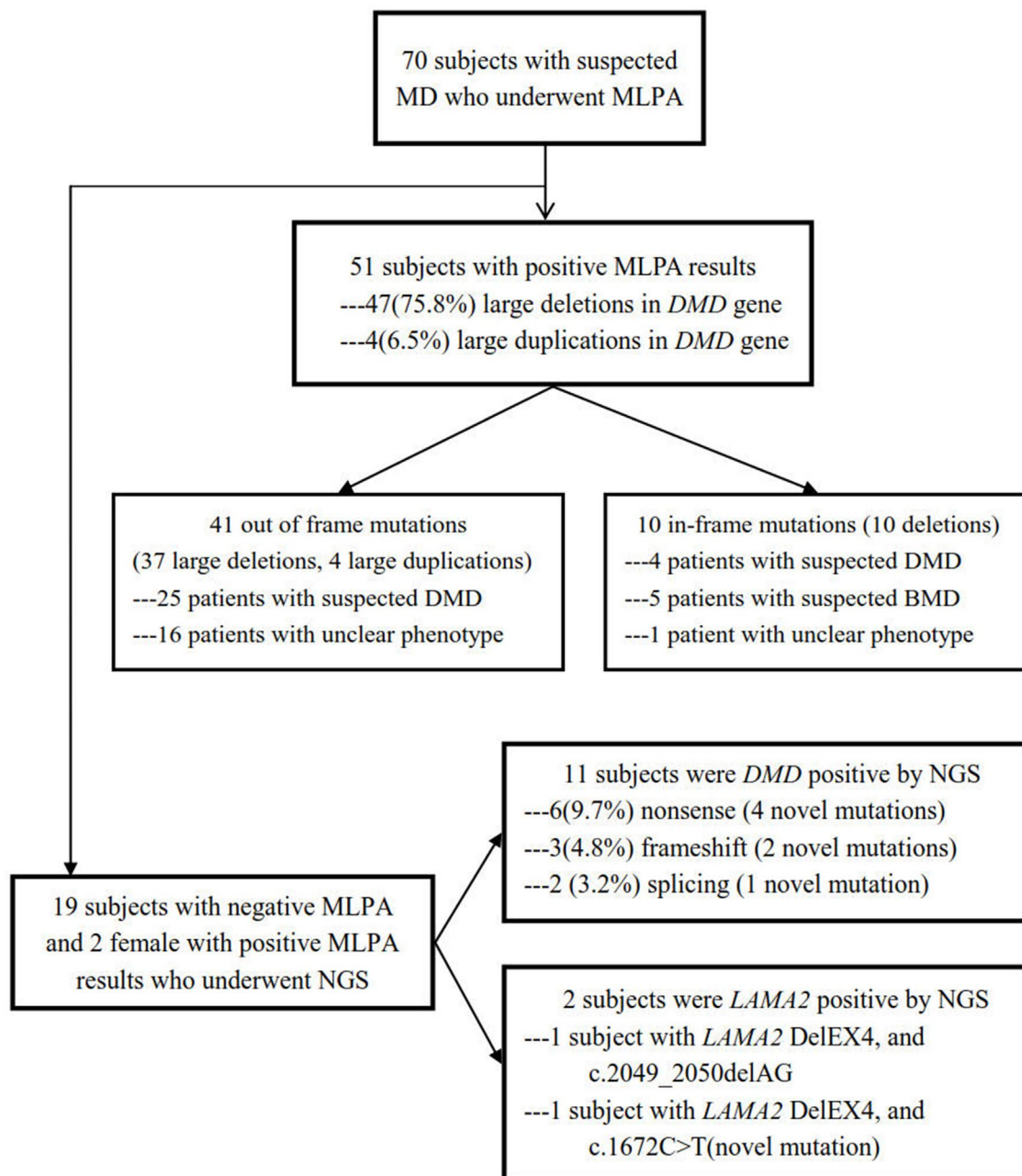


FIGURE 1 | Flow chart of this study design. A total of 70 subjects with suspected MD underwent MLPA, and 51 were positive including 41 out of frame mutations and 10 in-frame mutations, while in the rest of the 19 subjects with negative MLPA and two female with positive MLPA, NGS was performed, with 11 positive in *DMD* and 2 positive in *LAMA2*.

SureSelect captured Exome probes (Agilent, Santa Clara, CA, USA). The quality and quantity of the libraries were assessed by both Advanced Analytical Technologies Inc. (AATI) Fragment Analyzer (Ankeny, Des Moines, IA, USA) and qPCR. Purified sequencing libraries were pooled together and massively parallel-sequenced using Illumina HiSeq X Ten platform yielding an average of 1.5 Gb of total sequence per sample at an average sequencing depth of 100×. The data were then aligned to version

GRCh37/hg19 of the human genome in NextGENe Software v2.3.4 (SoftGenetics, State College, PA, USA) that aligned sequence reads to the reference. NextGENe Software uses a preloaded index alignment algorithm that employs a suffix array that is represented by the Burrows-Wheeler Transform (BWT).

Both the general pipelines for NGS data analyses of diseases and data analysis strategies for the pathogenic genes of MD were conducted according to the recommendations from Jin et al. (2018).

Variants were classified according to the 2015 ACMG guideline for the interpretation of sequence variants. All the mutations that can potentially cause the diseases were verified by Sanger sequencing on a 3130 Genetic Analyzer (ABI, Carlsbad, CA, USA).

Validation of Gene Mutations

Quantitative PCR (qPCR) and Sanger sequencing were then utilized to validate the potentially pathogenic variants of exon deletions/duplications and small mutations in the patients with designed specific primers. The qPCR was amplified according to the manufacturer's recommendations on a real-time PCR system (LightCycler 480 II, Roche, Foster City, USA). Copy number variations were determined based on the ratio of copies of the deletion fragment to a reference gene (*GAPDH*) with the SYBR Premix Ex Taq II PCR reagent kit (TakaRa Bio, Dalian, China) according to the manufacturer's protocol. The PCR products were purified and sequenced using an ABI Prism 3700 automated sequencer (Applied Biosystems, Foster City, CA).

Statistical Analysis

SPSS 17.0 (IBM, Armonk, NY, United States) was used for statistical analyses. Two-sided Fisher's test was used to compare CK value between BMD and UMD. Correlations between phenotypes and factors (age of examination, mutation type, and CK values as mean \pm SD) were analyzed using logistic regression. It was considered statistically significant when *P* value was less than 0.05, and the confidence interval was 95%.

RESULTS

Clinical Findings

In total, 70 Chinese families with 70 clinically suspected MD patients were enrolled in the study, including 42 with a clinically suspected DMD/BMD (36 suspected DMD and 6 suspected BMD), and 28 (named as UMD, mean age 0.64 ± 0.34 years) lack of clinical phenotype due to relatively young age, with only persistent hyperCKemia and typical myopathy manifestation on EMG but normal peripheral nerve conduction velocity.

The average CK values for these patients clinically diagnosed with DMD and BMD were $17,528.33 \pm 10,234.82$ U/L and $6,017.71 \pm 2,890.50$ U/L, respectively, which indicated a significant difference between both categories (*P* < 0.01) (Table 1).

Detection of Mutations in *DMD* by MLPA

Large rearrangements in *DMD* were detected in the 70 probands using MLPA and validated by qPCR. A total of 51 (72.9%) deletions and duplications were found in 51 patients including two girls, 47 (75.8%) of which were deletions, and 4 were duplications (6.5%). Overall, 38 different rearrangements were identified. Among the 51 positive results, 41 showed out of frame mutations (37 large deletions, 4 large duplications) and 10 showed in-frame mutations (deletions). Of the 41 out of frame mutations, 25 were found in clinically diagnosed DMD patients, and 16 in very young UMD patients. Among the patients with 10 in-frame

mutations, 5 BMD, 4 DMD, and 1 UMD were finally determined after reviewing their clinical manifestations (Figure 1, Table 1). There were 11 different single-exon deletions in *DMD* gene identified in 15 patients involving exons 2, 13, 44, 45, 50, 51, 52, 53, 55, 61, and 62, respectively, while 36 cases were found to have multiple exon deletions. The largest deletion covering the whole exons from 1 to 79 in *DMD* was found in a 2-month-old male DMD patient (P1) with a higher CK level of 14,982 U/L (Table 1). The hotspot region was demonstrated between exons 45 and 55 in which 70.6% large deletions were found most frequently, followed by deletions in exons 3–34 (24.4%) (Figure 2); three out of four duplications in *DMD* were detected in the proximal hotspot regions between exons 3 and 25 (Table 1). However, 19 cases failed to detect large deletions or duplications in *DMD* by MLPA.

Detection of Small Mutations Using NGS

NGS were then utilized to detect small mutations associated with MD in the 19 MLPA-negative patients as well as the two female patients showing MLPA-positive (P5 and P30; see Figure 1 and Table 1 for details). Sanger sequencing was used to validate the potentially pathogenic small mutations in the patients and carriers. Overall, 11 point mutations in *DMD* were found in 11 different probands, respectively, which were 3 *de novo* and 8 maternally inherited, including 6 (9.7%) nonsense mutations, 2 (3.2%) small deletions, 2 splice-site mutations (3.2%), and 2 small insertions (1.6%) (Figure 1, Table 2). The point mutations of c.2436C > T, c.7264dupG, c.1231A > T, c.5167G > T, 10187delC, c.7660+1C > G, and c.7792C > T in the patients of P52, P53, P54, P57, P58, P59, and P60 were novel, unreported previously. No small mutations were found in the two MLPA-positive female patients. Meanwhile, the known pathogenic mutation c.2049_2050delAG and novel mutation c.1672C > T in *LAMA2* were detected in two patients of P63 and P64, respectively (Figure 1, Table 2). All the *de novo* mutations were finally predicted to be pathogenic after analysis according to ACMG guideline (Table 2).

Confirmation of *LAMA2* Mutations in Two MD Patients

To further analyze and confirm the possible deletion in *LAMA2* in both patients of P63 and P64, MLPA was performed for detection of *LAMA2* deletion using SALSA MLPA probe sets P391 (MRC-Holland, Amsterdam, the Netherlands) and confirmed that two patients of P63 and P64 harbored DelEX 4 in *LAMA2*. Thus, both compound heterozygous mutations in *LAMA2* were detected in patients of P63 and P64, individually, with P63 carrying c.2049_2050delAG and DelEX 4, whereas P64 having c.1672C > T and DelEX 4, inherited from father and mother, respectively (Table 2). After analyzing and comparing the results with the databases, c.1672C > T (p.Arg2095Ter) was determined as a novel unreported mutation. The clinical symptoms of the two patients showed that they got apparent muscle weakness since the first 6 months of life, hypotonia, poor suck and cry, and delayed motor development, while their parents had normal phenotype. Moreover, clinical laboratory tests showed significantly increased serum CK levels (Table 1) and typical pattern of myopathy on

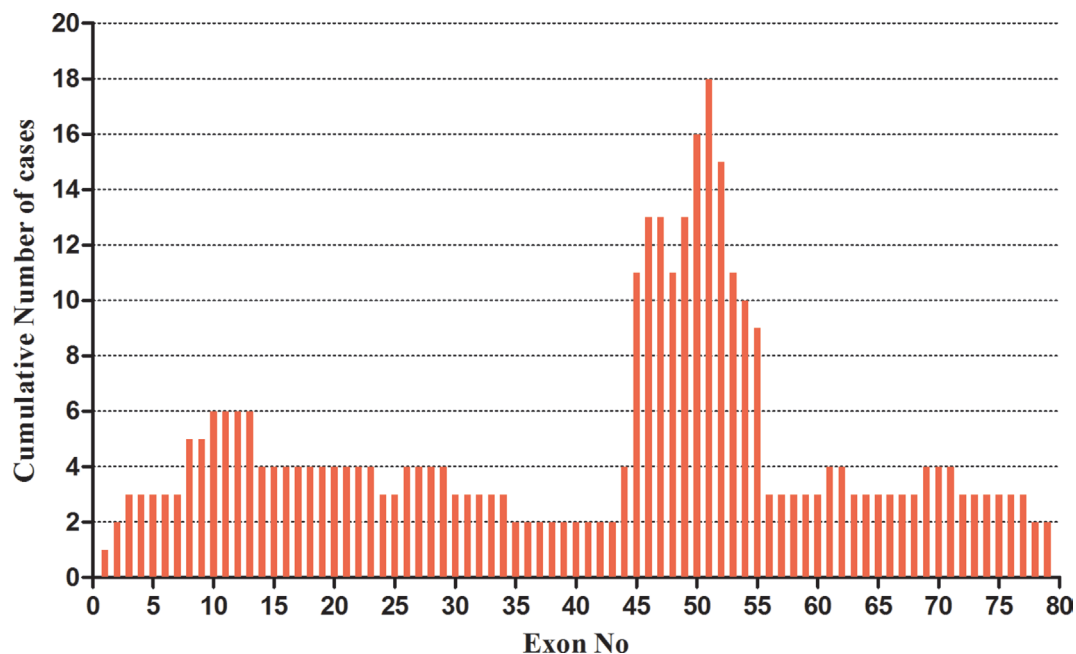


FIGURE 2 | The distribution of gross deletions identified in *DMD* gene by MLPA. The frequency of large deletions for each exon of *DMD* gene was shown. The abscissa showed exon number of *DMD* gene from exons 1 to 79; the ordinate demonstrated the frequency of large deletions for each exon. A region with a highest frequency of deletions was found in exons 45–55.

EMG. So, the two patients were finally diagnosed as Merosin-deficient congenital MD type 1A (MDC1A).

Data Summary

A total of 70 cases were included in this study, of which 62 were *DMD* gene mutations-positive, and 2 were *LAMA2* mutations-positive detected by MLPA plus NGS. Large deletions and duplications in *DMD* gene were detected in 51 patients by MLPA, of which deletions were found in 47 cases, and duplications were in 4 cases. The remaining 19 cases who were MLPA-negative undergone NGS, and 11 small mutations were identified in 15 male cases. All genetic mutations in *DMD* were shown in **Table 2**. The overall positive mutation rate was 91.4% (64/70), encompassing 47 (75.8%) large deletions, 4 (6.5%) large duplications, 6 (9.7%) nonsense mutations, 2 (3.2%) small deletions, 2 (3.2%) splice-site mutations, and 1 (1.6%) small insertion (**Figure 3**).

Final diagnosis was made based on the phenotypes and genotypes in the 70 patients. Of them, 34 DMD, 4 BMD, and 2 MDC1A were made. Additional 24 cases couldn't be differentiated between BMD and DMD due to very young age at present and were diagnosed as BMD/DMD. More research in the future need to be done in the remaining six undiagnosed patients (**Table 3**).

DISCUSSION

MDs are a clinically, genetically, and biochemically heterogeneous group of degenerative skeletal muscle disorders. As one of children's rare diseases, diagnosis of MDs still faces many challenges in

China (Fang et al., 2017; Ni and Shi, 2017). DMD/BMD are the most common types of MD in childhood and caused by the loss of dystrophin function completely or partially (Wein et al., 2015; Yiu and Kornberg, 2015). Muscle biopsy, an invasive technique, showing a specific absence of dystrophin protein in DMD patients while partially functional dystrophin protein in BMD, became the golden standard test in diagnosing MD after the 1960s (Vogel and Zamecnik, 2005; Skram et al., 2009). With the development and availability of genetic techniques, the muscle biopsy has been gradually replaced by gene tests that become the new golden standard as stated in the latest guideline (Birnkrant et al., 2018). MLPA has been used to detect large deletion/duplication (≥ 1 exon) of *DMD* in roughly 70% of cases, and NGS is applied to detect the remaining 25% to 30% small mutations (< 1 exon), such as point mutations and small deletions or insertions, which indicated that MLPA is a simple, rapid, and reliable technique for detection of deletion/duplication in *DMD* gene (Yiu and Kornberg, 2015; Wein et al., 2015; Suh et al., 2017). Moreover, MLPA has been suggested to use as a first-line screening test for detection of DNA rearrangements in *DMD* in clinically suspected DMD/BMD patients (Mohammed et al., 2018).

In this study, MLPA was firstly applied to detect deletions and duplications in *DMD* in 70 patients with suspected MD, and 38 DNA rearrangements in *DMD* gene were detected in 51 (72.9%) patients (comprised 31 suspected DMD/BMD and 20 UMD). Out of the mutations, 47 large deletions (47/70, 67.14%) and 4 large duplications (4/70, 5.71%) were found, indicating a higher proportion of *DMD* deletion in comparison with *DMD* duplication, which is similar as the most previous studies (Danieli

TABLE 2 | Summary of putative pathogenic mutations in *DMD/LAMA2* gene analyzed by NGS and validated by Sanger sequencing.

Patient ID	Gene/exon	DNA change	Effect	Mutation type	dbSNP/1000G/EVS/ExAC	Status	Inheritance	Pathogenicity prediction score		Pathogenic evaluation according to ACMG
								Mutation Taster	Human Splicing Finder	
P52	<i>DMD</i> /20	c.2436C > T	p.Trp812Ter	Nonsense	0/0/0/0	Novel	De novo	1		Pathogenic
P53	<i>DMD</i> /50	c.7264dupG	p.Ala2422GlyfsTer5	Frameshift	0/0/0/0	Novel	Maternal	1		Pathogenic
P54	<i>DMD</i> /11	c.1231A > T	p.Lys411Ter	Nonsense	0/0/0/0	Novel	Maternal	1		Pathogenic
P55	<i>DMD</i> /7	c.580C > T	p.Gln194Ter	Nonsense	0/0/0/0	Flanigan et al., 2009	Maternal	1		Pathogenic
P56	<i>DMD</i> /15	c.1812+1C > T	–	Splicing	0/0/0/0	Piton et al., 2013; Taylor et al., 2007	Maternal	1	1	Pathogenic
P57	<i>DMD</i> /37	c.5167G > T	p.Glu1723Ter	Nonsense	0/0/0/0	Novel	De novo	1		Pathogenic
P58	<i>DMD</i> /70	10187delC	p.Pro3396GlnfsTer6	Frameshift	0/0/0/0	Novel	Maternal	1		Pathogenic
P59	<i>DMD</i> /52	c.7660+1C > G	–	Splicing	0/0/0/0	Novel	Maternal	1	1	Pathogenic
P60	<i>DMD</i> /53	c.7792C > T	p.Gln2598Ter	Nonsense	0/0/0/0	Novel	De novo	1		Pathogenic
P61	<i>DMD</i> /20	c.2571delC	p.Pro857ProfsTer14	Frameshift	0/0/0/0	Li et al., 2015	Maternal	1		Pathogenic
P62	<i>DMD</i> /43	c.6283C > T	p.Arg2095Ter	Nonsense	0/0/0/0	Cho et al., 2017;	Maternal	1		Pathogenic
P63	<i>LAMA2</i> /exon14	c.2049_2050delAG	p.Arg683SerfsTer21	Nonsense	0/0/0/0		Paternal	1		Pathogenic
	<i>LAMA2</i> /exon 4	DelEX 4	–	In-frame	0/0/0/0		Maternal			Pathogenic
P64	<i>LAMA2</i> /exon 4	DelEX 4	–	In-frame	0/0/0/0		Maternal			Pathogenic
	<i>LAMA2</i> /exon 12	c.1672C > T	p.Gln558Ter	Nonsense	0/0/0/0	Novel	Paternal	1		Pathogenic

All the variants shown in **Table 2** are described using the NM_004006.2 of *DMD*/NM_000426.3 of *LAMA2* transcript references, respectively.

del, deletion; dup, duplication; fs, frameshift; Ter, termination; in-frame, in-frame mutation.

ACMG, American College of Medical Genetics and Genomics.

The mutation spectrum of *DMD* gene in the cohort

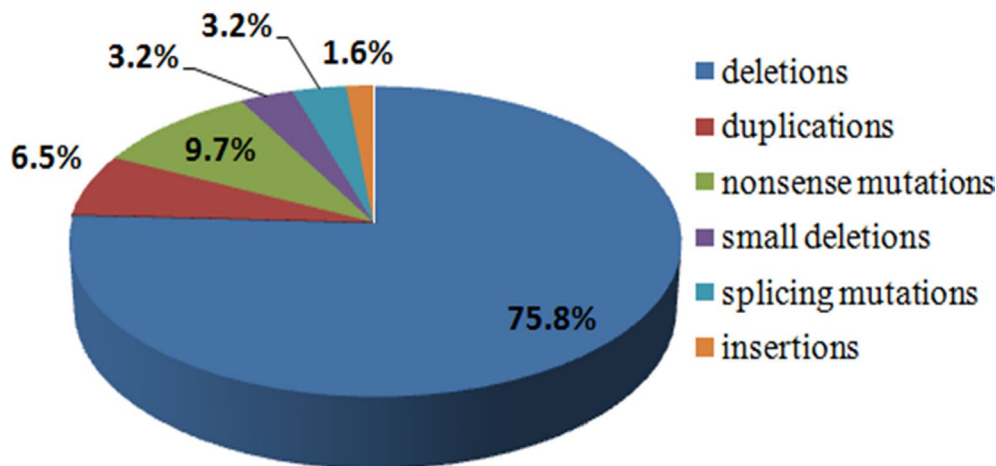


FIGURE 3 | The mutation spectrum of *DMD* gene in the cohort. The mutation spectrum of *DMD* gene in the cohort was shown, including 75.8% large deletions, 6.5% large duplications, 9.7% nonsense mutations, 3.2% small deletions, 3.2% splice-site mutations, and 1.6% small insertion.

TABLE 3 | Final diagnosis of all 70 patients after detection with MLPA and NGS.

Final diagnosis		Genotype	
Disease	Case no. %	Gene: mutation	Case no. %
DMD	P34, 48.6%	<i>DMD</i> : large deletion	24, 34.3%
		<i>DMD</i> : large duplication	4, 5.7%
		<i>DMD</i> : small mutation	6, 8.6%
BMD	P4, 5.7%	<i>DMD</i> : large deletion	3, 4.3%
		<i>DMD</i> : small mutation	1, 1.4%
BMD/DMD*	P24, 34.3%	<i>DMD</i> : large deletion	20, 28.6%
		<i>DMD</i> : small mutation	4, 5.7%
MDC1A	P2, 2.9%	<i>LAMA2</i> : large deletion	2, 2.9%
		<i>LAMA2</i> : small mutation	2, 2.9%
Undiagnosed†	P6, 8.6%	No mutation found	6, 8.6%

*Could not differentiate from each other at present due to very young age of the patients.

†Needs to do more research.

et al., 1993; Muntoni et al., 2003; Chen et al., 2014; Deepha et al., 2017; Okubo et al., 2017; Suh et al., 2017) but different from some reports that demonstrated about 40% deletion and about 25% duplication (Hwa et al., 2007; Wang et al., 2008; Lee et al., 2012). So, further investigation of the differences in different populations from various locations still needs to be done.

Despite distribution of deletions and duplications occurring in almost every exon in *DMD* gene, two deletion hot spot regions with a non-random style have been repeatedly reported at the 5'-end and the central region around exons 44 to 55 (Muntoni et al., 2003; Aartsma-Rus et al., 2006; Suh et al., 2017; Mohammed et al., 2018). In our cohort, we could detect the deletion hot spot in the central region of exons 45 to 55, with a distribution rate in this region of approximately 70.6%, which is in accordance with previous studies (Danieli et al., 1993; Hwa et al., 2007; Lee et al., 2012; Chen et al., 2014; Okubo et al., 2017). Out of four duplications detected in this

study, three showed in the proximal hot spot region from exons 3 to 25 similar as the previous report (Okubo et al., 2017).

With the great advantage of cost-effectiveness and accuracy, NGS has made the detection of small mutations in *DMD* revolutionized (Zhang et al., 2019; Jia and Shi, 2017). To minimize the cost of NGS for detection of mutations, we utilized NGS in the remaining 19 MLPA-negative (13 DMD/BMD and 6 UMD) patients and the two female patients of MLPA-positive. In total, 13 small mutations in 13 (7 DMD/BMD, 6 UMD) patients were found, including 11 *DMD* mutations (three *de novo* and eight maternal inheritances) and 2 *LAMA2* mutations. Among 11 point mutations of *DMD* gene, 6 were nonsense mutations (54.55%), which were the most common point mutations in this study, similar to the report (Okubo et al., 2017), and then followed by splice-site mutations (18.2%), small deletions (18.2%), and small insertions (9.1%); however, no missense mutations were detected in this cohort, which might be due to the small sample size. The distribution of mutations types (six nonsense, three frameshift, and two splicing) in this study was consistent with some studies using different methods (Flanigan et al., 2009; Takeshima et al., 2010; Lim et al., 2011). As expected, G:C to A:T transitions were the most prevalent stop mutation class; our results showed that the transversions of GC to AT accounted for 63.63% (7/11). After mutation verification of the probands' mothers, eight novel (seven in *DMD* and one in *LAMA2*) mutations were identified. All these novel mutations were predicted to generate a premature stop codon in the coding sequence of *DMD* causing premature termination of the protein product lacking key domains of dystrophin protein, and produce a non-functional dystrophin protein thereby leading to DMD. In this way, we got much higher detection rate of 91.4% than muscle biopsies with more precise and sensitive diagnosis of MD.

Two unrelated female patients clinically diagnosed as DMD-like MD (Table 1), whose parents showed normal phenotypes, were detected to carry a heterozygous deletion of exon 50 and exons 8–12 in *DMD* by MLPA, inherited maternally, but no disease-causing small mutations were found by the NGS. As we know, most heterozygous *DMD* mutations' female carriers were asymptomatic, however, although of which 2.5–7.8% were symptomatic with symptoms ranging from mild muscle weakness to a rapidly progressive DMD-like MD (Moser and Emery, 1974; Norman and Harper, 1989; Taylor et al., 2007). Fujii et al. (2009) summarized the mechanism for female DMD and BMD as the following five cases: uniparental disomy of the entire X chromosome with mutations, skewed X inactivation either balanced X-autosome translocation patients or in the female carriers with *dystrophin* mutation, Turner syndrome with a *dystrophin* mutation on the remaining X chromosome, co-occurrence of mutations in both *dystrophin* and androgen receptor genes, and double *dystrophin* mutations on both X chromosomes. In regard of the two female cases in this study, the most probable mechanism is skewed X inactivation in the female *dystrophin* mutation carriers. The further X chromosome inactivation analyses of the two patients are required to confirm the inference.

In this cohort, 28 UMD patients presented only persistent hyperCKemia (CK from 2,391 to 35,340 U/L) and typical myopathy performance on EMG but without typical clinical features due to very young age (average age at 0.64 ± 0.34 years). After detecting mutation using the MLPA plus NGS, 24 mutations in *DMD* were found, including 20 large deletions and 3 nonsense and 1 splice-site mutations. Additionally, two small mutations in *LAMA2* were detected. *Laminin- α 2* (*LAMA2*) is the causative gene to MDC1A (merosin-deficient congenital MD type 1A), which is a type of CMD, an autosomal recessive disorder. So far, more than 493 mutations in *LAMA2* gene have been listed in the locus specific database (LSDB) (April, 2018), in which a limited number of large deletions/duplications have been reported. In this study, deletions of exon 4 in *LAMA2* in the two patients with MDC1A were detected, respectively, which have been reported in five other Chinese patients by Xiong et al. (2015), but have not been found in other countries. So, we inferred that it might be a particular mutation type in Chinese population. In addition, the mutation c.2049_2050delAG in *LAMA2* in patient P63 was a known pathogenic mutation (Guicheney et al., 1998), while c.1672C > T (p.Q558X) in *LAMA2* in patient P64 was a *de novo* nonsense mutation, unreported in population databases (ExAC, dbSNP, and 1,000 genomes), leading to a truncated and nonfunctional laminin- α 2 protein, which was also predicted to be pathogenic by using MutationTaster. Moreover, many pathogenic truncating mutations downstream of this mutation site have been reported in the LSDB for the *LAMA2* gene (<http://www.lovd.nl/LAMA2>). Thus, we inferred that the mutation c.1672C > T (p.Q558X) in *LAMA2* was probably pathogenic.

To the best of our knowledge, this is the first report using the molecular genetics techniques to identify genes' mutations in young patients with MD before their clinical features appeared. And it is critical to perform molecular genetics detection for the UMD patients with indicators like persistent hyperCKemia and typical myopathy performance on EMG but without definite clinical manifestations at a relatively young age. The genotype-first approach

can provide a definitive diagnosis mainly based on molecular evidence (Shen, 2018). This study of using MLPA plus NGS provides a comprehensive and effective genetic approach for precise and early diagnosis of MD, with the diagnostic yield of 91.4% (64/70).

CONCLUSIONS

In this study, we identified 51 large deletions/duplications, 11 small mutations in *DMD* gene, and 2 mutations in *LAMA2* gene by MLPA plus NGS. We finally diagnosed 34 DMD, 4 BMD, 24 BMD/DMD, and 2 MDC1A in the cohort. Our data indicated that the MLPA plus NGS can be a comprehensive and effective tool for precision diagnosis of MD and is particularly necessary for the patients with only two clinical indicators (persistent hyperCKemia and typical myopathy performance on electromyogram) but no definite clinical manifestations at a relatively young age, which would benefit the very young patients for early diagnosis and treatment as well. Moreover, the method is suitable for early precise diagnosis of children with unclear clinical subtypes of MD.

ETHICS STATEMENT

This study was approved by Ethics Committee of Qilu Children's Hospital of Shandong University. Informed written consent has been provided by the guardians of the patients. All the patients' information was submitted anonymously. All the procedures performed in the study were in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

The study was conceived and designed by ZG and YiL. The experiments were conducted by DW, MG, KZ, YuL, JM and YW. Data analyzed by DW, MG, KZ and YL. RJ, YoL and ZG contributed clinical diagnosis of the patients. The paper was written by DW and YL.

FUNDING

This work was funded by Development Project of Science and Technology in Shandong Province (2013GSF11829) and Jinan outstanding scientific and technological innovation team fund (20150519). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank the patients and their families for their participation and contribution to the work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.00814/full#supplementary-material>

REFERENCES

- Aartsma-Rus, A., Van Deutekom, J. C., Fokkema, I. F., Van Ommen, G. J., and Den Dunnen, J. T. (2006). Entries in the Leiden Duchenne muscular dystrophy mutation database: an overview of mutation types and paradoxical cases that confirm the reading-frame rule. *Muscle Nerve* 34 (2), 135–144. doi: 10.1002/mus.20586
- Ahn, A. H., and Kunkel, L. M. (1993). The structural and functional diversity of dystrophin. *Nat. Genet.* 3 (4), 283–291. doi: 10.1038/ng0493-283
- Birnkrant, D. J., Bushby, K., Bann, C. M., Apkon, S. D., Blackwell, A., Brumbaugh, D., et al. (2018). Diagnosis and management of Duchenne muscular dystrophy, part 1: diagnosis, and neuromuscular, rehabilitation, endocrine, and gastrointestinal and nutritional management. *Lancet Neurol.* 17, 251–267. doi: 10.1016/S1474-4422(18)30024-3
- Carter, J. C., Sheehan, D. W., Prochoroff, A., and Birnkrant, D. J. (2018). Muscular dystrophies. *Clin. Chest Med.* 39 (2), 377–389. doi: 10.1016/j.ccm.2018.01.004
- Chen, C., Ma, H., Zhang, F., Chen, L., Xing, X., Wang, S., et al. (2014). Screening of Duchenne muscular dystrophy (DMD) mutations and investigating its mutational mechanism in Chinese patients. *PLoS One* 9 (9), e108038. doi: 10.1371/journal.pone.0108038
- Cho, A., Seong, M. W., Lim, B. C., Lee, H. J., Byeon, J. H., Kim, S. S., et al. (2017). Consecutive analysis of mutation spectrum in the dystrophin gene of 507 Korean boys with Duchenne/Becker muscular dystrophy in a single center. *Muscle Nerve* 55 (5), 727–734. doi: 10.1002/mus.25396
- Danieli, G. A., Mioni, F., Muller, C. R., Vitiello, L., Mostacciuolo, M. L., and Grimm, T. (1993). Patterns of deletions of the dystrophin gene in different European populations. *Hum. Genet.* 91 (4), 342–346. doi: 10.1007/BF00217354
- Darras, B. T., Urion, D. K., and Ghosh, P. S. (1993). “Dystrophinopathies,” in *GeneReviews(R)*. Eds. M. P. Adam, H. H. Ardinger, R. A. Pagon, S. E. Wallace, L. J. H. Bean, K. Stephens and A. Amemiya. Seattle, WA: University of Washington, Seattle.
- Deepha, S., Vengalil, S., Preethish-Kumar, V., Polavarapu, K., Nalini, A., Gayathri, N., et al. (2017). MLPA identification of dystrophin mutations and in silico evaluation of the predicted protein in dystrophinopathy cases from India. *BMC Med. Genet.* 18 (1), 67. doi: 10.1186/s12881-017-0431-6
- Fang, F., Liu, Z., Fang, H., Wu, J., Shen, D., Sun, S., et al. (2017). The clinical and genetic characteristics in children with mitochondrial disease in China. *Sci. China Life Sci.* 60 (7), 746–757. doi: 10.1007/s11427-017-9080-y
- Falsaperla, R., Pratico, A. D., Ruggieri, M., Parano, E., Rizzo, R., Corsello, G., et al. (2016). Congenital muscular dystrophy: from muscle to brain. *Ital. J. Pediatr.* 42 (1), 78. doi: 10.1186/s13052-016-0289-9
- Flanigan, K. M., Dunn, D. M., von Niederhausern, A., Soltanzadeh, P., Gappmaier, E., Howard, M. T., et al. (2009). Mutational spectrum of DMD mutations in dystrophinopathy patients: application of modern diagnostic techniques to a large cohort. *Hum. Mutat.* 30 (12), 1657–1666. doi: 10.1002/humu.21114
- Fujii, K., Minami, N., Hayashi, Y., Nishino, I., Nonaka, I., Tanabe, Y., et al. (2009). Homozygous female Becker muscular dystrophy. *Am. J. Med. Genet. A* 149A (5), 1052–1055. doi: 10.1002/ajmg.a.32808
- Gatta, V., Scariolla, O., Gaspari, A. R., Palka, C., De Angelis, M. V., Di Muzio, A., et al. (2005). Identification of deletions and duplications of the DMD gene in affected males and carrier females by multiple ligation probe amplification (MLPA). *Hum. Genet.* 117 (1), 92–98. doi: 10.1007/s00439-005-1270-7
- Guicheney, P., Vignier, N., Zhang, X., He, Y., Cruaud, C., Frey, V., et al. (1998). PCR based mutation screening of the laminin alpha2 chain gene (LAMA2): application to prenatal diagnosis and search for founder effects in congenital muscular dystrophy. *J. Med. Genet.* 35 (3), 211–217. doi: 10.1136/jmg.35.3.211
- Hegde, M. R., Chin, E. L., Mulle, J. G., Okou, D. T., Warren, S. T., and Zwick, M. E. (2008). Microarray-based mutation detection in the dystrophin gene. *Hum. Mutat.* 29 (9), 1091–1099. doi: 10.1002/humu.20831
- Human Gene Mutation Database [Online]. Available: <https://portal.biobase-international.com/cgi-bin/portal/login.cgi> [Accessed].
- Hwa, H. L., Chang, Y. Y., Chen, C. H., Kao, Y. S., Jong, Y. J., Chao, M. C., et al. (2007). Multiplex ligation-dependent probe amplification identification of deletions and duplications of the Duchenne muscular dystrophy gene in Taiwanese subjects. *J. Formos. Med. Assoc.* 106 (5), 339–346. doi: 10.1016/S0929-6646(09)60318-1
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60 (7), 686–691. doi: 10.1007/s11427-017-9099-3
- Jin, Y., Zhang, L., Ning, B., Hong, H., Xiao, W., Tong, W., et al. (2018). Application of genome analysis strategies in the clinical testing for pediatric diseases. *Pediatr. Invest.* 2 (2), 72–81. doi: 10.1002/ped4.12044
- Lee, B. L., Nam, S. H., Lee, J. H., Ki, C. S., Lee, M., and Lee, J. (2012). Genetic analysis of dystrophin gene for affected male and female carriers with Duchenne/Becker muscular dystrophy in Korea. *J. Korean Med. Sci.* 27 (3), 274–280. doi: 10.3346/jkms.2012.27.3.274
- Li, X., Zhao, L., Zhou, S., Hu, C., Shi, Y., Shi, W., et al. (2015). A comprehensive database of Duchenne and Becker muscular dystrophy patients (0-18 years old) in East China. *Orphanet. J. Rare Dis.* 10, 5. doi: 10.1186/s13023-014-0220-7
- Lim, B. C., Lee, S., Shin, J. Y., Kim, J. I., Hwang, H., Kim, K. J., et al. (2011). Genetic diagnosis of Duchenne and Becker muscular dystrophy using next-generation sequencing technology: comprehensive mutational search in a single platform. *J. Med. Genet.* 48 (11), 731–736. doi: 10.1136/jmedgenet-2011-100133
- Luce, L. N., Carcione, M., Mazzanti, C., Ferrer, M., Szijan, I., and Giliberto, F. (2018). Small mutation screening in the DMD gene by whole exome sequencing of an Argentine Duchenne/Becker muscular dystrophies cohort. *Neuromuscul. Disord.* 28 (12), 986–995. doi: 10.1016/j.nmd.2018.08.012
- Marden, F. A., Connolly, A. M., Siegel, M. J., and Rubin, D. A. (2005). Compositional analysis of muscle in boys with Duchenne muscular dystrophy using MR imaging. *Skeletal Radiol.* 34 (3), 140–148. doi: 10.1007/s00256-004-0825-3
- Mercuri, E., and Muntoni, F. (2013). Muscular dystrophies. *Lancet* 381, 845–860. doi: 10.1016/S0140-6736(12)61897-2
- Moser, H., and Emery, A. E. (1974). The manifesting carrier in Duchenne muscular dystrophy. *Clin. Genet.* 5 (4), 271–284. doi: 10.1111/j.1399-0004.1974.tb01694.x
- Mohammed, F., Elshafey, A., Al-Balool, H., Alaboud, H., Al Ben Ali, M., Bager, A., et al. (2018). Mutation spectrum analysis of Cuchenne/Becker muscular dystrophy in 68 families in Kuwait: the era of personalized medicine. *PLoS One* 13 (5), e0197205. doi: 10.1371/journal.pone.0197205
- Muntoni, F., Torelli, S., and Ferlini, A. (2003). Dystrophin and mutations: one gene, several proteins, multiple phenotypes. *Lancet Neurol.* 2, 731–740. doi: 10.1016/S1474-4422(03)00585-4
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60 (7), 681–685. doi: 10.1007/s11427-017-9100-1
- Niba, E. T., Tran, V. K., Tuan-Pham le, A., Vu, D. C., Nguyen, N. K., Ta, V. T., et al. (2014). Validation of ambiguous MLPA results by targeted next-generation sequencing discloses a nonsense mutation in the DMD gene. *Clin. Chim. Acta* 436, 155–159. doi: 10.1016/j.cca.2014.05.018
- Norman, A., and Harper, P. (1989). A survey of manifesting carriers of Duchenne and Becker muscular dystrophy in Wales. *Clin. Genet.* 36 (1), 31–37. doi: 10.1111/j.1399-0004.1989.tb03363.x
- Okubo, M., Goto, K., Komaki, H., Nakamura, H., Mori-Yoshimura, M., Hayashi, Y. K., et al. (2017). Comprehensive analysis for genetic diagnosis of Dystrophinopathies in Japan. *Orphanet J. Rare Dis.* 12 (1), 149. doi: 10.1186/s13023-017-0703-4
- Piton, A., Redin, C., and Mandel, J. L. (2013). XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. *Am. J. Hum. Genet.* 93 (2), 368–383. doi: 10.1016/j.ajhg.2013.06.013
- Skrum, M. K., Gulati, S., Larsson, E., Lindal, S., and Torp, S. H. (2009). Muscle biopsies in children—an evaluation of histopathology and clinical value during a 5-year period. *Ups J. Med. Sci.* 114 (1), 41–45. doi: 10.1080/03009730802604949
- Shen, Y. (2018). Next-generation sequencing based molecular testing is an equalizer for diagnostic service of rare genetic disorders in China. *Pediatr. Invest.* 2 (2), 96–97. doi: 10.1002/ped4.12036
- Singh, B., Mandal, K., Lallar, M., Narayanan, D. L., Mishra, S., Gambhir, P. S., et al. (2018). Next generation sequencing in diagnosis of MLPA negative cases presenting as duchenne/becker muscular dystrophies. *Indian J. Pediatr.* 85 (4), 309–310. doi: 10.1007/s12098-017-2455-5
- Stuppia, L., Antonucci, I., Palka, G., and Gatta, V. (2012). Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int. J. Mol. Sci.* 13 (3), 3245–3276. doi: 10.3390/ijms13033245
- Suh, M. R., Lee, K. A., Kim, E. Y., Jung, J., Choi, W. A., and Kang, S. W. (2017). Multiplex ligation-dependent probe amplification in X-linked recessive muscular dystrophy in Korean subjects. *Yonsei Med. J.* 58 (3), 613–618. doi: 10.3349/ymj.2017.58.3.613

- Takeshima, Y., Yagi, M., Okizuka, Y., Awano, H., Zhang, Z., Yamauchi, Y., et al. (2010). Mutation spectrum of the dystrophin gene in 442 Duchenne/Becker muscular dystrophy cases from one Japanese referral center. *J. Hum. Genet.* 55 (6), 379–388. doi: 10.1038/jhg.2010.49
- Taylor, P. J., Maroulis, S., Mullan, G. L., Pedersen, R. L., Baumli, A., Elakis, G., et al. (2007). Measurement of the clinical utility of a combined mutation detection protocol in carriers of Duchenne and Becker muscular dystrophy. *J. Med. Genet.* 44 (6), 368–372. doi: 10.1136/jmg.2006.047464
- Vogel, H., and Zamecnik, J. (2005). Diagnostic immunohistology of muscle diseases. *J. Neuropathol. Exp. Neurol.* 64 (3), 181–193. doi: 10.1093/jnen/64.3.181
- Wang, X., Wang, Z., Yan, M., Huang, S., Chen, T. J., and Zhong, N. (2008). Similarity of DMD gene deletion and duplication in the Chinese patients compared to global populations. *Behav. Brain Funct.* 4, 20. doi: 10.1186/1744-9081-4-20
- Wang, Y., Yang, Y., Liu, J., Chen, X. C., Liu, X., Wang, C. Z., et al. (2014). Whole dystrophin gene analysis by next-generation sequencing: a comprehensive genetic diagnosis of Duchenne and Becker muscular dystrophy. *Mol. Genet. Genomics* 289 (5), 1013–1021. doi: 10.1007/s00438-014-0847-z
- Wang, L., Xu, M., Li, H., He, R., Lin, J., Zhang, C., et al. (2019). Genotypes and phenotypes of DMD small mutations in chinese patients with dystrophinopathies. *Front. Genet.* 10, 114. doi: 10.3389/fgene.2019.00114
- Wein, N., Alfano, L., and Flanigan, K. M. (2015). Genetics and emerging treatments for Duchenne and Becker muscular dystrophy. *Pediatr. Clin. North Am.* 62 (3), 723–742. doi: 10.1016/j.pcl.2015.03.008
- Xiong, H., Tan, D., Wang, S., Song, S., Yang, H., Gao, K., et al. (2015). Genotype/phenotype analysis in Chinese laminin-alpha2 deficient congenital muscular dystrophy patients. *Clin. Genet.* 87 (3), 233–243. doi: 10.1111/cge.12366
- Yiu, E. M., and Kornberg, A. J. (2015). Duchenne muscular dystrophy. *J. Paediatr. Child Health* 51 (8), 759–764. doi: 10.1111/jpc.12868
- Zhang, K., Yang, X., Lin, G., Han, Y., and Li, J. (2019). Molecular genetic testing and diagnosis strategies for dystrophinopathies in the era of next generation sequencing. *Clin. Chim. Acta* 491, 66–73. doi: 10.1016/j.cca.2019.01.014
- Zimowski, J. G., Pawelec, M., Purzycka, J. K., Szirkowiec, W., and Zaremba, J. (2017). Deletions, not duplications or small mutations, are the predominante new mutations in the dystrophin gene. *J. Hum. Genet.* 62 (10), 885–888. doi: 10.1038/jhg.2017.70

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wang, Gao, Zhang, Jin, Lv, Liu, Ma, Wan, Gai and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



COL1A1/2 Pathogenic Variants and Phenotype Characteristics in Ukrainian Osteogenesis Imperfecta Patients

Lidiia Zhytnik^{1*}, Katre Maasalu^{1,2}, Andrey Pashenko³, Sergey Khmyzov³, Ene Reimann^{4,5}, Ele Prans⁵, Sulev Kõks⁶ and Aare Märtson^{1,2}

OPEN ACCESS

Edited by:

Mike Mikailov,
United States Food and Drug
Administration,
United States

Reviewed by:

Nelson L. S. Tang,
The Chinese University of
Hong Kong, China
Lars Folkestad,
Odense University Hospital,
Denmark
Samia Ali Temtamy,
National Research Centre (Egypt),
Egypt

*Correspondence:

Lidiia Zhytnik
Lidiia.zhytnik@ut.ee

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 23 November 2018

Accepted: 10 July 2019

Published: 09 August 2019

Citation:

Zhytnik L, Maasalu K, Pashenko A,
Khmyzov S, Reimann E, Prans E,
Kõks S and Märtson A (2019)
COL1A1/2 Pathogenic Variants
and Phenotype Characteristics in
Ukrainian Osteogenesis
Imperfecta Patients.
Front. Genet. 10:722.
doi: 10.3389/fgene.2019.00722

¹ Department of Traumatology and Orthopedics, University of Tartu, Tartu, Estonia, ² Clinic of Traumatology and Orthopedics, Tartu University Hospital, Tartu, Estonia, ³ Department of Pediatric Orthopedics, Sytenko Institute of Spine and Joint Pathology, AMS Ukraine, Kharkiv, Ukraine, ⁴ Centre of Translational Medicine, University of Tartu, Tartu, Estonia, ⁵ Department of Pathophysiology, University of Tartu, Tartu, Estonia, ⁶ Perron Institute for Neurological and Translational Science, QEII Medical Centre, Nedlands, WA, Australia

Osteogenesis imperfecta (OI) is a hereditary bone disorder caused by defects of type I collagen. Although up to 90% of patients harbor pathogenic variants in the COL1A1/2 gene, which codes for collagen α 1/2 chains, the spectrum of OI genotypes may differ between populations, and there is academic controversy around OI genotype-phenotype correlations. In the current study, 94 Ukrainian OI families were interviewed. Clinical and genealogical information was collected from patients in spoken form, and their phenotypes were described. To identify the spectrum of collagen I pathogenic variants, COL1A1/2 mutational analysis with Sanger sequencing was performed on the youngest affected individual of every family. Of the 143 patients investigated, 67 (46.85%) had type I OI, 24 (16.78%) had type III, 49 (34.27%) had type IV, and 13 (9.09%) had type V. The mean number of fractures suffered per patient per year was 1.32 ± 2.88 (type I 0.50 ± 0.43 ; type III 3.51 ± 6.18 ; type IV 1.44 ± 1.77 ; and type V 0.77 ± 0.23). 87.23% of patients had skeletal deformations of different severity. Blue sclera, dentinogenesis imperfecta, and hearing loss were present in 87%, 55%, and 22% of patients, respectively. COL1A1/2 pathogenic variants were harbored by 60 patients (63.83%). 27 pathogenic variants are described herein for the first time. The majority of the pathogenic variants were located in the COL1A1 gene (76.19%). Half (49.21%) of the pathogenic variants were represented by structural variants. OI phenotype severity was highly correlated with type of collagen I defect. The current article presents an analysis of the clinical manifestations and COL1A1/2 mutational spectrum of 94 Ukrainian OI families with 27 novel COL1A1/2 pathogenic variants. It is hoped that this data and its analysis will contribute toward the increased understanding of the phenotype development and genetics of the disorder.

Keywords: osteogenesis imperfecta, collagen I, COL1A1, COL1A2, Sanger sequencing, bone disorder

INTRODUCTION

Osteogenesis imperfecta (OI) is a group of rare congenital disorders of the connective tissue, also known as a brittle bone disease. Up to 90% of OI is caused by collagen type I structural (i.e. qualitative) or haploinsufficiency (i.e. quantitative) defects (Shapiro, 2014). OI patients suffer from low bone mass, which results in pathological fractures and skeletal deformities. According to the severity of a collagen defect, patients typically develop bowing of long bones, rib cage deformations, scoliosis and kyphosis, triangular shape of the head, and short stature. Being the most abundant structural protein in the body, collagen type I is also altered in other tissues, causing dentinogenesis imperfecta (DI), blue sclera, muscle weakness, ligamentous laxity, easy bruising, cardiac valve and pulmonary abnormalities, and conductive or sensory hearing loss (Marini et al., 2007; Marini et al., 2017).

OI is one of the most common skeletal dysplasias among orphan diseases. Its prevalence is estimated to be 1/20,000, although this may be affected by OI type and diagnostic practice, as many mild OI cases remain underdiagnosed (Byers and Steiner, 1992). OI phenotypes range from mild osteopenia to severe perinatal lethal forms (Kocher and Shapiro; Roughley et al., 2003; Sillence et al., 1979). The clinical classification distinguishes five main OI types (Sillence et al., 1979). The four classical Sillence OI types are: type I (mild non-deforming OI with blue sclera); type II (perinatal lethal); type III (severe progressive deforming); and type IV (moderate varied OI). Type V OI involves the ossification of interosseous membranes (Glorieux et al., 2000; Amor et al., 2011; Cho et al., 2012; Semler et al., 2012).

Regardless of the various genetic causes of the remaining 10% of OI cases, their clinical manifestations usually coincide with the phenotypes of individuals with *COL1A1/2* pathogenic variants. These OI forms are autosomal recessive and arise due to homozygous pathogenic variants in the genes, which alter collagen transport, folding, post-translational modification, bone mineralization, and cell signaling and bone cell function (Marini et al., 2017). However interconnections between OI phenotype and genotype exist to a certain extent, many carriers of the same mutations might develop different phenotypes and the factors influencing additional phenotype modification remain unidentified (Marini et al., 2007).

Interviews were conducted with 94 Ukrainian OI families in order to gather genealogical information and clinical history and to describe the phenotypes of affected individuals. A *COL1A1/2* mutational analysis of the youngest affected member of each family was performed, to reveal the spectrum of collagen type I pathogenic variants in the Ukrainian OI population. Afterwards, genotype-phenotype correlations in the Ukrainian OI cohort were examined.

This current study presents for the first time the clinical and molecular characteristics of the Ukrainian OI population. Ukrainian OI population enriches pool of known *COL1A1/2* pathogenic variants with 27 novel pathogenic variants. We suppose that this data may advance the understanding of the

OI genetic epidemiology; broaden current knowledge about the spectrum of *COL1A1/2* pathogenic variants, patient's phenotypes, and clinical manifestations; and help to estimate the scope of genotype-phenotype correlations.

SUBJECTS AND METHODS

Medical interviews were conducted with 143 individuals affected with OI, from 94 Ukrainian families. Patients were classified according to the updated Sillence OI classification types I–V (Warman et al., 2011). Clinical history collected by reviewing available medical documents and by patients declared medical history. Clinical examination and phenotypes description was done by UT medical team. Phenotypes were described on the basis of observation and available clinical documentation. Mutational analysis of the *COL1A1/2* genes was conducted with Sanger sequencing, in order to reveal the *COL1A1/2* mutational spectrum. OI skeletal and extraskeletal manifestations were compared with pathogenic variant type, to determine genotype-phenotype correlations.

The current study was conducted in accordance with the Helsinki Declaration and received approval from the Sytenko Institute of Spine and Joint Pathology of the Ukrainian Academy of Medical Sciences and the Ethical Review Committee on Human Research of the University of Tartu (permit no. 221/M-34). Informed written consent from the patients or their legal representatives was obtained prior to inclusion in the study.

Subjects

Ukrainian OI families from the Ukrainian Association of Crystal People participated in the study. In May 2016 and September 2017, Ukrainian OI patients and their relatives (from all regions of Ukraine) attended an interview and clinical examination with researchers from the University of Tartu, Estonia, in cooperation with Ukrainian medical staff. Patients with other skeletal disorders were excluded from the study during screening (five families).

A total number of 143 unrelated OI patients (66 males and 77 females; aged from 2 months to 65 years) from 94 unrelated families were included in the study. Mutational analysis of the *COL1A1/2* genes was performed on the youngest affected member of every OI kindred included in the study ($n = 94$).

Clinical Characteristics and Genealogical Description

In order to characterize clinical OI manifestations, patients underwent both clinical and physical examinations. Cases were classified as OI types I–V, according to the observed clinical features, based on severity of the symptoms (Sillence et al., 1979). Patients with mild, non-deforming OI were classified as type I. Patients with moderate variable OI were indicated as type IV. Patients with severe progressive deforming OI were designated as type III. Individuals with signs of calcification of interosseous membranes were enrolled as type V.

Clinical data was registered based on medical documentation. Genealogical data was recorded from the patients' spoken

Abbreviations: DI, dentinogenesis imperfecta; OI, osteogenesis imperfecta.

words. Blood samples were obtained (for DNA analysis) from all available affected family members and their close healthy relatives.

Phenotype description was assessed by clinical observation. Skeletal fractures and deformations (severity, location) and extra-skeletal OI features (sclera color, DI) were noted. All sclera shades on the blue-gray scale were defined as “blue.” Phenotypic data was provided by the medical records, patients, and their relatives, including birth data (weight, height, intrauterine and birth fractures, preterm pregnancy); anthropometric data (weight and height); fracture history (time and location of the first fracture, total number of fractures, number of fractures per year); patient physical mobility; occurrence of hearing loss; and joint laxity. In order to exclude bias, the registration of the phenotype and the OI type classification were performed by a single medical professional. Fracture per year and total fracture values are presented to make data comparable with previous studies. Data regarding additional OI symptoms and features, including pulmonary function, cardiovascular system, BMD was incomplete and incomparable thus excluded from analysis in current study.

Genealogical data included OI history in the family, consanguinity data, and miscarriages. Pedigree trees were constructed for every kindred with the “Kinship2” package of the R statistical program v3.3.2. (R team, Austria).

The Shapiro–Wilk’s test was used to check normality of continuous variables. Normally distributed continuous data was presented as mean and the standard deviation (SD), continuous data lacking normal distribution is presented as median and range. The Student’s *t*-test was used to compare normally distributed continuous data, and Mann–Whitney *U* test was used to compare data without normal distribution.

The categorical data was expressed as percentages. The significance of associations between the genotype and phenotype manifestations was tested with Fisher’s χ^2 -test for categorical variables. *P*-values less than 0.05 were considered to be statistically significant. All statistical analyses were performed using R v3.3.2. software (R Team, Austria) (Chen et al., 2012).

Mutational Analysis of the COL1A1/2 Genes

Genomic DNA purification was done with 3 ml of an EDTA-preserved whole blood sample, using the Gentra Puregene Blood Kit (Qiagen, Germany) in accordance with the manufacturer’s protocol and stored at -80°C . gDNA samples were amplified with PCR using a 25 specially designed primer pairs covering the 5’ and 3’ untranslated region and 51 exons of the *COL1A1* gene; 36 primer pairs covering the 5’ UTR and 3’ UTR regions and 52 exons of the *COL1A2* gene. The PCR reaction was performed in a total volume of 20 μl , which included 4 μl of 5 \times HOT FIREPol® Blend.

Master Mix Ready to Load with 7.5 mM MgCl₂ (Solis BioDyne, Estonia), 1 μl each of forward and reverse primer (5 pmol), and 1 μl of gDNA (50 ng). PCR reaction was performed with a Thermal Cycler (Applied Biosystems, USA) PCR machine. The PCR touchdown program has been previously described

(Ho Duy et al., 2016; Zhytnik et al., 2017). PCR products were electrophoresed through a 1.5% agarose gel, to control the quality of fragments. The PCR products then purified with exonuclease I and shrimp alkaline phosphatase (Thermo Fisher Scientific, USA). Sanger sequencing reactions were performed on the purified PCR fragments using a BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Reactions were processed on the ABI3730xl instrument. Applied Biosystems’ Sequence Scanner v1.0 and Mutation Surveyor DNA Variant analysis software v5.0.1. (SoftGenetics, USA) were used to analyze sequence products. Sequence products were further aligned to the GenBank human reference genome sequences of *COL1A1* (gDNA NG_007400.1, complementary (cDNA) NM_000088.3) and *COL1A2* (gDNA NG_007405.1, cDNA NM_000089.3). Sequencing data is available from the authors upon reasonable request. This study focused on the non-synonymous and splice-site variants absent from the publicly available normal datasets (including dbSNP135 and the 1000 Genomes Project) (Consortium T 1000 GP, 2015; Sherry et al., 2001). PolyPhen-2, SIFT, and MutationTaster software tools were used to assess the pathogenic nature and functional impact of discovered variants (Kumar et al., 2009; Adzhubei et al., 2010; Schwarz et al., 2010). Variants not described in osteogenesis imperfecta variant database (<http://www.le.ac.uk/ge/collagen/>) were classified as novel (Dagleish, 1997; Dagleish, 1998; Ho Duy et al., 2016; Zhytnik et al., 2017). Mutational analysis of the patients with OI type V will be reported elsewhere.

All laboratory procedures and data analyses were performed at the University of Tartu, Estonia. The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

RESULTS

Clinical Characteristics of Ukrainian OI Population

A total number of 143 individuals from 94 unrelated OI families (66 males and 77 females) were included in the study. Within this cohort, there were no cases of consanguineous families. Out of 93 families with known family history, 36 families had a previous history of OI. For one subject, there was no parent information, as he had grown up in an orphanage. Patients were classified in accordance with the new Silience classification, as follows: type I, *n* = 67 (46.85%); type III, *n* = 24 (16.78%); type IV, *n* = 49 (34.27%); and type V, *n* = 3 (2.10%). In the current study, there were no cases of OI type II (Figure 1A).

The age range of the patients was from 2 months to 65 years, with a mean age of 19.5 ± 14.4 years. Slightly more than half of the study subjects (*n* = 77, 54.23%) were in the “child” age group (i.e. between 0 and 17 years). A detailed description of the skeletal and nonskeletal clinical characteristics of the Ukrainian patient cohort is present in Table 1.

The median/(range) of total number of fractures per patient was 10.00 (0–300) (type I 5.00 [0–30], type III 20.00 [1–300], type IV 11.00 [0–60], and type V [median \pm SD] 12.67 ± 11.72) (Figures 1B and 2). The number of fractures per year per patient

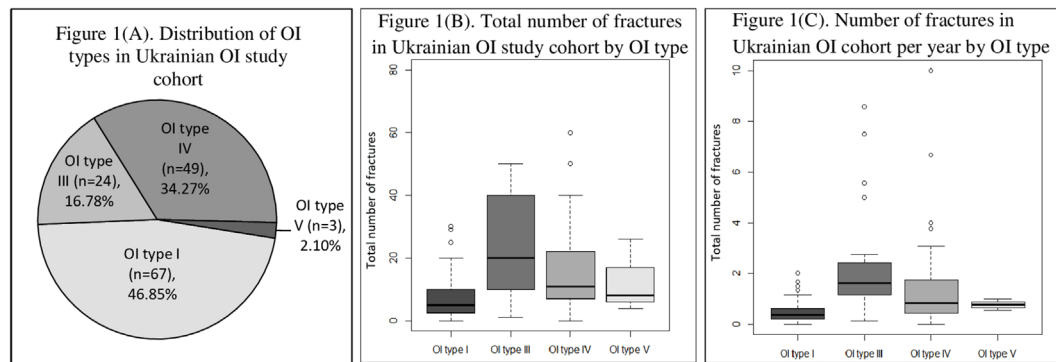


FIGURE 1 | (A) Distribution of osteogenesis imperfecta (OI) types in Ukrainian OI study cohort. **(B)** Total number of fractures in Ukrainian OI study cohort by OI type. **(C)** Number of fractures in Ukrainian OI study cohort per year by OI type.

was 1.32 ± 2.88 (type I 0.50 ± 0.43 , type III 3.51 ± 6.18 , type IV 0.83 ± 1.76 , and type V 0.77 ± 0.23) (Figure 1C). The highest number of fractures in an individual patient ($n = 300$) occurred in a 35-year old patient with OI type III. Nine patients suffered intrauterine fractures (type III 30.43%, $n = 7$; type IV 4.17%, $n = 2$). 19 patients suffered their first fracture during delivery (type I 8.96%, $n = 6$; type III 26.09%, $n = 6$; type IV 12.50%, $n = 6$; type V 33.33%, $n = 1$). The ages of first fracture for study participants was between 0 and 1 years for 22.70% of patients ($n = 32$); between 1 and 2 years for 24.82% ($n = 35$); between 3 and 6 years for 17.02% ($n = 24$); and at ≥ 7 years for 10.64% ($n = 15$). Seven patients (4.96%) did not experience fractures; of these, one was 50 years old (type I) and one was 33 years old (type IV). However, based on extraskeletal features, skeletal deformations, and positive OI history in the family, both these individuals were diagnosed with OI. The most commonly fractured bones were tubular bones: lower limbs (53.73%, $n = 72$); upper limbs (22.39%, $n = 30$); and both lower and upper limbs (17.16%, $n = 23$).

Deformities were present in the majority of the affected study participants (87.23%, $n = 123$). Type I patients mostly had either no deformities or mild deformities. The majority (81.56%, $n = 115$) of subjects had deformities of the lower limbs of varying severity: severe $n = 21$ (type III, $n = 16$; type IV, $n = 4$; type V, $n = 1$); moderate $n = 47$ (type I, $n = 12$; type III, $n = 7$; type IV, $n = 27$, type V, $n = 1$); and mild (type I, $n = 33$, type IV, $n = 14$). Spine deformations (scoliosis and kyphosis) were present in 77.30% ($n = 109$) of patients. More than half of the patients suffered from deformations of the upper limbs (65.96%, $n = 93$). Approximately one third of patients had chest deformities (34.75%, $n = 49$). Most patients (74.29%, $n = 104$) were able to walk independently; 12.86% ($n = 18$) were able to walk with a support; and 12.14% ($n = 17$) used a wheelchair. Two patients were immobile (type III).

DI was observed in 54.68% ($n = 76$) of the subjects (type I, 41.79%; type III, 70.83%; type IV, 61.22%; type V, 33.33%). Joint laxity was present in 26.62% ($n = 37$) of patients (type I, 24.24%; type III, 27.27%; type IV, 26.53%; type V, 66.67%). Two patients (types I and III) had contractures. The majority of the studied individuals (124 individuals, 87.32%) had blue eye sclera (type I, 85.07%; type 3, 86.96%; type 4, 89.80%; and type V, 100%).

Hearing loss was noted in 22.38% of patients ($n = 32$). Of these, 22.38% were type I ($n = 15$), 20.83% were type III ($n = 5$), 22.45% were type IV ($n = 11$), and 33.33% were type IV ($n = 1$). Congenital hearing loss was diagnosed in three patients (types I, III, and IV). Early hearing loss (at between 1 and 10 years) was noted in 12 patients. Nine patients had lost hearing between the ages of 11–20 years. Eight patients experienced hearing loss from maturity (aged over 20 years).

Spectrum of COL1A1/2 Pathogenic Variants in the Ukrainian OI Population

The mutational analysis highlighted that 60 (63.83%) of the 94 Ukrainian OI families harbored pathogenic COL1A1/2 variants (Figure 3A). The number of patients harboring COL1A1/2 pathogenic variants by OI type was as follows: type I, 23 (63.89%); type III, 14 (60.87%); type IV, 23 (69.70%). The number of pathogenic variants was 63, as three patients harbored double pathogenic variants (UA08, UA85 in both COL1A1/2; UA55 in COL1A1). A list of all observed variants is given in Table 2.

The number of COL1A1 pathogenic variants was 48/63 (76.19%), whereas in the COL1A2 gene, 15/63 (23.81%) variants were observed (Figure 3B). All pathogenic variants were in a heterozygous state, underlying the dominant inheritance pattern of the disorder in this cohort. Novel pathogenic variants were represented by 27 (42.85%) variants (20 in COL1A1; 7 in COL1A2) absent from the *osteogenesis imperfecta* variant database (Table 2).

Structural variants comprised slightly more than half of revealed pathogenic variants (31/63, or 49.21%), whereas haploinsufficiency variants were observed in 32/63 (50.79%) of cases (Figure 4A).

Of the 31 structural pathogenic variants, 18 (58.06%) were situated in the COL1A1 gene and 13 (41.94%) in the COL1A2 gene. Glycine substitution was present in 24 (77.42%) of the missense variants, of which 14 (58.33%) and 10 (41.67%) were in the COL1A1 and COL1A2 genes, respectively. Glycine substitution with serine was present in 11 (45.83%) Gly missense variants, seven in the COL1A1, and four in the COL1A2 genes (Figures 4B–D). Two individuals harbored missense non-Gly

TABLE 1 | Clinical characteristics of the study cohort of Ukrainian osteogenesis imperfecta (OI) patients divided by clinical OI type.

	OI type I	OI type III	OI type IV	OI type V	All OI types
Individuals	67 (46.85%)	24 (16.78%)	49 (34.27%)	3 (2.10%)	143 (100%)
Families with OI history (<i>n</i> = 93)	19 (52.78%)	2 (8.70%)	14 (42.42%)	1 (50%)	36 (38.71%)
Gender male/female	35/32	4/20	25/24	2/1	66/77
Age, years (mean ± SD, range) (<i>n</i> = 142)	20.3 ± 15.9 (2–65y)	16.3 ± 11.0 (2m–35y)	20.0 ± 13.9 (6m–50y)	17.7 ± 15.2 (4–34y)	19.5 ± 14.4 (2m–65y)
Children (0–17y)	37 (55.22%)	13 (54.17%)	25 (51.02%)	2 (66.67%)	77 (54.23%)
0–5y	17 (25.37%)	4 (16.67%)	8 (16.33%)	1 (33.33%)	30 (20.98%)
6–10y	8 (11.94%)	5 (20.83%)	8 (16.33%)	0 (0.00%)	21 (14.69%)
10–17y	12 (17.91%)	4 (16.67%)	9 (18.37%)	1 (33.33%)	26 (18.18%)
Total fracture number (<i>n</i> = 142) (median [range]/mean ± SD)	5.00/(0–30)	20.00/(1–300)	11.00/(0–60)	12.67 ± 11.72	10.00/(0–300)
Fractures per year (<i>n</i> = 142) (median [range]/mean ± SD)	0.41 (0–2)	1.63 (0.14–30)	0.83 (0–10)	0.77 ± 0.23	0.61 (0–30)
Time of the first fracture (<i>n</i> = 141)					
No fractures	6 (8.96%)	0 (0.00%)	1 (2.08%)	0 (0.00%)	7 (4.96%)
Intrauterine	0 (0.00%)	7 (30.43%)	2 (4.17%)	0 (0.00%)	9 (6.38%)
During delivery	6 (8.96%)	6 (26.09%)	6 (12.50%)	1 (33.33%)	19 (13.48%)
≤1y	11 (16.42%)	6 (26.09%)	14 (29.17%)	1 (33.33%)	32 (22.70%)
1–2y	19 (28.36%)	1 (4.35%)	14 (29.17%)	1 (33.33%)	35 (24.82%)
3–6y	14 (20.90%)	3 (13.04%)	7 (14.58%)	0 (0.00%)	24 (17.02%)
≥7y	11 (16.42%)	0 (0.00%)	4 (8.33%)	0 (0.00%)	15 (10.64%)
Most fractured body compartment (<i>n</i> = 134)					
All tubular bones	9 (14.75%)	3 (13.04%)	11 (23.40%)	0 (0.00%)	23 (17.16%)
Lower limb	25 (40.98%)	18 (78.26%)	27 (57.45%)	2 (66.67%)	72 (53.73%)
Upper limb	20 (32.79%)	2 (8.70%)	7 (14.89%)	1 (33.33%)	30 (22.39%)
Phalanx	6 (9.84%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	6 (4.48%)
Spine	1 (1.64%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.75%)
Ribs	0 (0.00%)	0 (0.00%)	2 (4.26%)	0 (0.00%)	2 (1.49%)
Deformations (<i>n</i> = 141)					
Upper limb	32 (47.76%)	21 (91.30%)	38 (79.17%)	2 (66.67%)	93 (65.96%)
Lower limb	45 (67.16%)	23 (100.00%)	45 (93.75%)	2 (66.67%)	115 (81.56%)
Spine	40 (59.70%)	23 (100.00%)	44 (91.67%)	2 (66.67%)	109 (77.30%)
Chest	6 (8.96%)	22 (95.65%)	19 (39.58%)	2 (66.67%)	49 (34.75%)
Mobility (<i>n</i> = 140)					
Immobile	0 (0.00%)	1 (4.35%)	0 (0.00%)	1 (33.33%)	2 (1.43%)
Wheelchair	0 (0.00%)	13 (56.52%)	4 (8.51%)	0 (0.00%)	17 (12.14%)
Walking with support	2 (2.99%)	9 (39.13%)	7 (14.89%)	0 (0.00%)	18 (12.86%)
Walking independently	66 (98.51%)	0 (0.00%)	36 (76.60%)	2 (66.67%)	104 (74.29%)
DI (yes) (<i>n</i> = 139)	28 (41.79%)	17 (70.83%)	30 (61.22%)	1 (33.33%)	76 (54.68%)
Blue sclera (yes)	57 (85.07%)	20 (86.96%)	44 (89.80%)	3 (100%)	124 (87.32%)
Hearing loss (yes)	15 (22.38%)	5 (20.83%)	11 (22.45%)	1 (33.33%)	32 (22.38%)
Time of hearing loss start (<i>n</i> = 32)					
at birth	1 (6.67%)	1 (20.00%)	1 (9.09%)	0 (0.00%)	3 (9.38%)
1–10y	5 (33.33%)	2 (40.00%)	5 (45.45%)	0 (0.00%)	12 (37.50%)
11–20y	5 (33.33%)	2 (40.00%)	1 (9.09%)	1 (100.00%)	9 (28.13%)
21–30y	3 (20.00%)	0 (0.00%)	4 (36.36%)	0 (0.00%)	7 (21.88%)
31–40y	1 (6.67%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (3.13%)

Data are *n* (%) unless otherwise indicated. SD is standard deviation.

substitution pathogenic variants in the C-terminal propeptide of the collagen α1 chain: UA21 c.3655G > GT (p.[Asp1219Tyr]) and UA71 c.4356G > GC (p.[Gln1452His]).

Thirty (93.75%) of the haploinsufficiency variants arose in the *COL1A1* gene and 2 (6.25%) in the *COL1A2*. Frameshift pathogenic variants were represented by 12 (37.50%) variants, 11 (91.67%) in the *COL1A1* gene, and one (8.33%) in the *COL1A2*. Furthermore, 11 (34.38%) splice site variants were discovered (nine of *COL1A1*; one of *COL1A2*), as well as 10 (31.25%) nonsense variants, all of which altered the *COL1A1* gene (**Figures 4B–D**).

Correlation Between Clinical Characteristics and Genotypes of Ukrainian OI Patients

The results of the analyses show a clear correlation between OI type and collagen defect. Haploinsufficiency collagen pathogenic variants correlate with milder OI types (*p* = 0.007). Interestingly, haploinsufficiency pathogenic variants in the *COL1A2* gene both caused severe OI, whereas the *COL1A1* variants caused moderate and mild OI (*p* = 0.003). However, there was no difference in OI types between structural *COL1A1* and *COL1A2* pathogenic variants (*p* = 0.895) (**Table 3**).

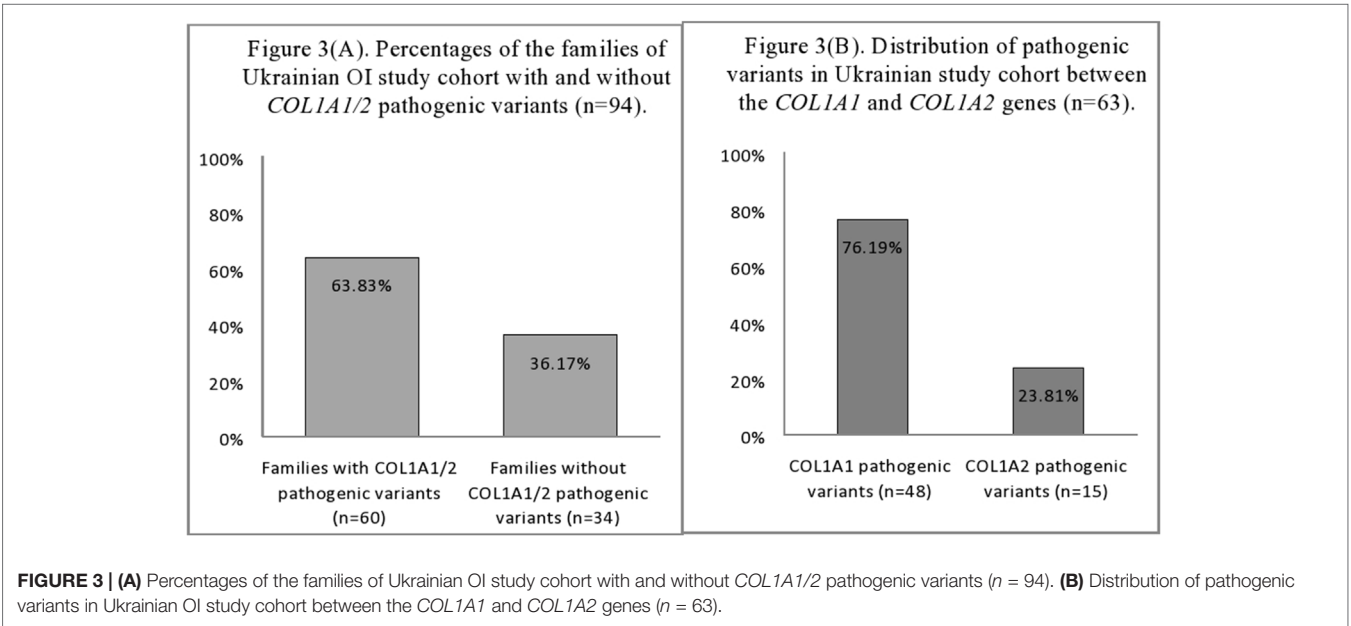
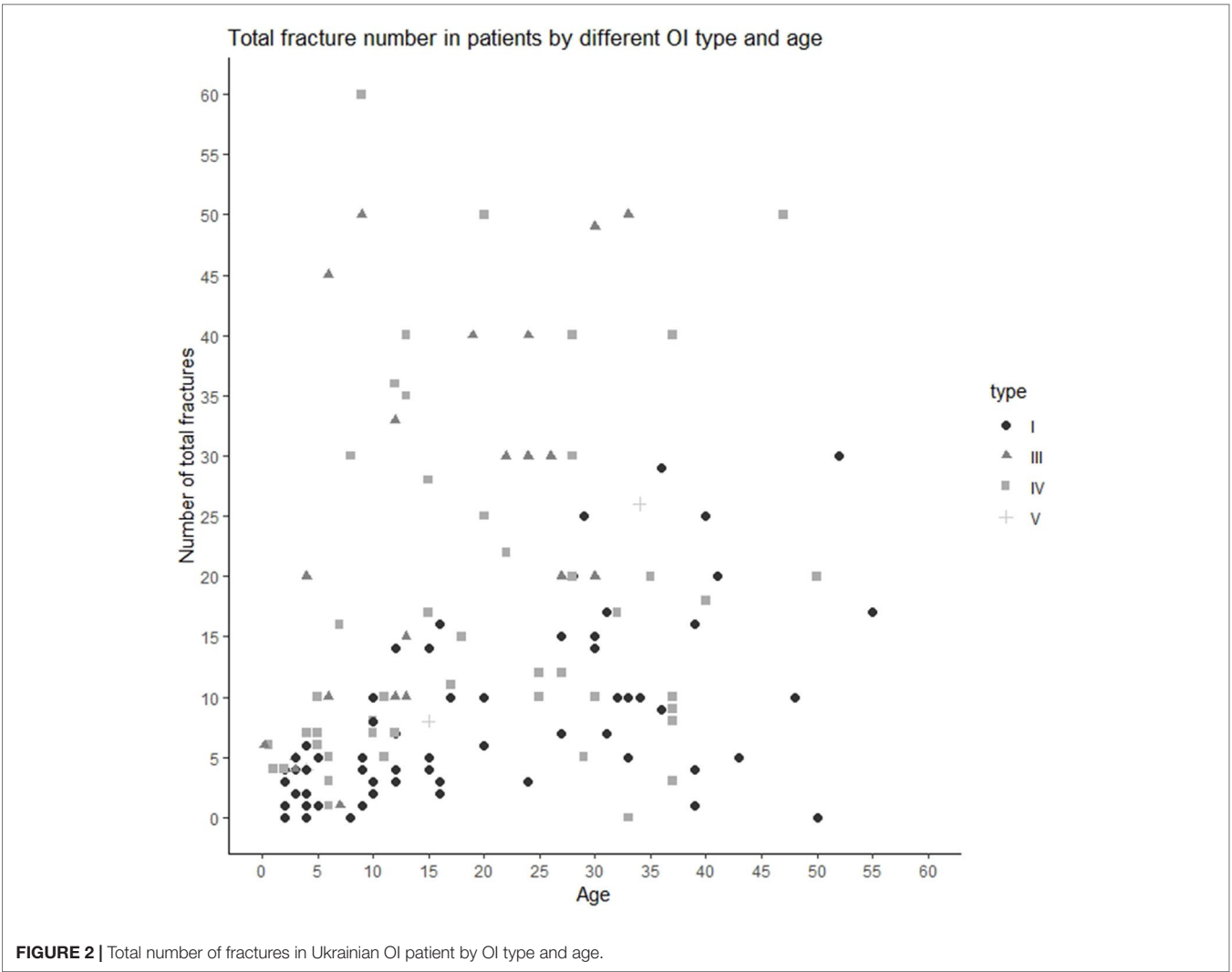


TABLE 2 | The COL1A1 (gDNA NG_007400.1, cDNA NM_000088.3) and COL1A2 (gDNA NG_007405.1, cDNA NM_000089.3) mutational spectrum among studied Ukrainian OI families.

#	Patient ID	Gene	Pathogenic variant	Exon	Pathogenic variant type	Protein alteration	Protein structural domain	OI type
1. COL1A1 pathogenic variants								
a. Haploinsufficiency COL1A1 pathogenic variants								
1	UA48†	COL1A1	c.1675_delG‡	25	Frameshift	p.Ala559Profs*21	Helical	I
2	UA91	COL1A1	c.1816_del G‡	26	Frameshift	p.Ala606Leufs*160	Helical	I
3	UA95†	COL1A1	c.3489_del G‡	48	Frameshift	p.Pro1165Leufs*74	Helical	IV
4	UA72†	COL1A1	c.3759_delG‡	50	Frameshift	p.Asp1254Thrfs*77	C-terminal propeptide	I
5	UA15†	COL1A1	c.459_delT	5	Frameshift	p.Gly154Alafs*111	N-terminal propeptide	I
6	UA14†	COL1A1	c.579_delT	7	Frameshift	p.Gly194Valfs*71	Chain	I
7	UA61	COL1A1	c.2393_dupC‡	35	Frameshift	p.Gly800Argfs*5	Helical	I
8	UA67	COL1A1	c.2523_delT	37	Frameshift	p.Gly842Alafs*266	Helical	I
9	UA31	COL1A1	c.2821_delG‡	40	Frameshift	p.Gly941Valfs*167	Helical	IV
10	UA44†	COL1A1	c.1789G > T	26	Nonsense	p.Glu597*	Helical	I
11	UA50	COL1A1	c.3076C > T	43	Nonsense	p.Arg1026*	Helical	I
12	UA19	COL1A1	c.658C > T	9	Nonsense	p.Arg220*	Helical	I
13	UA45†	COL1A1	c.1081C > T	17	Nonsense	p.Arg361*	Helical	IV
14	UA64	COL1A1	c.1243C > T	19	Nonsense	p.Arg415*	Helical	I
15	UA17†	COL1A1	c.1426G > T‡	21	Nonsense	p.Arg476*	Helical	I
16	UA93	COL1A1	c.2179C > T‡	32	Nonsense	p.Gln727*	Helical	IV
17	UA73†	COL1A1	c.3807G > A	49	Nonsense	p.Trp1269*	C-terminal propeptide	I
18	UA22†	COL1A1	c.495T > A‡	6	Nonsense	p.Tyr165*	Helical	I
19	UA25	COL1A1	c.904-9G > A	13i	Splice site	–	–	I
20	UA38	COL1A1	c.2613+6T > C	38i	Splice site	–	–	IV
21	UA40†	COL1A1	c.1614+1G > A	23i	Splice site	–	–	IV
22	UA41†	COL1A1	c.3815-1G > A‡	48i	Splice site	–	–	IV
23	UA43†	COL1A1	c.858+1G > A	12i	Splice site	–	–	IV
24	UA56†	COL1A1	c.858+1G > A	12i	Splice site	–	–	IV
25	UA80	COL1A1	c.804+1G > A	11i	Splice site	–	–	IV
b. Structural COL1A1 pathogenic variants								
26	UA101	COL1A1	c.3226G > A	45	Missense	p.Gly1076Ser	Helical	IV
27	UA23	COL1A1	c.769G > A	11	Missense	p.Gly257Arg	Helical	I
28	UA27†	COL1A1	c.2101G > T	31	Missense	p.Gly701Cys	Helical	III
29	UA62	COL1A1	c.1057G > T	17	Missense	p.Gly353Cys	Helical	IV
30	UA39	COL1A1	c.3652G > A	49	Missense	p.Ala1218Thr	Helical	IV
31	UA57†	COL1A1	c.757C > T	11	Nonsense	p.Arg253*	Helical	IV
32	UA21†	COL1A1	c.3655G > T‡	49	Missense	p.Asp1219Tyr	C-terminal propeptide	IV
33	UA71†	COL1A1	c.4356G > C‡	52	Missense	p.Gln1452His	C-terminal propeptide	III
34	UA89†	COL1A1	c.653G > A	9	Missense	p.Gly218Asp	Helical	I
35	UA53†	COL1A1	c.734G > A‡	10	Missense	p.Gly245Glu	Helical	I
36	UA33†	COL1A1	c.742G > A	10	Missense	p.Gly284Arg	Helical	III
37	UA30†	COL1A1	c.1319G > C‡	20	Missense	p.Gly440Ala	Helical	III
38	UA76†	COL1A1	c.1192G > A‡	18	Missense	p.Gly398Ser	Helical	III
39	UA94†	COL1A1	c.1588G > A	23	Missense	p.Gly530Ser	Helical	III
40	UA10	COL1A1	c.2362G > A	35	Missense	p.Gly788Ser	Helical	IV
41	UA78	COL1A1	c.2434G > A‡	36	Missense	p.Gly812Ser	Helical	IV
42	UA05†	COL1A1	c.2461G > A	37	Missense	p.Gly821Ser	Helical	III
43	UA96	COL1A1	c.2560G > A	38	Missense	p.Gly854Ser	Helical	I
44	UA32	COL1A1	c.1A > C‡	1	Missense	p.Met1Leu	Signal peptide	I
2. COL1A2 pathogenic variants								
a. Haploinsufficiency COL1A2 pathogenic variants								
45	UA82	COL1A2	2093_2110_dup‡	35	Frameshift	p.Leu699_Leu704dup	Helical	III
46	UA92†	COL1A2	c.2026-1_2042_dup‡	33i-34	Splice site	–	–	III
b. Structural COL1A2 pathogenic variants								
47	UA86†	COL1A2	c.2045G > T‡	34	Missense	p.Gly682Val	Helical	III
48	UA37	COL1A2	c.2233G > C	37	Missense	p.Gly745Arg	Helical	III
49	UA83†	COL1A2	c.2027G > A	34	Missense	p.Gly676Asp	Helical	IV
50	UA51†	COL1A2	c.3034G > A	46	Missense	p.Gly1012Ser	Helical	III
51	UA54†	COL1A2	c.982G > A	19	Missense	p.Gly328Ser	Helical	III
52	UA65	COL1A2	c.1009G > A	19	Missense	p.Gly337Ser	Helical	IV
53	UA102†	COL1A2	c.2224G > A‡	37	Missense	p.Gly742Arg	Helical	I
54	UA74	COL1A2	c.2288G > T	37	Missense	p.Gly763Val	Helical	III
55	UA47	COL1A2	c.2314G > A	38	Missense	p.Gly772Ser	Helical	IV

(Continued)

TABLE 2 | Continued

#	Patient ID	Gene	Pathogenic variant	Exon	Pathogenic variant type	Protein alteration	Protein structural domain	OI type
56	UA69†	COL1A2	c.2324G > A	38	Missense	p.Gly775Glu	Helical	IV
57	UA90†	COL1A2	c.1220T > C‡	22	Missense	p.Leu407Pro	Helical	I
3. Double pathogenic variants in the COL1A1 and COL1A2								
58	UA55†	COL1A1	c.2195_delA‡	32	Frameshift	p.Glu732Aspfs*34	Helical	IV
		COL1A1	c.2548_2549_dupC‡	36	Frameshift	p.Gly851Luefs*258	Helical	
59	UA85	COL1A1	c.370-1G > A	4i	Splice site	-	-	I
		COL1A2	c.2642A > C‡	41	Missense	p.Glu881Ala	Helical	
60	UA08†	COL1A1	c.334-1G > A‡	3i	Splice site	-	-	IV
		COL1A2	c.2642A > C‡	41	Missense	p.Glu881Ala	Helical	

Patients with sporadic pathogenic variants are marked with an obelisk (†). Novel pathogenic variants unreported in the collagen type I variant database (<http://www.le.ac.uk/ge/collagen/>) are marked with a diesis (§).

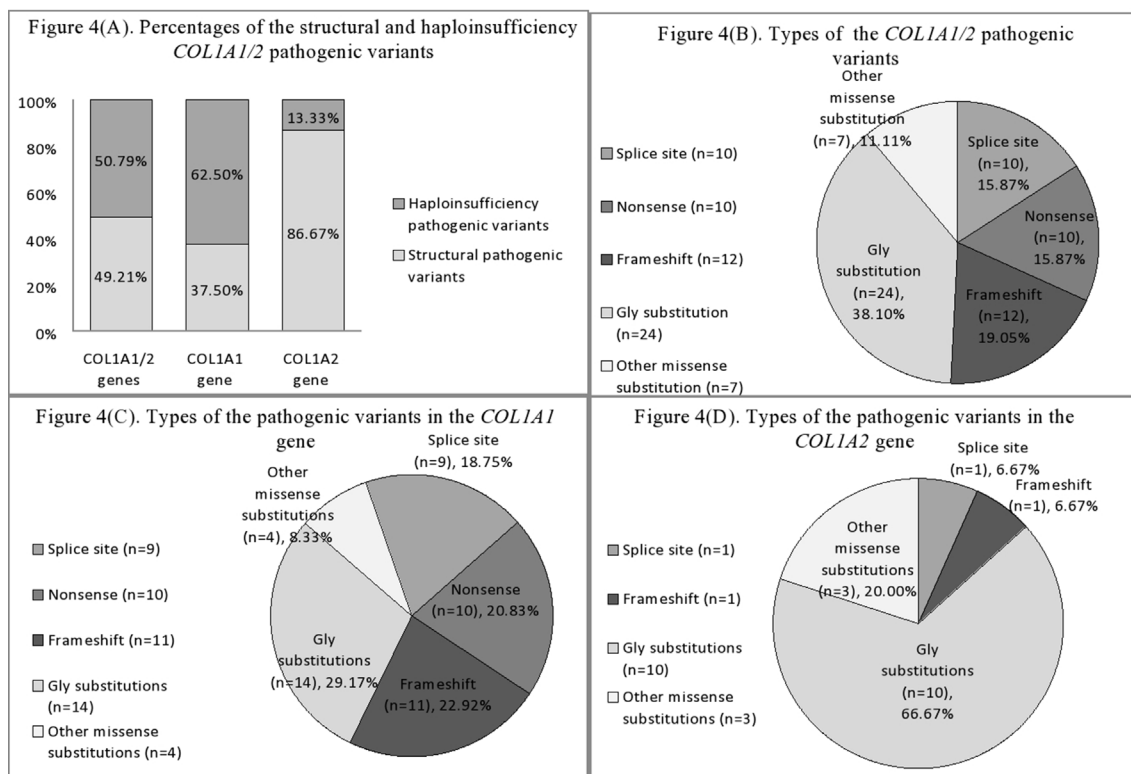


FIGURE 4 | (A) Percentages of the structural and haploinsufficiency COL1A1/2 pathogenic variants in Ukrainian OI study cohort. **(B)** Mutational spectrum of the COL1A1/2 genes in Ukrainian OI study cohort. **(C)** Types of pathogenic variants in the COL1A1 gene in Ukrainian OI study cohort. **(D)** Types of pathogenic variants in the COL1A2 gene in Ukrainian OI study cohort.

Type of pathogenic variant appeared to be crucially important for total fracture number, as well as for number of fractures per year. Patients with structural OI pathogenic variants had more fractures than did patients with haploinsufficiency defect. For those patients with structural defect (Gly substitutions), patients with COL1A1 pathogenic variants had more total fractures than did COL1A2 patients ($p = 0.0247$) (Table 3).

There were clear correlations between skeletal deformations and collagen defect. Patients with a structural collagen defect suffered from more severe skeletal deformations than did patients

with a haploinsufficiency OI defect ($p = 0.001$; $p = 5.80e-05$; $p = 2.37e-05$; $p = 2.0e-04$) (Table 3). Interestingly, patients with structural OI type I revealed more severe deformations in the lower limbs and spine, compared to those with haploinsufficiency OI type I ($p = 0.022$, $p = 0.029$). Gly substitutions in COL1A2 genes led to more severe cases of chest and spine deformations, compared to Gly substitutions in the COL1A1 gene ($p = 0.010$, $p = 0.004$). Patients with structural pathogenic variants were less mobile than patients with haploinsufficiency pathogenic variants ($p = 0.023$). Also those with structural COL1A2 pathogenic

TABLE 3 | Relationships between clinical characteristics and collagen I pathogenic variant types for cohort of Ukrainian OI patients.

	Structural OI type I/ haploin- sufficiency OI type I	All structural/ all haploin- sufficiency	Haploin- sufficiency COL1A1/ haploin- sufficiency COL1A2	Structural COL1A1/ structural COL1A2	Gly COL1A1/ Gly COL1A2	Gly > Ser COL1A1/ Gly > Ser COL1A2	Collagen OI/ non-collagen OI
All OI patients	7/16	29/31	27/2	18/11	20/10	7/4	60/34
Type I	–	7/16	15/0	5/2	8/1	1/0	23/13
Type III	–	12/2	0/2	7/5	6/5	3/2	14/9
Type IV	–	10/13	12/0	6/4	6/4	3/2	23/10
p-value	–	0.003*	0.004*	0.895	0.271	1	0.317
Blue sclera	5/15	24/30	26/2	15/9	16/8	4/3	54/23
p-value	0.481	0.329	1	1	1	1	0.011*
Mobility							
Wheelchair	–	8/2	1/1	2/6	1/6	1/1	10/5
Walking with support	1/1	7/5	5/0	6/1	7/1	3/1	12/4
Walking independently	5/15	12/23	21/0	9/3	11/2	2/2	35/22
Immobile	–	–	–	–	–	–	0/1
p-value	0.481	0.023*	0.169	0.038*	0.002*	1	0.465
Deformation severity							
(p-values)	0.065	0.001*	0.734	0.162	0.051	0.714	0.558
Upper limb	0.022*	5.803e-05*	0.048*	0.765	0.138	0.2	0.473
Lower limb	0.029*	2.37e-05*	0.285	0.224	0.010*	0.543	0.441
Spine	1	2.0e-04*	0.598	0.190	0.004*	0.2	0.125
Chest							
Total fracture number	6.00/4.00	28.00/6.00	6.00/13.00	10.00/30.00	7.00/30.00/	8.00/12.50	10/6.5
(median [range]/mean ± SD)	(4–14)/(0–25)	(4–300)/(0–40)	(0–40)/(6–20)	(4–300)/(7–50)	(2–49)/(7–50)	(4–49)/(7–30)	(0–300)/(0–50)
p-value	0.1184	0.001*	0.4899	0.3131	0.0247*	0.5918	0.046*
Fracture number per year	1.00/0.73	1.52/0.83	0.8/15.37	1.63/1.39	1.33/1.52	1.48/0.69	1.18/0.54
(median [range]/	(0.4–1.17)	(0.4–8.57)/	(0–2.69)/	(0.4–8.57)/	(0.25–7.5)/	(1–3.75)/	(0–30)/(0–5.56)
mean ± SD)	(0–2)	(0–30)	(0.74–30)	(0.7–3.08)	(0.77–3.08)	(0.78–1.75)	
p-value	0.483	0.001*	0.282	0.651	0.538	0.352	0.004*

Statistically significant p-values are marked with *. Data are n, unless otherwise indicated. SD is standard deviation.

variants were less mobile than patients with structural *COL1A1* pathogenic variants ($p = 0.038$).

There was no correlation between pathogenic variant type and presence of DI or hearing loss in the cohort of Ukrainian patients. However, the correlation of blue sclera with the presence of collagen I pathogenic variants ($p = 0.011$) was notable, as the majority of non-collagenous OI cases had white eye sclera.

DISCUSSION

Phenotype Characteristics of Ukrainian OI Cohort

In current study, the clinical and molecular characteristics of individuals from 94 osteogenesis imperfecta families from Ukraine were presented. Of the 143 subjects, 46.85% were diagnosed with OI type I. 16.78% of the patients had OI type III and 34.27% were described as having type IV. Three patients were classified as OI type V due to enlargement of the hyperplastic callus and were guided toward further analysis of the *IFITM5* gene, which will be presented elsewhere (Table 1, and Figure 1A). Previous studies showed that, amongst Polish OI patients ($n = 123$), patients with OI types I, III, and IV comprised 44%, 33%, and 21% of the cohort, respectively (Rusinska et al., 2017). In a Russian pediatric study (including

data from 31 regions of the Russian Federation), patients ($n = 117$) were classified as follows: OI type I, 59%; type III, 27%; and type IV, 14% (Yakhyayeva et al., 2016).

Interestingly, the current study revealed a low percentage of patients with OI type III. However, this may be connected with the difficulty involved for patients with severe forms of OI forms to travel to the study venue. Nevertheless, OI patients with milder OI types are usually rarer in population cross-sectional studies compared with nationwide register-based studies, due to poor diagnostics and less patient interest. The current study illustrates the effective work of the Ukrainian Association of Crystal People, which enabled the involvement in research of those individuals with milder OI. In studies of other populations, the distribution by types (I/III/IV) were as follows: Vietnamese 31.5%/31.5%/37%, Taiwanese 58%/7%/35%, Chinese 28%/38%/34%, Israeli 61%/21%/14%, Swedish 68%/13%/19%, Norwegian 77%/9%/11%, and Finnish 72%/4%/20% (Hartikka et al., 2004; Lin et al., 2015; Lindahl et al., 2015; Binh et al., 2017). There may be various reasons for the observed contrasts in the proportion of OI types across different populations, such as sample sizes, diagnostics, and classification. Additionally, methods of patient recruitment might affect distribution of OI types in our cohort, as patients were enrolled *via* OI patients' organization and thus OI individuals with more severe OI forms might be overrepresented in contrast to individuals with mild

OI forms. Being a spectrum of disorders, OI type classification remains subjective, as some of the individuals might develop border phenotypes and thus might be classified by different health professionals differently. Current circumstances might add bias and contrast to results of different studies. However, the influence of genetic factors cannot be excluded.

In the current study, the number of fractures per year was strongly aligned with those of a previous Swedish OI study (type I 0.57 ± 0.68 ; type III 3.83 ± 9.32 ; and type IV 1.33 ± 1.38) (Lindahl et al., 2015). In contrast to Swedish register-based study, clinical features in our cohort were partially self-reported and it might bias the data. However, similar to previous studies, more deformed and fractured body compartments occurred in the lower limbs, which may be due to the greater loading of the lower limbs (Binh et al., 2017). We discovered two adult patients without history of fractures, but with other OI symptoms and OI history in the family. Both cases lacked pathogenic variants in *COL1A1/2* genes. Therefore, it would be particularly interesting to identify causative pathogenic variants in these cases and further investigate reasons of fracture absence.

The proportion of patients suffering from DI (55%) was higher than in Swedish (25%) and Brazilian (27%) OI cohorts, but similar to the proportions in Taiwanese (44%) and Vietnamese (61%) OI populations (Lin et al., 2015; Lindahl et al., 2015; Binh et al., 2017; Brizola et al., 2017). The percentage of patients with blue sclera (87%) aligns with the proportions observed in previous studies (e.g. 82%, 93%, 90%, and 80% of patients in Swedish, Brazilian, Taiwanese, and Vietnamese cohorts, respectively) (Lin et al., 2015; Lindahl et al., 2015; Binh et al., 2017; Brizola et al., 2017).

Genotype Characteristics of Ukrainian OI Cohort

According to previous studies, collagen I pathogenic variants account for between 60 and 90% of OI cases (Dagleish, 1998). In a recent major genetic study of 598 OI individuals from 487 families of Caucasian, Hispanic, Arab, and Asian origin, the proportion of collagen pathogenic variants was 86% (Bardai et al., 2016). However, in population-based OI studies, *COL1A1/2* percentages vary. In this current study, the proportion of *COL1A1/2* pathogenic variants in Ukrainian OI patients comprised 63.83%, which appears to be lower than in Northern Europe [e.g. in Estonian and Swedish (87%) and Finnish (91%) populations] (Hartikka et al., 2004; Lindahl et al., 2015; Zhytnik et al., 2017). At the same time, these results show that Ukrainian OI patients harbor a higher number of *COL1A1/2* pathogenic variants than do patients from Russia and Asian populations. Amongst patients from Russia (Yakutia and Bashkortostan regions), the percentage of collagen I pathogenic variants was 41% (Khusainova et al., 2012). Asian populations from Vietnam, Taiwan, and Korea were characterized by *COL1A1/2* pathogenic variants of 59%, 51%, and 52%, respectively (Lee et al., 2006; Lin et al., 2015; Ho Duy et al., 2016).

We have previously analyzed Estonian OI patients, using same analysis methods and laboratory techniques (Zhytnik et al., 2017). Regarding the fact, that proportion of the *COL1A1/2* mutations in the Estonian OI cohort composed ~90%, we suppose that

analysis methods and techniques could not influence proportion of the collagen I mutations in the Ukrainian OI cohort.

Panel sequencing of the autosomal recessive OI genes in Ukrainian OI patients without collagen I mutations is ongoing and will be reported elsewhere. Patients with contractures will be scanned for presence of variants in the *PLOD2* and *FKBP10* genes to specify presence of a Bruck syndrome diagnosis.

The current analysis also illustrated that the proportion of pathogenic variants of the *COL1A2* gene was less than that of the *COL1A1* (23.81% and 76.19%, respectively). Similar results were observed in Estonian, Swedish, Finnish, and Taiwanese populations (Hartikka et al., 2004; Lin et al., 2015; Lindahl et al., 2015; Zhytnik et al., 2017). However, in a Russian study of 83 OI patients of Turkic and Slavic origin, and in a study of 11 Egyptian patients, pathogenic *COL1A2* variants were not observed (Khusainova et al., 2012; Aglan et al., 2015).

Another interesting result of the current study was that the amounts of structural and haploinsufficiency variants were almost equal (49.21% and 50.79%, respectively). This was similar to results reported for a Swedish OI cohort (Lindahl et al., 2015). In contrast, in Estonian and Finnish cohorts, the proportions of structural OI pathogenic variants were significantly lower (31% and 30%, respectively), whereas in Taiwanese and Vietnamese cohorts, the proportion was higher (78%) (Lin et al., 2015; Ho Duy et al., 2016; Zhytnik et al., 2017).

The reasons for differences in genotypes may be similar to those for differences in OI type distributions: i.e. sample sizes, methods of patient recruitment, and potential variations in OI genetic epidemiology between different populations.

As expected, the majority of haploinsufficiency variants caused OI types I and IV. Only two individuals with haploinsufficiency variants had OI type III (Table 2). Investigation of genotypes in patients with OI type III and *COL1A2* haploinsufficiency variants for presence of mutations in other OI genes is essential and current patients are included into panel sequencing cohort. The majority of patients with structural pathogenic variants were represented by individuals with OI types IV and III. However, seven individuals had OI type I, of which five harbored glycine substitutions. Patients with double pathogenic variants had moderate and mild phenotypes IV and I. Patient UA08 had blue sclera, hearing loss (started at the age of 10). Patients UA55 and UA85 had blue sclera, DI, but no hearing loss. It is known that phenotype severity depends not only on the type of the pathogenic variant, but also on helical location and substituted residue, which contribute to the development of the phenotype (Marini et al., 2007; Rauch et al., 2010). It could be proposed, that patient UA55 who had two frameshift variants in the *COL1A1* gene might have had a haploinsufficiency of the collagen $\alpha 1$ chain and suffer from quantitative collagen defect.

Unexpectedly mild phenotypes were harbored by patients UA89, UA53, UA96, UA23, UA102. All of these patients suffered OI I, however had Gly structural substitutions, which generally are associated with severe OI. Interestingly, *COL1A1* c.769G > A (p.[Gly257Arg] variant (UA23) was reported for 37 times, causing all range of classical non-lethal OI types I, III, IV. Number of observations might have underlined phenotype diversity evidence, stating broader spectrum of affection compared to

variants with lack of data. Another *COL1A1* variant c.653G > A (p.[Gly218Asp]) (UA89) was also described in a patient with OI IV. Whereas variant c.2560G > A was previously found in a patient with OI I/IV (Marini et al., 2007). Both patients had apparently mild types and differences in phenotype presentation might be connected to classification bias of border OI forms (i.e. type I/IV). Moreover, phenotypes of OI patients might be affected by the treatment, and real effect of the mutation might remain unresolved.

The above discussion indicates that, although the general effect of collagen I pathogenic variants on phenotype is understood, current knowledge does not reflect all of the nuances of genotype-phenotype correlations. OI is known to show incomplete penetrance, thus the clinical presentation can vary to a great degree even between same family members (Van Dijk and Sillence, 2014). Recent studies show over hundred loci in the human genome influencing bone mineral density (Estrada et al., 2012; Rivadeneira and Mäkitie, 2016). In addition bone morphology and quality, bone strength, and toughness, also influence fracture risk (Chesnut and Rosen, 2001; Rivadeneira and Mäkitie, 2016). Bone is influenced by metabolic (glucose, lipids, calcium, hormones), biomechanical (muscle strain, body weight, and composition), material (collagen, mineralization), cellular (osteoclast, osteoblast, osteocyte activity, and differentiating), growth, and remodeling factors (Rivadeneira and Mäkitie, 2016). All these factors can potentially influence the phenotypic expression of collagen gene mutations by yet undescribed molecular pathways. Further investigations of the clinical manifestations and molecular characteristics of patients may contribute to the understanding of OI phenotypes' diversity.

In current study we present 27 novel OI pathogenic variants, harbored by Ukrainian OI patients. Some novel pathogenic variants alter previously-reported positions with new substitutions; this could be of particular interest to the exploration of the OI phenotypical spectrum and its interconnections with genotypes. Moreover, current pathogenic variants enrich database of OI pathogenic variants and have practical use for OI genetic diagnostics. The percentage of novel pathogenic variants in Ukrainian OI cohort was 42.85%. Despite the numerous reported variants, there are still many novel collagen type I pathogenic variants, each of which improves the understanding of OI genotype-phenotype correlations.

As it is known, splice site mutations might result in exon skipping, intronic inclusion, or activation of cryptic splice sites (Marini et al., 2007). Changes in mRNA and protein depend on whether these alterations are in frame or produce translational frameshifts. Pathogenic mechanism of the identified novel OI haploinsufficiency variants is yet to be investigated with further functional studies. In general, haploinsufficiency mutations result in a severely truncated mRNA molecules, which cause activation of pretermination stop codon, followed by mediated mRNA decay (Chang et al., 2007; Fang et al., 2013; Symoens et al., 2014).

Structural pathogenic variants cause synthesis of an abnormal protein. Defective protein is secreted into the extracellular matrix and interferes with fibrillogenesis, collagen-matrix, bone cells, and hydroxyapatite. The structural pathogenic variants

affect extracellular matrix more severely than haploinsufficiency variants (Bodian et al., 2008; Marini et al., 2017). Although exact mechanism of the structural pathogenic mutations has to be further elucidated with functional studies, according to *in silico* analysis and description of functional domains in the collagen type I protein, following consequences of the novel structural variants can be predicted: *COL1A1*, c.3655G > T, p.(Asp1219Tyr), UA21, OI IV

Substitution in identical position was described previously by Lindahl et al. c.3655G > A (p.[Asp1219Asn]) (Lindahl et al., 2011). However, in contrast to Ukrainian patient, Swedish patient had mild OI phenotype. It can be partly explained by the hydrophobic side chain of the Tyr, which compared to uncharged side chain of Asn, has larger effect on procollagen properties.

Both variants alter C-cleavage site, what causes severe processing defect, alteration of the C-propeptide cleavage, and increased mineralization (Lindahl et al., 2011; Marini et al., 2017). However, we do not have bone mineral density data of the UA21 patient to confirm high bone mass phenotype.

COL1A1, c.4356G > C, p.(Gln1452His), UA71, OI III

Current structural variant also alters C-cleavage site of the procollagen, what is critical for mineralization. In addition, according to *in silico* analysis, current variant causes gain of donor splice site and might result in a cryptic splice site. Current mutation might be particularly interesting for further functional investigation, as patient suffered more than 300 fractures during lifetime and developed hearing loss at the age of 2.

COL1A1, c.734G > A, p.(Gly245Glu), UA53, OI I

According to collagen type I functional domains, current variant might affect keratan sulfate proteoglycans binding region, $\alpha 2\beta 1$ integrin, and interleukin 2 (IL2) binding sites (Sweeney et al., 2008). In this way, variant might alter protein function.

COL1A1, c.1319G > C, p.(Gly440Ala), UA30, OI III

Variant affects keratan sulfate proteoglycans binding domain and probably creates a new donor splice site (Sweeney et al., 2008).

COL1A1, c.1192G > A, p.(Gly398Ser), UA76, OI III

Variant is located in the proposed site of discoidin domain receptor 2 (DDR2) binding site, von Willebrand Factor binding site, and dermatan/chondroitin sulfate proteoglycans/decorin binding region (Sweeney et al., 2008). Marini et al. has described patient with OI III/IV harboring variant c.1192G > T (p.[Gly398Cys]) (Marini et al., 2007). More severe phenotypes of UA76 patient might be caused by more distant properties of Ser compared to Cys in relation to Gly and have greater effect on protein function.

COL1A1, c.2434G > A, p.(Gly812Ser), UA78, OI IV

Current variant is located in the major ligand binding region (MLBR) 2, proposed region for lethal OI mutations (Marini et al., 2007). Variant overlaps with $\alpha 2\beta 1$ integrin binding site and glycation region (Sweeney et al., 2008).

COL1A1, c.1A > C, p.(Met1Leu), UA32, OI I

The variant alters initiating methionine, causing activation of potential downstream translation initiation site with new reading frame. Substitutions of the Met in the *COL1A1* were previously described and resulted in OII.

COL1A2, c.2045G > T, p.(Gly682Val), UA86, OI III

Variant is located in the MLBR2 and interrupts cartilage oligomeric matrix protein and phosphoryn, secreted protein, acidic, and rich in cysteine (SPARC) binding sites, keratan sulfate proteoglycans binding region (Sweeney et al., 2008).

COL1A2, c.2224G > A, p.(Gly742Arg), UA102, OI I

Variant is located in the MLBR2, overlaps with SPARC binding site, cell interaction domain, cartilage oligomeric matrix protein, phosphoryn binding sites (Sweeney et al., 2008).

COL1A2, c.1220T > C, p.(Leu407Pro), UA90, OI I

Variant is located in the dermatan/chondroitin sulfate proteoglycans/decorin binding region.

COL1A2, c.2642A > C, p.(Glu881Ala), UA85, OI IV, and UA08, OI I

Variant is located at the delineate clusters of lethal OI mutations on the $\alpha 2(I)$ chain. Current variant might alter proposed site of dermatan/chondroitin sulfate proteoglycans/decorin, IL2, amyloid precursor protein binding regions (Sweeney et al., 2008). Both variants are harbored by individuals with additional *COL1A1* variants.

Genotypes were assessed with Sanger sequencing, which is considered to be a powerful and accurate method of mutational analysis. Special design of primers allowed the capture of all inter-exon junction regions and 5'UTR, 3'UTR regions. However, potential limitations include the inability of Sanger sequencing to detect whole gene or exon deletions and duplications. The number of *COL1A1/2* pathogenic variants may therefore be underestimated. Differences in sample sizes may also contribute to the variation of results between studies.

Genotype-Phenotype Analysis

OI genotype-phenotype correlations have been a subject of interest for many decades. Although the association of pathogenic variant with OI phenotype remains elusive, previous studies showed *COL1A1* pathogenic variants to be associated with more severe phenotypes than were *COL1A2* (Marini et al., 2007; Mrosk et al., 2018). In the current study, a higher fracture number was observed in structural *COL1A1* cases; among *COL1A2* cases there were more cases with severe deformations of the spine and chest.

These results show the presence of a strong correlation between collagen type I defect and severity of OI, which aligns well with previous studies. Haploinsufficiency collagen I defects cause milder, less fragile, and less deformed OI types (Marini et al., 2007; Amor et al., 2011; Lindahl et al., 2015). Although in a Swedish OI cohort, the difference between type I patients with structural and haploinsufficiency collagen defects were insignificant, in the current study, structural type I patients had more deformed lower limbs and more fractures. This underlies the vital importance of a normal collagen helical structure (rather than amount).

In contrast to previous studies, the current study did not find an association between DI and pathogenic variant type. This apparent lack of correlation may be attributed to the presence of the additional non-OI etiology of dental issues in the Ukrainian cohort (Lin et al., 2009; Amor et al., 2011; Lindahl et al., 2015).

Although there are numerous different genes associated with OI, the differences in phenotypes between collagen and

non-collagen OI cases were insignificant, except that of blue sclera (which was also associated with collagen I pathogenic variants in previous studies) (Lindahl et al., 2015). The overlap of OI phenotypes with different genotypes remains a focus for future research.

The current study's results corroborate the absence of a correlation between collagen pathogenic variants and hearing loss. No previous studies have detected factors, which could explain hearing loss in OI families, and this remains a vital topic for future research. Number of pediatric patients with hearing loss in Ukrainian OI cohort was high, so it would be especially interesting to identify non-collagen OI genes in remained patients (Hartikka et al., 2004; Amor et al., 2011).

CONCLUSIONS

The current paper presents the phenotype and genotype characteristics of 94 Ukrainian OI families for the first time. These patients exhibited OI phenotypes I (46.85%), III (16.78%), IV (34.27%), and V (2.10%). *COL1A1/2* pathogenic variants were identified in 63.83% of the 94 screened unrelated patients, which were equally divided between structural (49.21%) and haploinsufficiency (50.79%) pathogenic variants. 27 novel OI causing pathogenic variants in the *COL1A1/2* genes were presented in this study. Genotype-phenotype analysis supported previous findings on the dependence of OI phenotype severity on type of defect. Future research will focus on the performance of a whole exome sequencing analysis of patients negative for *COL1A1/2* pathogenic variants, in order to identify OI genetic causes.

There appears to be very little data available on OI patients from East Slavic populations. The results of this current research will enrich the OI variant database and contribute to the understanding of genotype-phenotype correlations in osteogenesis imperfecta. The results of this research may also be used to promote further research, treatment, and diagnostics of OI in Ukraine, which will result in the improvement of patients' quality of life and accessibility to treatment.

ETHICS STATEMENT

The current study was conducted in accordance with the Helsinki Declaration and received approval from the Sytenko Institute of Spine and Joint Pathology of the Ukrainian Academy of Medical Sciences and the Ethical Review Committee on Human Research of the University of Tartu (Permit no. 221/M-34). Informed written consent from the patients or their legal representatives was obtained prior to inclusion to the study.

AUTHOR CONTRIBUTIONS

LZ conceived the study, carried out the genetic studies, interacted with the patients, performed the data analysis, participated in the design of the study, and drafted the manuscript. KM participated in the design of the study, interacted with the patients, coordinated the blood sample collection, interacted with the patients, performed

the data analysis, participated in the design of the study, performed analysis and helped to draft the manuscript. AP and SKh interacted with the patients and participated in the designing of the study and sample collection. EP, SKö and ER carried out the genetic studies, performed the data analysis, and helped to draft the manuscript. AM participated in the designing of the study, coordinated the data interpretation and statistical analysis, and helped to draft the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Estonian Science Agency project IUT20-46 (TARBS14046I), the European Regional Development Fund and the Archimedes Foundation support for the Centre of Excellence on Translational Medicine, the University of Tartu's

Development Fund, University of Tartu's Baseline Funding, and the HypOrth Project funded by the European Union's 7th Framework Programme grant agreement no. 602398.

ACKNOWLEDGMENTS

We would like to thank all patients and their relatives who participated in the study. We would also like to show our appreciation to the following people and organizations for their help and support with data collection: Lyuba Petrova and The Ukrainian Association of Crystal People; Anneli Truupõld and workers of the Department of Traumatology and Orthopedics and Department of Pathophysiology, University of Tartu; and Ardo Birk and Madis Karu for the development of the online OI database of the Clinic of Traumatology and Orthopedics, TU Hospital.

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249. doi: 10.1038/nmeth0410-248
- Aglan, M. S., Abdel-Hamid, M. S., Otaify, G. A., Ismail, S. M., Effat, L. K., and Temtam, S. A. (2015). Mutational spectrum of COL1A1 and COL1A2 in Egyptian patients with autosomal dominant osteogenesis imperfecta with clinical severity score and genotype/phenotype correlation. *Middle East J. Med. Genet.* 4, 7–12. doi: 10.1097/01.MXE.0000457061.88602.72
- Amor, I. M., Rauch, F., Gruenwald, K., Weis, M., Eyre, D. R., Roughley, P., et al. (2011). Severe osteogenesis imperfecta caused by a small in-frame deletion in CRTAP. *Am. J. Med. Genet. A* 155A, 2865–2870. doi: 10.1002/ajmg.a.34269
- Bardai, G., Moffatt, P., Glorieux, F. H., and Rauch, F. (2016). DNA sequence analysis in 598 individuals with a clinical diagnosis of osteogenesis imperfecta: diagnostic yield and mutation spectrum. *Osteoporos. Int.* 27, 3607–3613. doi: 10.1007/s00198-016-3709-1
- Ben Amor, I. M., Glorieux, F. H., and Rauch, F. (2011). Genotype-phenotype correlations in autosomal dominant osteogenesis imperfecta. *J. Osteoporos.* 2011, 540178. doi: 10.4061/2011/540178
- Binh, H. D., Maasalu, K., Dung, V. C., Ngoc, C. T. B., Hung, T. T., Nam, T. V., et al. (2017). The clinical features of osteogenesis imperfecta in Vietnam. *Int. Orthop.* 41, 21–29. doi: 10.1007/s00264-016-3315-z
- Bodian, D. L., Madhan, B., Brodsky, B., and Klein, T. E. (2008). Predicting the clinical lethality of osteogenesis imperfecta from collagen glycine mutations †. *Biochemistry* 47, 5424–5432. doi: 10.1021/bi800026k
- Brizola, E., Zambrano, M. B., Pinheiro, B. S., Vanz, A. P., and Félix, T. M. (2017). Clinical features and pattern of fractures at the time of diagnosis of osteogenesis imperfecta in children. *Rev. Paul. Pediatr.* 35, 171–177. doi: 10.1590/1984-0462/2017/35;2;00001
- Byers, P. H., and Steiner, R. D. (1992). Osteogenesis imperfecta. *Annu. Rev. Med.* 43, 269–282. doi: 10.1146/annurev.me.43.020192.001413
- Chang, Y.-F., Imam, J. S., and Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.* 76, 51–74. doi: 10.1146/annurev.biochem.76.050106.093909
- Chen, R., Mias, G. I., Li-Pook-Than, J., Jiang, L., Lam, H. Y. K., Chen, R., et al. (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 148, 1293–1307. doi: 10.1016/j.cell.2012.02.009
- Chesnut, C. H., and Rosen, C. J. (2001). Reconsidering the effects of antiresorptive therapies in reducing osteoporotic fracture. *J. Bone Miner. Res.* 16, 2163–2172. doi: 10.1359/jbmr.2001.16.12.2163
- Cho, T.-J., Lee, K.-E., Lee, S.-K., Song, S. J., Kim, K. J., Jeon, D., et al. (2012). A single recurrent mutation in the 5'-UTR of IFITM5 causes osteogenesis imperfecta type V. *Am. J. Hum. Genet.* 91, 343–348. doi: 10.1016/j.ajhg.2012.06.005
- Consortium T 1000 GP. (2015) A global reference for human genetic variation. *Nature* 526, 68–74.
- Dalgleish, R. (1997). The human type I collagen mutation database. *Nucleic Acids Res.* 25, 181–187. doi: 10.1093/nar/25.1.181
- Dalgleish, R. (1998). The human collagen mutation database 1998. *Nucleic Acids Res.* 26, 253–255. doi: 10.1093/nar/26.1.253
- Estrada, K., Styrkarsdottir, U., Evangelou, E., Hsu, Y.-H., Duncan, E. L., Ntzani, E. E., et al. (2012). Genome-wide meta-analysis identifies 56 bone mineral density loci and reveals 14 loci associated with risk of fracture. *Nat. Genet.* 44, 491–501. doi: 10.1038/ng.2249
- Fang, Y., Bateman, J. F., Mercer, J. F., and Lalande, S. R. (2013). Nonsense-mediated mRNA decay of collagen -emerging complexity in RNA surveillance mechanisms. *J. Cell Sci.* 126, 2551–2560. doi: 10.1242/jcs.120220
- Glorieux, F. H., Rauch, F., Plotkin, H., Ward, L., Travers, R., Roughley, P., et al. (2000). Type V osteogenesis imperfecta: a new form of brittle bone disease. *J. Bone Miner. Res.* 15, 1650–1658. doi: 10.1359/jbmr.2000.15.9.1650
- Hartikka, H., Kuurila, K., Kärkkö, J., Kaitila, I., Grénman, R., Pynnönen, S., et al. (2004). Lack of correlation between the type of COL1A1 or COL1A2 mutation and hearing loss in osteogenesis imperfecta patients. *Hum. Mutat.* 24, 147–154. doi: 10.1002/humu.20071
- Ho Duy, B., Zhytnik, L., Maasalu, K., Kändla, I., Prans, E., Reimann, E., et al. (2016). Mutation analysis of the COL1A1 and COL1A2 genes in Vietnamese patients with osteogenesis imperfecta. *Hum. Genomics* 10, 27. doi: 10.1186/s40246-016-0083-1
- Khusainova, R., Nadyrshina, D., Gilyazova, I., Karunas, A., Akhmetova, V., Khidiyatova, I., et al. (2012). Analysis of COL1A1 and COL1A2 genes in osteogenesis imperfecta patients from Russia. *Bone* 50, S109. doi: 10.1016/j.bone.2012.02.333
- Kocher, M. S., and Shapiro, F. (1998). Osteogenesis imperfecta. *J. Am. Acad. Orthop. Surg.* 6, 225–236. doi: 10.5435/00124635-199807000-00004
- Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat. Protoc.* 4, 1073–1081. doi: 10.1038/nprot.2009.86
- Lee, K.-S., Song, H.-R., Cho, T.-J., Kim, H. J., Lee, T.-M., Jin, H.-S., et al. (2006). Mutational spectrum of type I collagen genes in Korean patients with osteogenesis imperfecta. *Hum. Mutat.* 27, 599–599. doi: 10.1002/humu.9423
- Lin, H.-Y., Chuang, C.-K., Su, Y.-N., Chen, M.-R., Chiu, H.-C., Niu, D.-M., et al. (2015). Genotype and phenotype analysis of Taiwanese patients with osteogenesis imperfecta. *Orphanet J. Rare Dis.* 10, 152. doi: 10.1186/s13023-015-0370-2
- Lin, H.-Y., Lin, S.-P., Chuang, C.-K., Chen, M.-R., Chang, C.-Y., and Niu, D.-M. (2009). Clinical features of osteogenesis imperfecta in Taiwan. *J. Formos. Med. Assoc.* 108, 570–576. doi: 10.1016/S0929-6646(09)60375-2
- Lindahl, K., Åström, E., Rubin, C.-J., Grigeliuniene, G., Malmgren, B., Ljunggren, Ö., et al. (2015). Genetic epidemiology, prevalence, and genotype-phenotype correlations in the Swedish population with osteogenesis imperfecta. *Eur. J. Hum. Genet.* 23, 1042–1050. doi: 10.1038/ejhg.2015.81
- Lindahl, K., Barnes, A. M., Fratzl-Zelman, N., Whyte, M. P., Hefferan, T. E., Makareeva, E., et al. (2011). COL1 C-propeptide cleavage site mutations

- cause high bone mass osteogenesis imperfecta. *Hum. Mutat.* 32, 598–609. doi: 10.1002/humu.21475
- Marini, J. C., Forlino, A., Bächinger, H. P., Bishop, N. J., Byers, P. H., Paepe, A., et al. (2017). Osteogenesis imperfecta. *Nat. Rev. Dis. Prim.* 3, 17052. doi: 10.1038/nrdp.2017.52
- Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., et al. (2007). Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.* 28, 209–221. doi: 10.1002/humu.20429
- Mrosk, J., Gandham, S. B., Shah, H., Hecht, J., Krüger, U., Shukla, A., et al. (2018). Diagnostic strategies and genotype-phenotype correlation in a large Indian cohort of osteogenesis imperfecta. *Bone* 110, 368–377. doi: 10.1016/j.bone.2018.02.029
- Rauch, F., Lalic, L., Roughley, P., and Glorieux, F. H. (2010). Genotype-phenotype correlations in nonlethal osteogenesis imperfecta caused by mutations in the helical domain of collagen type I. *Eur. J. Hum. Genet.* 18, 642–647. doi: 10.1038/ejhg.2009.242
- Rivadeneira, F., and Mäkitie, O. (2016). Osteoporosis and Bone Mass Disorders: From Gene Pathways to Treatments. *Trends Endocrinol. Metab.* 27, 262–281. doi: 10.1016/j.tem.2016.03.006
- Roughley, P. J., Rauch, F., and Glorieux, F. H. (2003). Osteogenesis imperfecta—clinical and molecular diversity. *Eur. Cell. Mater.* 5, 41–47. doi: 10.22203/eCM.v005a04
- Rusinska, A., Michalus, I., Jakubowska-Pietkiewicz, E., Beska, K., Adamiecka, P., and Chlebna-Sokol, D. (2017). Difficulties in diagnostics and clinical classification of osteogenesis imperfecta in Poland. *Bone* 6, P049. doi: 10.1530/boneabs.6.P049
- Schwarz, J. M., Rödelberger, C., Schuelke, M., and Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* 7, 575–576. doi: 10.1038/nmeth0810-575
- Semler, O., Garbes, L., Keupp, K., Swan, D., Zimmermann, K., Becker, J., et al. (2012). A mutation in the 5'-UTR of IFITM5 creates an in-frame start codon and causes autosomal-dominant osteogenesis imperfecta type V with hyperplastic callus. *Am. J. Hum. Genet.* 91, 349–357. doi: 10.1016/j.ajhg.2012.06.011
- Shapiro, J. R. (2014). "Clinical and genetic classification of osteogenesis imperfecta and epidemiology," in *Osteogenesis Imperfecta* (Amsterdam, The Netherlands: Elsevier), 15–22. doi: 10.1016/B978-0-12-397165-4.00002-2
- Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29(1), 308–311.
- Sillence, D. O., Senn, A., and Danks, D. M. (1979). Genetic heterogeneity in osteogenesis imperfecta. *J. Med. Genet.* 16, 101–116. doi: 10.1136/jmg.16.2.101
- Sweeney, S. M., Orgel, J. P., Fertala, A., McAuliffe, J. D., Turner, K. R., Di Lullo, G. A., et al. (2008). Candidate cell and matrix interaction domains on the collagen fibril, the predominant protein of vertebrates. *J. Biol. Chem.* 283, 21187–21197. doi: 10.1074/jbc.M709319200
- Symoens, S., Hulmes, D. J. S., Bourhis, J.-M., Coucke, P. J., De Paepe, A., and Malfait, F. (2014). Type I procollagen C-propeptide defects: study of genotype-phenotype correlation and predictive role of crystal structure. *Hum. Mutat.* 35, 1330–1341. doi: 10.1002/humu.22677
- Van Dijk, F. S., and Sillence, D. O. (2014). Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. *Am. J. Med. Genet. A* 164A, 1470–1481. doi: 10.1002/ajmg.a.36545
- Warman, M. L., Cormier-Daire, V., Hall, C., Krakow, D., Lachman, R., LeMerrer, M., et al. (2011). Nosology and classification of genetic skeletal disorders: 2010 revision. *Am. J. Med. Genet. Part A* 155, 943–968. doi: 10.1002/ajmg.a.33909
- Yakhyayeva, G. T., Namazova-Baranova, L. S., Margiyeva, T. V., and Chumakova, O. V. (2016). Osteogenesis imperfecta in children in the Russian Federation: results of the federal registry audit. *Pediatr. Pharmacol.* 13, 44–48. doi: 10.15690/pf.v13i1.1514
- Zhytnik, L., Maasalu, K., Reimann, E., Prans, E., Köks, S., and Märtson, A. (2017). Mutational analysis of COL1A1 and COL1A2 genes among Estonian osteogenesis imperfecta patients. *Hum. Genomics* 11, 19. doi: 10.1186/s40246-017-0115-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zhytnik, Maasalu, Pashenko, Khmyzov, Reimann, Prans, Köks and Märtson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Novel Neonatal Variants of the Carbamoyl Phosphate Synthetase 1 Deficiency: Two Case Reports and Review of Literature

Beibei Yan¹, Chao Wang², Kaihui Zhang³, Haiyan Zhang³, Min Gao³, Yuqiang Lv³, Xiaoying Li¹, Yi Liu^{3*} and Zhongtao Gai^{3*}

¹ Neonatology Department, Qilu Children's Hospital of Shandong University, Ji'nan, China, ² Shandong Freshwater Fisheries Research Institute, Ji'nan, China, ³ Pediatric Research Institute, Qilu Children's Hospital of Shandong University, Ji'nan, China

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University,
China

Reviewed by:

Dong Dong,
East China Normal University,
China
Abdallah El-Sayed Allam,
Tanta University, Egypt

*Correspondence:

Yi Liu
liuyi-ly@126.com
Zhongtao Gai
gaizhongtao@sina.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 29 June 2018

Accepted: 09 July 2019

Published: 22 August 2019

Citation:

Yan B, Wang C, Zhang K, Zhang H,
Gao M, Lv Y, Li X, Liu Y and Gai Z
(2019) Novel Neonatal Variants of the
Carbamoyl Phosphate Synthetase 1
Deficiency: Two Case Reports and
Review of Literature.
Front. Genet. 10:718.
doi: 10.3389/fgene.2019.00718

Carbamoyl phosphate synthetase I (CPS1) deficiency (CPS1D), is a rare autosomal recessive disorder, characterized by life-threatening hyperammonemia. In this study, we presented the detailed clinical features and genetic analysis of two patients with neonatal-onset CPS1D carrying two compound heterozygous variants of c.1631C > T (p.T544M)/c.1981G > T (p.G661C), and c.2896G > T (p.E966X)/c.622-3C > G in *CPS1* gene, individually. Out of them, three variants are novel, unreported including a missense (c.1981G > T, p.G661C), a nonsense (c.2896G > T, p.E966X), and a splicing change of c.622-3C > G. We reviewed all available publications regarding *CPS1* mutations, and in total 264 different variants have been reported, with majority of 157 (59.5%) missense, followed by 35 (13.2%) small deletions. This study expanded the mutational spectrum of *CPS1*. Moreover, our cases and review further support the idea that most ($\geq 90\%$) of the mutations were "private" and only $\sim 10\%$ recurred in unrelated families.

Keywords: carbamoyl phosphate synthetase 1 deficiency, carbamoyl phosphate synthetase 1, urea cycle disorders, next-generation sequencing, missense, nonsense, deletion, splicing

INTRODUCTION

Carbamoyl phosphate synthetase I (CPS1) deficiency (CPS1D) is a rare autosomal recessive urea cycle disorder, characterized by hyperammonemia with the incidence of 1/50,000 to 1/300,000 (Díez-Fernández et al., 2015). CPS1D is currently divided into two types of neonatal onset and late onset, whereas CPS1D with severe manifestations of hyperammonemia is common in neonatal-onset patients (Choi et al., 2017; Rokicki et al., 2017; Yang et al., 2017; Zhang et al., 2018). Typically, the neonatal-onset patient with CPS1D appears to be healthy at birth, but deteriorates rapidly into severe hyperammonemia, presenting poor feeding, vomiting, hypotonia, irritability, seizures, hypothermia, lethargy, coma, apnea, and even death after first feeding (Funghini et al., 2012; Choi et al., 2017; Rokicki et al., 2017; Zhang et al., 2018).

The function of urea cycle is to transform toxic ammonia into non-toxic urea. CPS1 catalyzes the initial and limiting step of the urea cycle, which is critical in the detoxification of excess ammonia, so CPS1D patient suffering from hyperammonemia will present a decreased level of citrulline but elevated glutamine in blood amino acid analysis, and a low level of orotic acid in urine test (Funghini et al., 2012; de Cima et al., 2015; Ali et al., 2016).

It is difficult to timely diagnose CPS1D due to atypical manifestations like sudden onset, rapid progress, and low morbidity, as well as complicated and non-recurrent genetic mutations in *CPS1* gene (Choi et al., 2017; Rokicki et al., 2017; Zhang and Li, 2017). For more than a decade, the diagnosis of CPS1D has been mainly relied on the laboratory tests of tandem mass spectrometry (MS/MS) including liquid chromatography-tandem mass spectrometry (LC-MS/MS) and gas chromatography mass spectrometry (GC/MS). MS/MS is a high-throughput technique for measurement of the intermediate metabolites and has been widely used to distinguish dozens of metabolic diseases (Lehotay et al., 2011; Janecková et al., 2012; Hao et al., 2018). However, this technology cannot differentiate CPS1D from *N*-acetylglutamate synthase deficiency (NAGSD) in UCDs due to their similar intermediate metabolites. Until recently, next-generation sequencing (NGS), a powerful DNA sequencing technology, has revolutionized genomic research with great utility in the molecular diagnosis of genetic disorders (Choi et al., 2017; Jia et al., 2017; Rokicki et al., 2017; Li et al., 2018; Zhang et al., 2018), and has been proven reliable and important to detect *CPS1* mutation for early diagnosis of CPS1D, as the severity of clinical manifestations in CPS1D patients is determined by the extent of *CPS1* deficiency (Choi et al., 2017; Chen et al., 2018; Zhang et al., 2018).

In this study, we performed the clinical examinations and mutation analysis on two neonatal patients with CPS1D. The LC-MS/MS and GC/MS were carried out to detect amino acids in blood and organic acidurias in the urine, and then NGS was utilized to test the gene mutation. Strikingly, for the first time, we identified three novel pathogenic mutations of *CPS1*. To our knowledge, there have been so far only three reports of using NGS to detect *CPS1* mutations for CPS1D diagnosis (Choi et al., 2017; Chen et al., 2018; Zhang et al., 2018). Our novel findings further expanded the mutational spectrum of *CPS1* and demonstrated additional evidences of using NGS for precise identification of *CPS1* mutations in patients.

MATERIALS AND METHODS

Patients, Samples and Ethical Approval

This study was approved by Medical Ethics Committee of Qilu Children's Hospital of Shandong University. The written informed consents were obtained from the parents of each study participant, and the patients' information was anonymized before submission. All the procedures performed in the study were in accordance with the Declaration of Helsinki.

Two patients from two unrelated families who were from the neonatal intensive care unit (NICU) of Qilu Children's Hospital of Shandong University (QCHSU) were firstly screened by LC/MS-MS and GC/MS. The parents of both patients were healthy and non-consanguineous. Blood samples were obtained from the patients and their parents in accordance with informed consents in the study.

In addition, 100 blood samples from healthy children were collected as control samples for mutations validation.

Routine Examination and Biochemical Laboratory Tests

Routine physical examination, complete blood count (CBC), C-reactive protein (CRP), hemoculture, and biochemical laboratory tests, such as liver function, kidney function, glucose, ammonia, lactic acid, and blood gas analyses, were carried out.

The level of orotic acids in urine was measured by GC/MS with GCMS-QP2010 analyzer (Shimadzu, Tokyo, Japan) and analyzed by the Inborn Errors of Metabolism Screening System software (Shimadzu), whereas amino acids level in blood was detected by LC-MS/MS with Applied Biosystems API 3200 analyzer (ABSCIEX, Foster City, CA) and analyzed by the ChemoView software (ABSCIEX).

Next-Generation Sequencing and Variant Discovery

Genomic DNA was extracted and purified from peripheral blood of the two patients and their parents using TIANamp Blood Genomic DNA Purification Kit (Tiagen Biotech, Beijing, China). Whole-exome sequencing was applied to test mutation of genes in both patients. Approximately 3 µg of genomic DNA was randomly fragmented. An exome enriched kit (Agilent, Santa Clara, CA) was used to obtain the coding exons and flanking intronic regions. The sequencing was performed using HiSeq2000 sequencer (Illumina, San Diego, CA). The obtained mean exome coverage was over 99.2%, and average sequencing depth of each sample was 100%. Raw data obtained from the sequencer were further analyzed including read alignment, variant calling, and annotation by SinoPath Enterprise Ltd (Beijing, China). Low-quality reads (quality score ≤ 20 and sequencing depth ≤ 5) in the raw data were removed. Filtered reads were aligned to the human reference genome (UCSC hg19 Feb.2009) using the Burrows-Wheeler Aligner (Raney et al., 2014). Single-nucleotide variants (SNVs) and small insertions/deletions (indels) can be detected. Annotation was carried out by ANNOVAR for gene information, protein functional predictions, and population allele frequencies (Wang et al., 2010). Variants outside of coding regions and greater than 1% MAF (minor allele frequency) in the population were excluded.

Bioinformatic Analysis and Verification of Mutations

All known variants were reported according to the following databases: OMIM (<http://www.ncbi.nlm.nih.gov/omim/limits>), UCSC Genome Bioinformatics (<http://genome.ucsc.edu/>), Human Gene Mutation Database (<http://www.hgmd.cf.ac.uk/ac/index.php>), Single Nucleotide Polymorphism Database (dbSNP) (<http://www.ncbi.nlm.nih.gov/SNP/>), 1000 Genomes Database (<http://browser.1000genomes.org>), ExAC (<http://exac.broadinstitute.org/about>), and gnomAD (<http://gnomad.broadinstitute.org/>). *In silico* analysis of the variants was carried out using PolyPhen-2 and SIFT and Mutation Taster to predict the pathogenicity. Human Splicing Finder (HSF) was applied to predict the effect of splicing variant. The multiple-sequence alignments were carried out by ClustalX. Modeling of affected protein structure was processed using

SWISS-MODEL. The data analysis was conducted referring to the document (Jin et al., 2018). All the selected variants were classified as pathogenic, likely pathogenic, a variant of unknown significance (VUS), likely benign, or benign according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015). The potential pathogenic mutations were validated by Sanger sequencing.

RESULTS

Clinical Characteristics of Two Patients

The clinical manifestations and laboratory data from the two patients were summarized in **Table 1**. Both patients were neonatal-onset type presenting fulminant symptoms due to serious hyperammonemia so that the life-sustaining mechanical ventilation, medications of vasopressors, liquid infusion, and ammonia scavengers were administered.

Patient 1 (P1), a full-term female, the first child of healthy unrelated parents, was vaginally delivered. Her mother had regular prenatal care starting from 12 weeks of pregnancy. She was apparently healthy at birth with weight of 2.95 kg, Apgar score of 10 at 1 min and 5 min after birth. The following day, however, she had a fever, and then gradually developed hyporeactiveness presenting respiratory distress, seizures, and acute circulatory collapse so she was immediately transported to NICU in QCHSU from local hospital. Laboratory tests revealed the abnormal blood indexes of ammonia, 1,404 $\mu\text{mol/L}$ (reference, 18–72 $\mu\text{mol/L}$); citrulline, 3.82 $\mu\text{mol/L}$ (reference, 4–30 $\mu\text{mol/L}$); alanine, 1,264.4 $\mu\text{mol/L}$ (reference, 62.9–328 $\mu\text{mol/L}$); lactic acid, 5.8 mmol/L (reference, 0.7–2.1 mmol/L); glucose, 0.3 mmol/L (reference, 3.3–6.1 mmol/L); and white blood cells $22.06 \times 10^9/\text{L}$ (reference, $5.0\text{--}14.5 \times 10^9/\text{L}$), as well as abnormal urinary indexes of undetected orotic acid and elevated 3-MGA 15.7 mmol/L (ref. 0–4 mmol/L). The chest radiograph result reported pneumonia and possible atelectasis. Patient heart rate reached up to 180 beats/min but the ejection fraction was only 38% and no signs of congenital heart disease (**Table 1**).

The mechanical ventilation and medications of vasopressors, liquid infusion, and antibiotics, ammonia scavengers, such as dopamine, dobutamine, dilator, meropenem, lactulose, and L-arginine were administered immediately. Meanwhile, oral feeding was forbidden, and total parenteral nutrition with lower amino acid was administered. Unfortunately, the patient deteriorated continually into multiple-organ failure and even had cardiac arrest with no spontaneous breathing. Considering the poor prognosis, her parents gave up her treatment, and she died at age of 5 days.

Patient 2 (P2), a full-term girl, the second child of healthy unrelated parents, was vaginally delivered. The first child of the family died suddenly at the third day after birth without a definite diagnosis. Her mother had regular prenatal care, and she was normal at birth with a weight of 2.9 kg, Apgar score of 10 at 1 and 5 min after birth. On the third day, however, she had a sudden onset of hyperlactacidemia and deteriorated even faster than P1 did. At the beginning, she was hyporeactive presenting grunting and anorectic, but no manifestations of

TABLE 1 | Clinical and laboratory data of the two patients with CPS1D.

Patients	P1	P2
Gender	Female	Female
Age at onset	2D	3D
Age deceased	5D	4D
Complications during pregnancy	Uneventful	uneventful
Weeks gestation at delivery	40	39 ⁺⁵
Birth weight (kg), Apgar score	2.95, 10	2.9, 10
Clinical course	fulminant	fulminant
Clinical features		
Fever	+	–
Seizure	+	–
Coma	+	+
Cyanosis	+	+
Breathing	weak	grunting
Hemorrhage	–	Lung, stomach
Cardiac failure	+	+
Poor feeding	+	+
Abdominal distention		+
Urine	oliguria	anuria
Arterial blood gas analysis		
PH (reference, 7.25–7.45)	7.13↓	7.10↓
PO ₂ (reference, 50–80 mm Hg)	67	72
PCO ₂ (reference, 40–60 mm Hg)	42	62↑
HCO ₃ (reference, 19–30 mmol/L)	12↓	19.2
BE (reference, –3 to +3)	–18↓	–10.5↓
Blood routine test		
Red blood cells (reference, $3.5\text{--}5.5 \times 10^{12}/\text{L}$)	5.68↑	3.35↓
White blood cells (reference, $5.0\text{--}14.5 \times 10^9/\text{L}$)	22.06↑	24.77
Platelet count (reference, $100\text{--}300 \times 10^9/\text{L}$)	354↑	290
Hemoglobin (reference, 138–218 g/L)	195	117↓
Urine routine test		
BLD (negative)	++	+++
PRO (negative)	++	+
KET (negative)	–	–
Blood biochemical tests		
ALT (reference, 0–38 U/L)	37	24
Lactic acid (reference, 0.7–2.1 mmol/L)	5.8↑	5.6↑
Glucose (reference, 3.3–6.1 mmol/L)	0.3↓	12↑
Potassium (reference, 3.5–5.5 mmol/L)	7.4↑	4.6
Sodium (reference, 135–145 mmol/L)	154↑	142
AST (reference, 0–38 U/L)	130↑	78↑
CK (reference, 21–220 U/L)	936↑	780↑
CK-MB (reference, 0–25 U/L)	30↑	67↑
PCT (reference, ≤ 0.5 ng/ml)	67.322↑	0.794↑
ammonia (reference, 18–72 $\mu\text{mol/L}$)	1404↑	823↑
Blood mass spectrometry profile		
Citrulline (reference, 4–30 $\mu\text{mol/L}$)	3.82↓	3.08↓
Alanine (reference, 62.9–328 $\mu\text{mol/L}$)	1264.4↑	3337.99↑
Proline (reference, 72–293 $\mu\text{mol/L}$)	634.45↑	413.38↑
Ornithine (reference, 42–358 $\mu\text{mol/L}$)	69.01	106.95
Urinary organic acids		
Urinary orotic acid (reference, 0–2 mmol/L)	0	0
Urinary 3-MGA (reference, 0–4 mmol/L)	15.7↑	45.75↑
Chest X-ray		
Pneumonia	+	–
Pneumorrhagia	–	+
Echocardiography		
Ejection fraction	38%↓	37%↓
Patent ductus arteriosus	+	–
Patent foramen ovale	+	+
CPS1 sequencing		
Allele 1 (from father)	c.1631C > T (p.T544M)	c.2896G > T (p.E966X)
Allele 2 (from mother)	c.1981G > T (p.G661C)	c.622-3C > G

+positive, ↑elevated, ↓decreased.

fever, vomiting, and seizures. Five hours later, she developed pneumorrhagia, gastrointestinal hemorrhage, and anuria, so that she was immediately transferred to NICU in QCHSU from her local hospital. On the way to the hospital, her heart rate and oxygen saturation could not be maintained; cardio-pulmonary resuscitation and mechanical ventilator had to be administered. Nevertheless, she deteriorated very quickly, presenting coma, shock, and irregular respirations. When she was admitted, she looked pale with reduced perfusion and a low ejection fraction (37.9%) in her echocardiography. The blood flowed out of her intratracheal tube and nose. Her pupil diameter was about 4 mm, and pupillary reflex disappeared. Laboratory tests revealed abnormal blood indexes of ammonia, 823 $\mu\text{mol/L}$; citrulline, 3.08 mmol/L; alanine, 3,337.99 $\mu\text{mol/L}$ (reference, 62.9–328 $\mu\text{mol/L}$); lactic acid, 5.6 mmol/L; glucose, 12 mmol/L; and white blood cells, $24.77 \times 10^9/\text{L}$, as well as abnormal urinary indexes of undetected orotic acid and increases 3-MGA 45.75 mmol/L. The chest radiograph showed exudative lesions, which matched her pulmonary hemorrhage.

This patient received immediate treatment that was similar as P1 with mechanical ventilation, vasopressors, liquid infusion, ammonia scavengers (such as lactulose and L-arginine), dopamine, dobutamine, dilator, as well as total parenteral nutrition with lower amino acid. She continued to deteriorate with tremendous speed and no sign of improvement after 13 h of admission and died at the age of 4 days.

Genetic Analysis and Pathogenicity Prediction

Whole-exome sequencing showed two compound heterozygous variants of the CPS1 gene in both P1 and P2, individually, including two missense variants of c.1631C > T (p.T544M) and c.1981G > T (p.G661C) found in P1, a nonsense variant c.2896G > T (p.G966X) and a splicing variant c.622-3C > G detected in P2. Of which, the variant c.1631C > T (p.T544M) was a known pathogenic mutation causing CPS1D (Finckh et al., 1998; Häberle et al., 2011) (**Table 1** and **Figures 1A, B**), whereas the remaining three variants of c.1981G > T (p.G661C), c.622-3C > G, and c.2896G > T (p.G966X) were novel and unreported in publications and public databases of OMIM, UCSC, HGMD, dbSNP, 1000 genomes, ExAC, and gnomAD. The missense c.1981G > T (p.G661C) occurred with an amino acid change from a nonpolar amino acid of glycine (G) to a polar amino acid of cysteine (C); the nonsense c.2896G > T (p.G966X) would create a premature stop codon; the splicing change c.622-3C > G was predicted to affect acceptor splice site. There were no mutations found in the control samples by using Sanger sequencing.

The pathogenicity of three novel variants was further analyzed using various prediction online tools. In brief, HSF was applied to assess the potential impacts on the splicing of three novel variants as all these variants located near intron–exon junction. The predicted results showed that all three variants in the exon 17 (c.1981G > T), intron 7 (c.622-3C > G), and exon 24 (c.2896G > T) probably affect the splice sites (**Figures 2A–C**). The missense mutations of c.1981G > T (p.G661C)

were predicted to be pathogenic by SIFT, MutationTaster, and PolyPhen-2 (**Figure 3A**). The conservation analysis of the variants of c.1981G > T (p.G661C) and c.2896G > T (p.G966X) in CPS1 showed that both sites were highly conservative in different species by using ClustalX (**Figures 3B, C**), whereas the missense variant c.1981G > T (p.G661C) was predicted to change the highly evolutionary conserved amino acid in CPS1, and the nonsense variant c.2896G > T (p.G966X) causing a premature stop could generate a truncated protein with missing conserved site of CPS1. In addition, the CPS1 protein crystallographic structure of both mutant types (p.G661C and p.G966X) revealed the changes of side strand structure and H-bond in variant of p.G661C, and a truncated protein with loss of 534 amino acids in variant of p.G966X (**Figures 4A, B**). All the mutation information and clinical data were uploaded into eRAM (Jia et al., 2018).

DISCUSSION

Pediatric rare diseases are often rapid deterioration with high mortality, which can be obviously improved by early diagnosis and treatment (Ni et al., 2017). CPS1D is a rare inborn error of UCD caused by CPS1 deficiency manifesting sudden onset, rapid progress, and low morbidity. In this study, we presented the detailed clinical manifestations and mutation analysis of two neonatal CPS1D cases. First, the blood ammonia, amino acids, and urine organic acids test results reported severe hyperammonemia on both patients, and the patients have very high levels of alanine and decreased levels of citrulline in blood, as well as the increased levels of 3-MGA and decreased levels of orotic acid in urine. We therefore referred patients for whole-exome sequencing to determine the genetic cause of this inborn error of metabolism. After validation of Sanger sequencing, two compound heterozygous variants in CPS1 were identified in both patients, and one missense variant (c.1631C > T, p.T544M) was of known pathogenicity (Finckh et al., 1998; Häberle et al., 2011), whereas other three were novel and predicted to be pathogenic. Therefore, both patients were finally diagnosed as neonatal-onset CPS1D caused by CPS1 mutations. To our knowledge, this study is the fifth case report of CPS1D in China and the 262–264th novel mutations in CPS1 documented in the world (Chen et al., 2013; Yang et al., 2017; Chen et al., 2018; Zhang et al., 2018), which expands the mutation spectrum of CPS1 (**Supplementary Tables S1 to S4**).

CPS1 is an enzyme that catalyzes the first and rate-limiting reaction of three steps in ammonia detoxification of the urea cycle from ammonia to carbamoyl phosphate (de Cima et al., 2015; Ali et al., 2016). Normal function of the urea cycle requires six enzymes, including CPS1 as well as two mitochondrial transporters (Helman et al., 2014). CPS1 deficiency caused by CPS1 gene mutation usually leads to accumulation of ammonia in the blood and thereby presents severe hyperammonemia, which is neurotoxic resulting in neonatal death or severe and irreversible brain damage in the developing and mature brain (Funghini et al., 2012; Choi

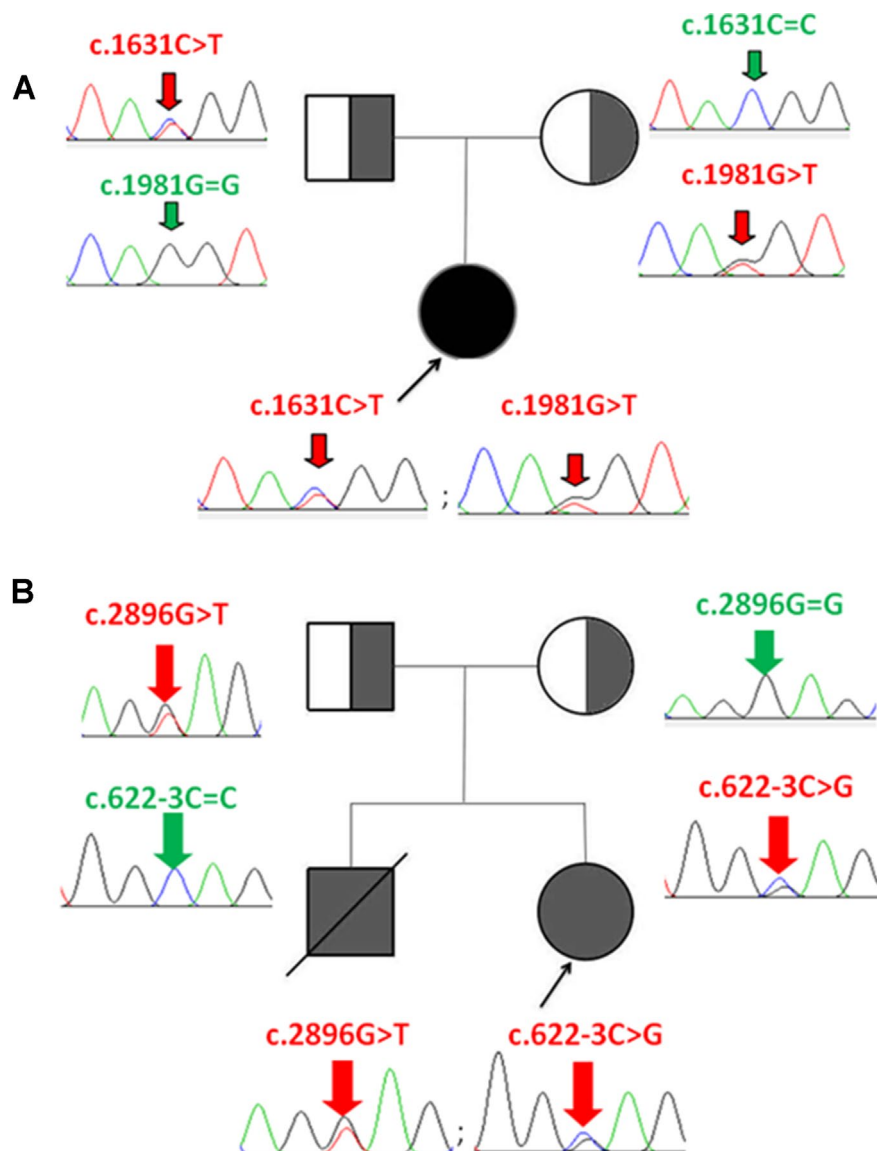


FIGURE 1 | Identification of CPS1 mutations and pedigree of the two families with CPS1D probands. **(A)** P1 has two compound heterozygous mutations of c.1631C > T and c.1981G > T, inherited from her father and mother, respectively. **(B)** P2 has two compound heterozygous mutations of c.2896G > T and c.622-3C > G, inherited from her father and mother, respectively.

et al., 2017; Rokicki et al., 2017; Yang et al., 2017). CPS1D are divided into two types of lethal neonatal-onset or less severe late-onset based on the age of onset, clinical features, and severity of CPS1 deficiency (Diez-Fernandez et al., 2017). To date, most of the reported CPS1D cases are neonatal-onset with severe hyperammonemia and usually died of multiorgan failure. Signs and symptoms in patients with CPS1D are often atypical with rapid progression and extremely low morbidity, which makes the clinical diagnosis difficult (Funghini et al., 2012; Choi et al., 2017; Rokicki et al., 2017).

As a result, the diagnosis of CPS1D is heavily dependent on laboratory data, such as blood ammonia, blood amino acids, urine organic acids, and genetic testing. The determination of blood ammonia concentration is critical for early clinical

evaluation as it often reaches 150 $\mu\text{mol/L}$ or higher in acute stage (László et al., 1991). Abnormal level of amino acids can be detected by mass spectrometry, like elevated blood glutamate, glutamine, and alanine, and reduced citrulline and arginine, whereas decreased urinary orotic acid and increased urinary 3-MGA (Lehotay et al., 2011; Janecková et al., 2012). Next-generation sequencing has been increasingly accessible in clinical laboratory for precise diagnosis of inborn errors of metabolism, including CPS1D (Choi et al., 2017; Yang et al., 2017; Li et al., 2018); for the biochemical tests mentioned above, one cannot distinguish from different types of UCD, particularly N-acetylglutamate synthase deficiency (NAGSD) from CPS1D due to similar intermediary metabolites (Choi et al., 2017).

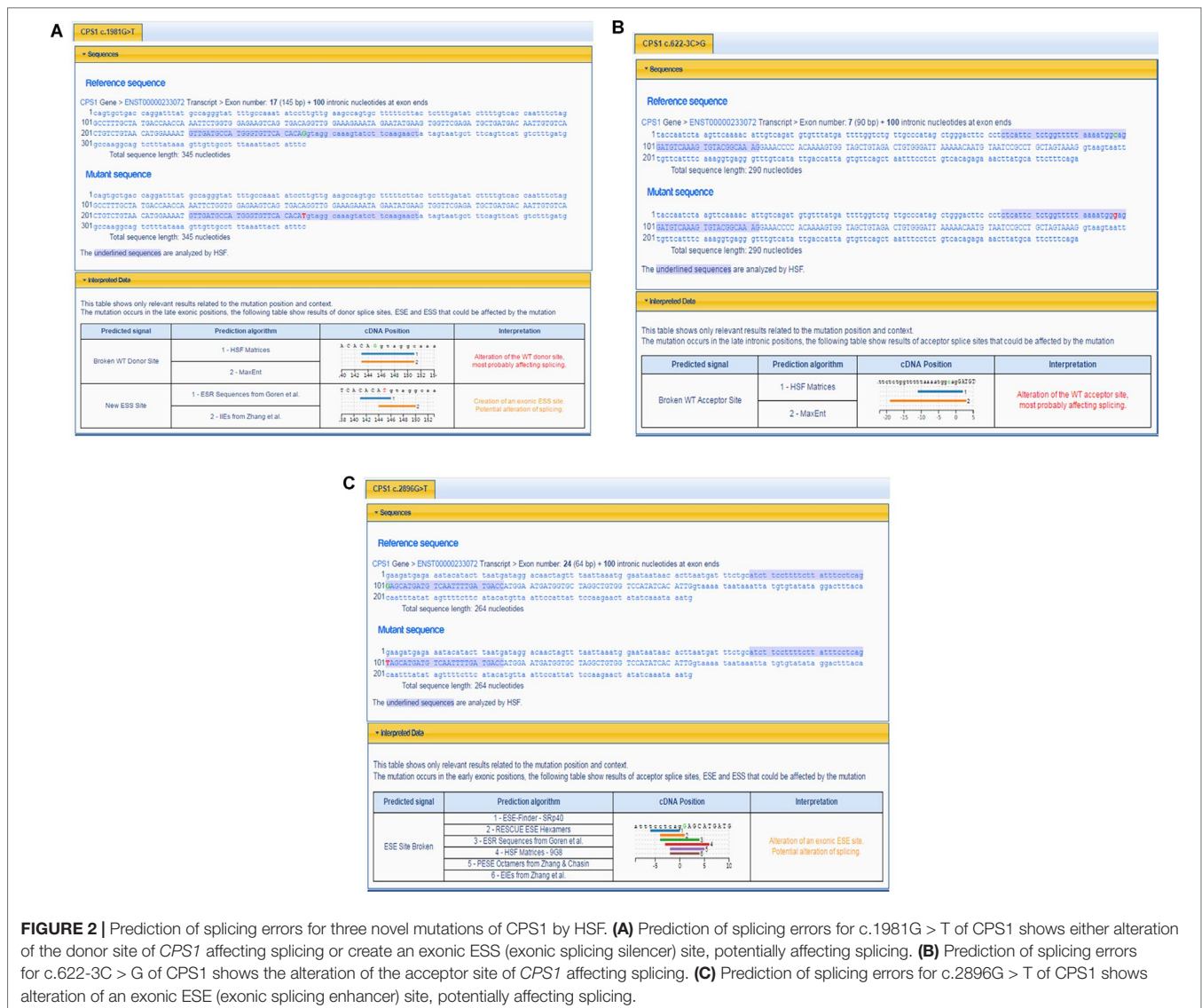


FIGURE 2 | Prediction of splicing errors for three novel mutations of CPS1 by HSF. **(A)** Prediction of splicing errors for c.1981G > T of CPS1 shows either alteration of the donor site of CPS1 affecting splicing or create an exonic ESS (exonic splicing silencer) site, potentially affecting splicing. **(B)** Prediction of splicing errors for c.622-3C > G of CPS1 shows the alteration of the acceptor site of CPS1 affecting splicing. **(C)** Prediction of splicing errors for c.2896G > T of CPS1 shows alteration of an exonic ESE (exonic splicing enhancer) site, potentially affecting splicing.

CPS1 (NM_001875.4) located on chromosome 2q34, spans over 122 kb consisting of 38 exons, which encode a polypeptide of 1,500 amino acids. Up to now, different variations of CPS1 have been reported, including missense, nonsense, small deletions, small insertions, small indels (insertions+deletions), and large deletions. As far as both cases of this study are concerned, two missense variants of c.1631C > T (p.T544M) and c.1981G > T (p.G661C) were found in a 2-day neonate girl (P1) with severe hyperammonemia. The variant of c.1631C > T (p.T544M) was a previously reported mutation (Finckh et al., 1998; Häberle et al., 2011) and was proven in the expression study causing large decrease of the enzyme activity due to hampering of the cross-talk between the bicarbonate phosphorylation domain (BPSD) and the allosteric NAG binding domain (ASD) (Diez-Fernandez et al., 2013). Another missense mutation (c.1981G > T, p.G661C) in P1 was novel and unreported. The pathogenicity of c.1981G > T (p.G661C) was analyzed and predicted as “deleterious” by SIFT, “protein features affected”

by MutationTaster, and “probably damaging” with a score of 1.000 (sensitivity: 0.00; specificity: 1.00) by PolyPhen-2 (Figure 3A). As the variant occurred at the last exon base of the 17 exon, the potential splicing effect was assessed by HSF, which revealed a possible donor site error affecting the splicing of mRNA (Figure 2A). Moreover, the amino acid substitution site was highly conservative in different species analyzed by ClustalX (Figure 3B), so the alteration of p.G661C from glycine to cysteine was predicted to not only interrupt the conserved position of glycine but also change the side strand structure in the CPS1 protein crystallographic model, which produced a defective protein (Figure 4A). In addition, the mutation site of c.1981G > T (p.G661C) is identical as that of a known variant c.1981G > C (p.G661R) with the substitution of a small glycine residue by a large side chain, which was predicted to decrease structural stability of the protein (Funghini et al., 2012). Thus, we inferred that the substitution of amino acids from glycine to cysteine (p.G661C) might have a similar damage to the CPS1

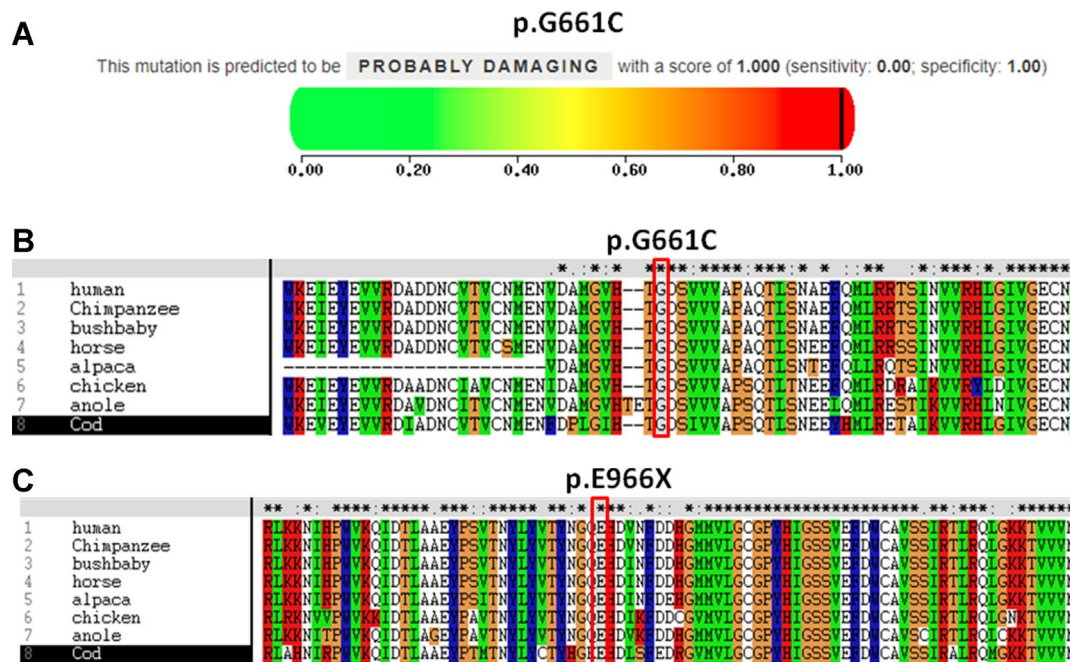


FIGURE 3 | Pathogenicity and conservation analysis of the missense and nonsense mutations of CPS1. **(A)** Pathogenicity analysis of c.1981G > T (p.G661C) shows “probably damaging” with a score of 1.000 (sensitivity: 0.00; specificity: 1.00) by PolyPhen-2. **(B)** *In silico* analysis of c.1981G > T (p.G661C) in CPS1 shows the site p.G661 highly conservative in different species of human, chimpanzee, bushbaby, horse, alpaca, chicken, anole, and cod. **(C)** *In silico* analysis of c.2896G > T (p.E966X) in CPS1 shows the site p.E966 highly conservative in different species of human, chimpanzee, bushbaby, horse, alpaca, chicken, anole, and cod.

stability, and both missense mutations in *CPS1*, leading to defects of enzyme function, are therefore the genetic cause of the patient with CPS1D.

A nonsense variant of c.2896G > T (p.G966X) and a splicing site change of c.622-3C > G were detected in a 3-day neonatal girl (P2). Since the nonsense variant site (c.2896G > T, p.G966X) is the first exon base of the 24th exon, its probable impact on splicing was assessed by HSF and revealed a potential effect of splicing with the alteration of an ESE (exonic splicing enhancer) site (**Figure 2C**). Besides, this position of glycine was highly conservative in different species analyzed by ClustalX (**Figure 3C**); the variant was predicted to generate a truncated protein with a loss of 534 amino acids and abolish the activity of the enzyme. The crystallographic structure model of the G966X further demonstrated the truncated protein (**Figure 4B**). The mutation of c.622-3C > G in intron 7 was a splicing site change that was predicted to alter the acceptor site of *CPS1* gene and affect mRNA splicing, which would produce a nonfunctional enzyme (**Figure 2B**). The severe phenotype of P2 with more rapid progress to multiple-organ failure within 13 h from her onset suggested that both alleles encode a nonfunctional protein. Additionally, an unusual death of the first boy in the family drew our attention. The first child was born at term and apparently healthy after a normal pregnancy. He had sudden deterioration and died on the third day without a definite diagnosis. The retrospective analysis of the first boy from this family demonstrated that he had similar features as P2. We conjectured that the first child might carry

the identically compound heterozygous variants inherited from their father and mother as his sister P2. On this occasion, genetic counseling and prenatal genetic testing are necessary for the subsequent pregnancy. Considering the positions of the novel mutations and their potential splicing defects, RT-PCR should be used to identify the possibly aberrant transcripts; unfortunately, we failed to get the RNA from both patients due to their parents’ refusal.

We reviewed all publications of *CPS1* variants in cases of CPS1D, and a total of 264 *CPS1* different variations (including the 3 variants in this study) have been reported. Among them, the missense variants were the majority accounting for 157 (59.5%), followed by small deletions of 35 (13.2%), splice site changes of 25 (9.5%), and nonsense of 22 (8.3%), whereas the minority of the variants were 4 (1.5%) small indels and 5 (1.9%) large deletions with missing 1,000 bp to 767 kb, which were detected by genomic microarray (**Figure 5**). Of the variants, 81 (30.7%) were predicted to cause protein truncation, including 22 nonsense, 31 small deletions, 16 small insertions, 4 small indels, 6 splicing site changes, and 2 large deletions. Our reviewing data further clarified that most *CPS1* variants ($\geq 90\%$) were “private” with non-recurrence, and the few recurrent mutations tended to occur at CpG dinucleotides, which made the diagnosis more complicated (**Supplementary Tables S1 to S4**) (Hoshida et al., 1993; Finckh et al., 1998; Summar et al., 1998; Ihara et al., 1999; Aoshima et al., 2001a; Aoshima et al., 2001b; Rapp et al., 2001; Wakutani et al., 2001; Häberle et al., 2003; Eeds et al., 2006; Kurokawa et al., 2007;

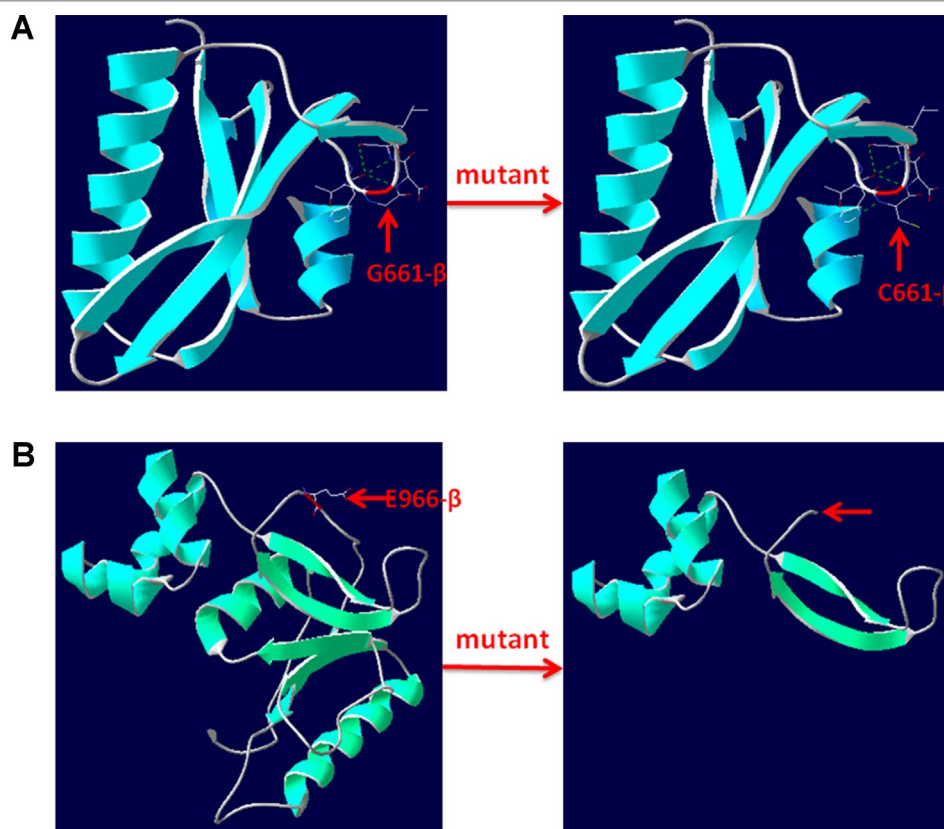


FIGURE 4 | 3D structure of wild type and mutant type of p.G661C and p.E966X in CPS1. **(A)** Wild type and mutant type of p.G661 in CPS1 shows the mutation of G661C changes the side strand structure and H-bond. **(B)** Wild type and mutant type of E966 in CPS1 shows the mutation of E661X leading to a truncated protein with missing 534 amino acids. The red arrows marked the sites of wild-type G661, E966, and mutated C661, X966.

Distribution of mutation spectrum in *CPS1*

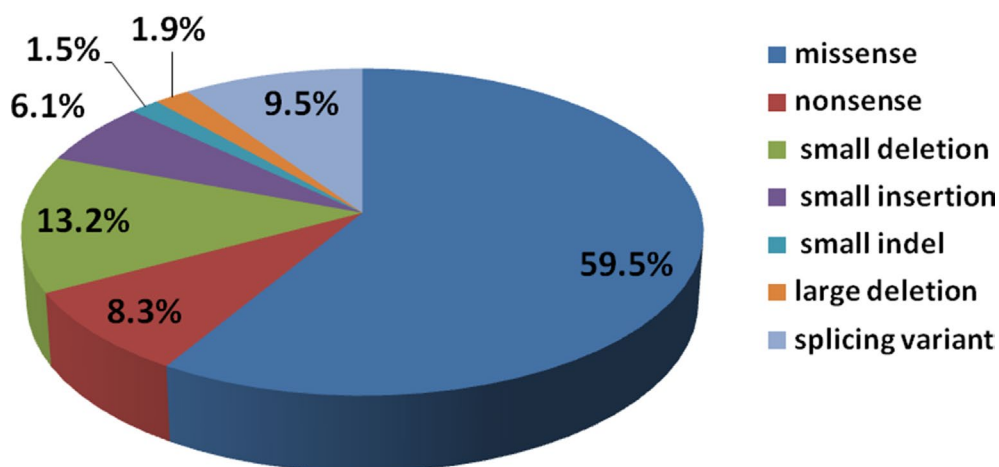


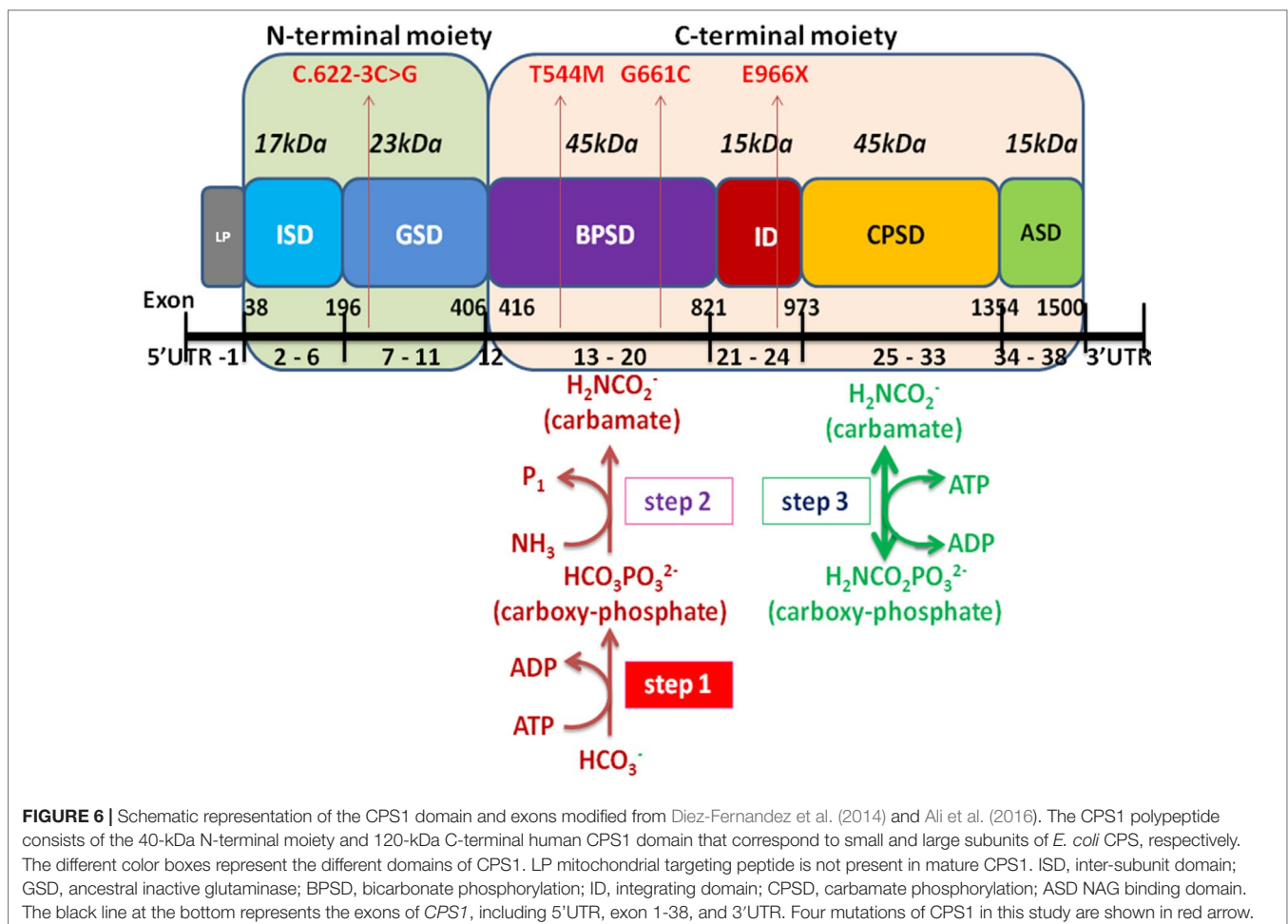
FIGURE 5 | Distribution of mutation spectrum in CPS1 reported in both cases and literature. A total of 259 mutations in *CPS1* include 58.7% missense, 8.5% nonsense, 13.5% small deletion, 6.2% small insertion, 9.6% splicing, 1.6% small indel, and 1.9% large deletion.

Khayat, 2009; Ono et al., 2009; Pekkala et al., 2010; Häberle et al., 2011; Wang et al., 2011; Funghini et al., 2012; Kretz et al., 2012; Diez-Fernandez et al., 2014; Ali et al., 2016; Choi et al., 2017; Rokicki et al., 2017; Yang et al., 2017; Chen et al., 2018; Zhang et al., 2018).

Encoded by *CPS1* gene, CPS1 is a complex multidomain enzyme, composed of a 40-kDa N-terminal moiety with two unknown function domains and a 120-kDa C-terminal moiety involving four domains of bicarbonate phosphorylation (BPSD), integrating (ID), carbamate phosphorylation (CPSD), and allosteric NAG binding (ASD) (Diez-Fernandez et al., 2014; de Cima et al., 2015). The C-terminal moiety contains two ATP-binding sites, catalyzing the synthesis of carbamoyl phosphate from bicarbonate, ATP, and ammonia and has been discovered to possess missense mutations of *CPS1* with high frequency and plays a critical integrating role in folding of structural elements leading to decreased yield of CPS1 (Figure 6) (Häberle et al., 2011; Diez-Fernandez et al., 2013; Diez-Fernandez et al., 2014). We analyzed the distribution of the 264 mutations and found that 66 (25%) variants were located at N-terminal moiety, whereas 198 (75%) mutations were at C-terminal moiety involving 76 variants at BPSD, 29 at ID, 75 at CPSD, and 18 at ASD.

Three novel mutations were found in the study, two located at BPSD of C-terminal and one at N-terminal moiety, which further supported the importance of the C-terminal moiety in maintaining the function of CPS1.

At present, the treatment of CPS1D is to strictly follow the recommendations of UCDs, which focuses on reduction of ammonia production by a restricted protein diet and management of ammonia scavengers, such as sodium benzoate, sodium phenylbutyrate, and sodium phenylacetate, as well as drugs of L-arginine and L-citrulline to improve the residual urea cycle function and the renal excretion of ammonia (Diez-Fernandez and Häberle, 2017). In case of severe hyperammonemia, hemodialysis or peritoneal dialysis can be administered (Häberle, 2012; Diez-Fernandez and Häberle, 2017). However, these approaches cannot cure CPS1D, and the only available cure currently is liver transplantation, which has demonstrated excellent results with approximately 90% survival rate in UCD children, though it is limited by donor sources (Diez-Fernandez and Häberle, 2017; Zhang and Li, 2017). To date, most of the CPS1D patients died before receiving the confirmed diagnosis, so the detection of blood ammonia, blood amino acids, urine organic acid, and next-generation sequencing should be performed as early as possible.



CONCLUSION

In this study, we presented the detailed clinical features and genetic analysis of two patients with neonatal-onset CPS1D and discovered three novel pathogenic variants in *CPS1* by whole-exome sequencing with a comprehensive outline of available publications regarding *CPS1* gene mutations. A total 264 different variants of *CPS1* have been reported with the majority of 157 (59.5%) missense, followed by small deletions of 35 (13.2%), and the minority of 5 (1.9%) large deletions and 4 (1.5%) indels, of which 81 (30.7%) were predicted to cause protein truncation. Our data further expand the spectrum of *CPS1* mutation and support the clinical applicability of whole-exome sequencing for genetic diagnosis of UCD.

ETHICS STATEMENT

The work was approved by Medical Ethics Committee of Qilu Children's Hospital of Shandong University. Written informed consents was obtained from the patients' parents and the patients' information was anonymized before submission. All the procedures performed in the study were in accordance with the Declaration of Helsinki.

REFERENCES

- Ali, E. Z., Khalid, M. K. N. M., Yunus, Z. M., Yakob, Y., Chin, C. B., Latif, K. A., et al. (2016). Carbamoyl phosphate synthetase 1 (CPS1) deficiency: clinical, biochemical, and molecular characterization in Malaysian patients. *Eur. J. Pediatr.* 175 (3), 339–346. doi: 10.1007/s00431-015-2644-z
- Aoshima, T., Kajita, M., Sekido, Y., Kikuchi, S., Yasuda, I., Saheki, T., et al. (2001a). Novel mutations (H337R and 238-362del) in the CPS1 gene cause carbamoyl phosphate synthetase I deficiency. *Hum. Hered.* 52 (2), 99–101. doi: 10.1159/000053360
- Aoshima, T., Kajita, M., Sekido, Y., Mimura, S., Itakura, A., Yasuda, I., et al. (2001b). Carbamoyl phosphate synthetase I deficiency: molecular genetic findings and prenatal diagnosis. *Prenat. Diagn.* 21 (8), 634–637. doi: 10.1002/pd.123
- Chen, M., Zhang, A., and Xu, Q. (2013). A case report of carbamoyl phosphate synthetase I deficiency. *Jilin Med. (in Chinese)* 34 (35), 7563–7564.
- Chen, X., Yuan, L., Sun, M., Liu, Q., and Wu, Y. (2018). Two novel CPS1 mutations in a case of carbamoyl phosphate synthetase 1 deficiency causing hyperammonemia and leukodystrophy. *J. Clin. Lab. Anal.* 32 (5), e22375. doi: 10.1002/jcla.22375
- Choi, R., Park, H. D., Yang, M., Ki, C. S., Lee, S. Y., Kim, J. W., et al. (2017). Novel pathogenic variant (c.580C > T) in the CPS1 gene in a newborn with carbamoyl phosphate synthetase 1 deficiency identified by whole exome sequencing. *Ann. Lab. Med.* 37 (1), 58–62. doi: 10.3343/alm.2017.37.1.58
- de Cima, S., Polo, L. M., Díez-Fernández, C., Martínez, A. I., Cervera, J., Fita, I., et al. (2015). Structure of human carbamoyl phosphate synthetase: deciphering the on/off switch of human ureagenesis. *Sci. Rep.* 5, 16950. doi: 10.1038/srep16950
- Díez-Fernández, C., and Häberle, J. (2017). Targeting CPS1 in the treatment of carbamoyl phosphate synthetase 1 (CPS1) deficiency, a urea cycle disorder. *Expert Opin. Ther. Targets* 21 (4), 391–399. doi: 10.1080/14728222.2017.1294685
- Díez-Fernández, C., Martínez, A. I., Pekkala, S., Barcelona, B., Perez-Arellano, I., Guadalajara, A. M., et al. (2013). Molecular characterization of carbamoyl-phosphate synthetase (CPS1) deficiency using human recombinant CPS1 as a key tool. *Hum. Mutat.* 34 (8), 1149–1159. doi: 10.1002/humu.22349
- Díez-Fernández, C., Hu, L., Cervera, J., Häberle, J., and Rubio, V. (2014). Understanding carbamoyl phosphate synthetase (CPS1) deficiency by using the recombinantly purified human enzyme: effects of CPS1 mutations that

AUTHOR CONTRIBUTIONS

This study was conceived and designed by ZG and YL. The experiments were conducted by KZ, MG, YQL, and HZ. Data analyzed by KZ and YL. BY, XL and ZG contributed clinical diagnosis of the patients. The paper was written by BY, CW and YL.

ACKNOWLEDGMENTS

This work was financially supported by Science and Technology Foundation of Shandong Province (2013GSF11829) and Jinan Excellent Science and Technology Innovation Team Project (20150515). The authors are grateful to the patients and their parents for their contribution to the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00718/full#supplementary-material>

- concentrate in a central domain of unknown function. *Mol. Genet. Metab.* 112 (2), 123–132. doi: 10.1016/j.ymgme.2014.04.003
- Díez-Fernández, C., Gallego, J., Häberle, J., Cervera, J., and Rubio, V. (2015). The study of carbamoyl phosphate synthetase 1 deficiency sheds light on the mechanism for switching on/off the Urea Cycle. *J. Genet. Genomics* 42 (5), 249–260. doi: 10.1016/j.jgg.2015.03.009
- Eeds, A. M., Hall, L. D., Yadav, M., Willis, A., Summar, S., Putnam, A., et al. (2006). The frequent observation of evidence for nonsense-mediated decay in RNA from patients with carbamyl phosphate synthetase I deficiency. *Mol. Genet. Metab.* 89 (1-2), 80–86. doi: 10.1016/j.ymgme.2006.04.006
- Finckh, U., Kohlschütter, A., Schaefer, H., Sperhake, K., Colombo, J. P., and Gal, A. (1998). Prenatal diagnosis of carbamoyl phosphate synthetase I deficiency by identification of a missense mutation in CPS1. *Hum. Mutat.* 12 (3), 206–211. doi: 10.1002/(SICI)1098-1004(1998)12:3<206::AID-HUMU8>3.0.CO;2-E
- Funghini, S., Thusberg, J., Spada, M., Gasperini, S., Parini, R., Ventura, L., et al. (2012). Carbamoyl phosphate synthetase 1 deficiency in Italy: clinical and genetic findings in a heterogeneous cohort. *Gene* 493 (2), 228–234. doi: 10.1016/j.gene.2011.11.052
- Hao, C., Guo, J., Guo, R., Qi, Z., Li, W., and Ni, X. (2018). Compound heterozygous variants in POR gene identified by whole-exome sequencing in a Chinese pedigree with cytochrome P450 oxidoreductase deficiency. *Pediatr. Invest.* 2 (2), 90–95. doi: 10.1002/ped4.12035
- Häberle, J. (2012). Carglumic acid for the treatment of N-acetylglutamate synthase deficiency and acute hyperammonemia. *Expert Rev. Endocrinol. Metab.* 7 (3), 263–271. doi: 10.1586/eem.12.17
- Häberle, J., Schmidt, E., Pauli, S., Rapp, B., Christensen, E., Wermuth, B., et al. (2003). Gene structure of human carbamylphosphate synthetase 1 and novel mutations in patients with neonatal onset. *Hum. Mutat.* 21 (4), 444. doi: 10.1002/humu.9118
- Häberle, J., Shchelochkov, O. A., Wang, J., Katsonis, P., Hall, L., Reiss, S., et al. (2011). Molecular defects in human carbamoyl phosphate synthetase I: mutational spectrum, diagnostic and protein structure considerations. *Hum. Mutat.* 32 (6), 579–589. doi: 10.1002/humu.21406
- Helman, G., Pacheco-Colon, I., and Gropman, A. L. (2014). The urea cycle disorders. *Semin. Neurol.* 34 (3), 341–349. doi: 10.1055/s-0034-1386771
- Hoshida, R., Matsuura, T., Haraguchi, Y., Endo, F., Yoshinaga, M., and Matsuda, I. (1993). Carbamylphosphate synthetase I deficiency. One base substitution in an exon of the CPS1 gene causes a 9-basepair deletion due to aberrant splicing. *J. Clin. Invest.* 91 (5), 1884–1887. doi: 10.1172/JCI116405

- Ihara, K., Nakayama, H., Hikino, S., and Hara, T. (1999). Mutation in CPS1. *Hum. Genet.* 105, 375.
- Janecková, H., Hron, K., Wojtowicz, P., Hlídková, E., Baresová, A., Friedecký, D., et al. (2012). Targeted metabolomics analysis of plasma samples for the diagnosis of inherited metabolic disorders. *J. Chromatogr. A* 1226, 11–17. doi: 10.1016/j.chroma.2011.09.074
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60 (7), 686–691. doi: 10.1007/s11427-017-9099-3
- Jia, J., An, Z., Ming, Y., Guo, Y., and Li, W. (2018). eRAM: encyclopedia of rare disease annotations for precision medicine. *Nucleic Acids Res.* 46, D937–D943. doi: 10.1093/nar/gkx1062
- Jin, Y., Zhang, L., Ning, B., Hong, H., Xiao, W., Tong, W., et al. (2018). Application of genome analysis strategies in the clinical testing for pediatric diseases. *Pediatr. Invest.* 2 (2), 72–81. doi: 10.1002/ped4.12044
- Khayat, M. (2009). Novel human pathological mutations. Gene symbol: CPS1. Disease: carbamoyl phosphate synthetase I deficiency. *Hum. Genet.* 125 (3), 336. doi: 10.1007/s00439-009-0642-9
- Kretz, R., Hu, L., Wettstein, V., Leiteritz, D., and Häberle, J. (2012). Phytohemagglutinin stimulation of lymphocytes improves mutation analysis of carbamoylphosphate synthetase I. *Mol. Genet. Metab.* 106 (3), 375–378. doi: 10.1016/j.ymgme.2012.04.011
- Kurokawa, K., Yorifuji, T., Kawai, M., Momoi, T., Nagasaka, H., Takayanagi, M., et al. (2007). Molecular and clinical analyses of Japanese patients with carbamoylphosphate synthetase I (CPS1) deficiency. *J. Hum. Genet.* 52 (4), 349–354. doi: 10.1007/s10038-007-0122-9
- László, A., Karsai, T., and Várkonyi, A. (1991). Congenital hyperammonemia: symptomatic carrier girl patient and her asymptomatic heterozygous mother for ornithine transcarbamylase (OTC) deficiency: specific enzyme diagnostic and kinetic investigations for the detection of heterozygous genostatus. *Acta Paediatr. Hung.* 31 (3), 291–297.
- Lehotay, D. C., Hall, P., Lepage, J., Eichhorst, J. C., Etter, M. L., and Greenberg, C. R. (2011). LC–MS/MS progress in newborn screening. *Clin. Biochem.* 44 (1), 21–31. doi: 10.1016/j.clinbiochem.2010.08.007
- Li, X., Yang, Y., Gao, Q., Gao, M., Lv, Y., Dong, R., et al. (2018). Clinical characteristics and mutation analysis of five Chinese patients with maple syrup urine disease. *Metab. Brain Dis.* 33 (3), 741–751. doi: 10.1007/s11011-017-0168-0
- Ni, X., and Shi, T. L. (2017). The Challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60 (7), 681–685. doi: 10.1007/s11427-017-9100-1
- Ono, H., Suto, T., Kinoshita, Y., Sakano, T., Furue, T., and Ohta, T. (2009). A case of carbamoyl phosphate synthetase I deficiency presenting symptoms at one month of age. *Brain Dev.* 31 (10), 779–781. doi: 10.1016/j.braindev.2008.12.013
- Pekkala, S., Martinez, A. I., Barcelona, B., Yefimenko, I., Finckh, U., Rubio, V., et al. (2010). Understanding carbamoyl-phosphate synthetase I (CPS1) deficiency by using expression studies and structure-based analysis. *Hum. Mutat.* 31 (7), 801–808. doi: 10.1002/humu.21272
- Raney, B. J., Dreszer, T. R., Barber, G. P., Clawson, H., Fujita, P. A., Wang, T., et al. (2014). Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. *Bioinformatics* 30, 1003–1005. doi: 10.1093/bioinformatics/btt637
- Rapp, B., Häberle, J., Linnebank, M., Wermuth, B., Marquardt, T., Harms, E., et al. (2001). Genetic analysis of carbamoylphosphate synthetase I and ornithine transcarbamylase deficiency using fibroblasts. *Eur. J. Pediatr.* 160 (5), 283–287. doi: 10.1007/s004310100725
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* 17, 405–424. doi: 10.1038/gim.2015.30
- Rokicki, D., Pajdowska, M., Trubicka, J., Thong, M. K., Ciara, E., Piekutowska-Abramczuk, D., et al. (2017). 3-Methylglutaconic aciduria, frequent but underrecognized finding in carbamoyl phosphate synthetase I deficiency. *Clin. Chim. Acta* 471, 95–100. doi: 10.1016/j.cca.2017.05.023
- Summar, M. L. (1998). Molecular genetic research into carbamoyl-phosphate synthase I: molecular defects and linkage markers. *J. Inher. Metab. Dis.* 21 Suppl 1, 30–39. doi: 10.1023/A:1005349306311
- Wakutani, Y., Nakayasu, H., Takeshima, T., Mori, N., Kobayashi, K., Endo, F., et al. (2001). A case of late-onset carbamoyl phosphate synthetase I deficiency, presenting periodic psychotic episodes coinciding with menstrual periods. *Rinsho Shinkeigaku* 41 (11), 780–785.
- Wang, J., Shchelochkov, O. A., Zhan, H., Li, F., Chen, L. C., Brundage, E. K., et al. (2011). Molecular characterization of CPS1 deletions by array CGH. *Mol. Genet. Metab.* 102 (1), 103–106. doi: 10.1016/j.ymgme.2010.08.020
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVER: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic. Acids. Res.* 38, e164. doi: 10.1093/nar/gkq603
- Yang, X. Y., Shi, J., Lei, H. H., Xia, B., and Mu, D. Z. (2017). Neonatal-onset carbamoyl phosphate synthetase I deficiency. *Medicine (Baltimore)* 96 (26), e7365. doi: 10.1097/MD.00000000000007365
- Zhang, Y., and Li, B. (2017). Cell and gene therapy for carbamoyl phosphate synthetase I deficiency. *J. Pediatr. Neonatal Care* 7 (1), 00273.
- Zhang, G., Chen, Y., Ju, H., Bei, F., Li, J., Wang, J., et al. (2018). Carbamoyl phosphate synthetase I deficiency diagnosed by whole exome sequencing. *J. Clin. Lab. Anal.* 32 (2), e22241. doi: 10.1002/jcla.22241

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yan, Wang, Zhang, Zhang, Gao, Lv, Li, Liu and Gai. The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Association of *SERPINC1* Gene Polymorphism (rs2227589) With Pulmonary Embolism Risk in a Chinese Population

Yongjian Yue^{1†}, Qing Sun^{2†}, Lu Xiao^{1,3}, Shengguo Liu¹, Qijun Huang¹, Minlian Wang¹, Mei Huo⁴, Mo Yang⁴ and Yingyun Fu^{1*}

¹ Key Laboratory of Shenzhen Respiratory Diseases, Department of Pulmonary and Critical Care Medicine, Shenzhen Institute of Respiratory Disease, The First Affiliated Hospital of Southern University of Science and Technology, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Shenzhen, China, ² Shenzhen Key Laboratory of Reproductive Immunology for Peri-implantation, Fertility Center, Shenzhen Zhongshan Urology Hospital, Shenzhen, China, ³ Research Centre, The Seventh affiliated Hospital of Sun Yat-sen University, Shenzhen, China, ⁴ Department of Clinical Laboratory, The First Affiliated Hospital of Southern University of Science and Technology, The Second Clinical Medical College of Jinan University, Shenzhen People's Hospital, Shenzhen, China

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Mariza De Andrade,
Mayo Clinic, United States
María Eugenia De La Morena-Barrio,
University of Murcia, Spain
Javier Corral,
University of Murcia, Spain

*Correspondence:

Yingyun Fu
yingyunfu2017@163.com

[†]These authors have contributed
equally to this work

Specialty section

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 29 January 2019

Accepted: 14 August 2019

Published: 13 September 2019

Citation:

Yue Y, Sun Q, Xiao L, Liu S, Huang Q,
Wang M, Huo M, Yang M and Fu Y
(2019) Association of *SERPINC1*
Gene Polymorphism (rs2227589)
With Pulmonary Embolism Risk in a
Chinese Population.
Front. Genet. 10:844.
doi: 10.3389/fgene.2019.00844

Background and Aims: Genetic variants in the gene *SERPINC1* have been shown to be associated with antithrombin deficiency, which subsequently contributes to the susceptibility to venous thrombosis. However, several other studies have shown conflicting results regarding the association of *SERPINC1* gene polymorphisms (rs2227589) with the risk of thrombosis. Hence, in the present study, we conducted a case-control study to further evaluate the association between the variant rs2227589 with antithrombin deficiency in pulmonary embolism (PTE). A pooled systematic analysis was also conducted to evaluate the risk of rs2227589 in venous thromboembolism (VTE) among multiple populations.

Methods: This case-control study involved 101 patients and 199 healthy controls. The allele frequency of *SERPINC1* variant rs2227589 was analyzed by Sequenom assay. Antithrombin anticoagulant activity was detected using an automatic coagulation analyzer. In addition, a pooled systematic analysis on 10 cohorts consisting of 5,518 patients with VTE and 8,935 controls was performed.

Results: In total, 27 (26.7%) PTE subjects were diagnosed as having antithrombin deficiency. Our results showed that antithrombin plasma activity was slightly lower in T allele carriers than that in C allele carriers. However, there was no significant correlation between rs2227589 genotype and antithrombin anticoagulant activity. The recessive model showed that rs2227589 was significantly associated ($p = 0.026$) with an increased risk {odds ratio [OR]: 2.31, 95% confidence interval [CI] (1.09–4.89)} of Chinese PTE. The pooled systematic analysis of all case-control study and meta-analysis showed that rs2227589 polymorphism was associated with an increased risk of VTE in the additive model [OR: 1.09, 95% CI (1.01–1.18), $P = 0.029$] and dominant model [OR: 1.10, 95% CI (1.01–1.20), $P = 0.034$].

Conclusions: Our study demonstrated that variant rs2227589 is associated with an increased risk of PTE in a Chinese population but no correlation with antithrombin anticoagulant activity. However, pooled systematic analysis of multiple populations showed a significant association between rs2227589 and the risk of VTE in the additive and dominant genetic model.

Keywords: *SERPINC1*, rs2227589, pooled systematic analysis, antithrombin anticoagulant activity, pulmonary embolism

INTRODUCTION

Venous thromboembolism (VTE) is a complex and common cardiovascular disease that includes deep venous thrombosis (DVT), cerebral infarction, and pulmonary embolism (PTE), which is caused by multiple factors. The incidence of VTE is around 0.1–0.2% in Caucasian and American populations (Beckman et al., 2010). The hospitalization rates of VTE increased from 3.2 to 17.5 per 100,000 population, and the mortality decreased from 4.7% to 2.1% in China (Zhang et al., 2019). The acquired or inherited risk factors for the development of VTE include surgery, pregnancy, and cancer (Heit, 2015; Hotoleanu, 2017). Familial cohort and case-control studies have demonstrated that VTE is often familial or hereditary (Mili et al., 2011; Holzhauer et al., 2012). Over 60% of the variations in susceptibility to common thrombosis are attributable to genetic factors (Souto et al., 2000).

Genetic variations in coagulation system genes (such as *F5*, *PROC*, and *SERPINC1*) contribute to susceptibility to venous thrombosis (Garcia de Frutos et al., 2007; Rosendaal and Reitsma, 2009). Protein C (PROC), protein S (PROS1), and antithrombin have been demonstrated to play important roles in the anticoagulation process (Lee et al., 2017). *SERPINC1* is the gene encoding antithrombin. Deficiency of antithrombin is usually caused by rare or private variations of *SERPINC1* gene. Deficiency among coagulation system factors can increase the risk of developing thrombosis. Inherited antithrombin deficiency is a rare, autosomal dominant disorder (MIM#107300). Antithrombin exerts its physiological function by inhibiting procoagulation factors, such as thrombin, factor Xa, factor IIa, and other factors of the blood coagulation system (Zeng et al., 2015). Antithrombin belongs to the serine protease inhibitor superfamily and regulates clot formation both by inhibiting thrombin activity directly and by interfering with earlier stages of the clotting cascade (Rosenberg and Bauer, 1987). There are two types of antithrombin deficiency. In type I antithrombin deficiency, functional and antigenic levels are proportionally decreased. In type II antithrombin deficiency, antigenic levels are normal while the functional activity is abnormal. In around 0.02–0.25% of a healthy population with antithrombin deficiency, there is a 5- to 50-fold increased risk of developing venous thrombosis (Zhu et al., 2011; Kim et al., 2014; Luxembourg et al., 2014).

The first variation linked to antithrombin deficiency was characterized in 1983 and, to date, more than 200 variants have been reported to be associated with the risk of thrombosis (Corral et al., 2007; Navarro-Fernandez et al., 2016). The homozygous

variant (Phe229Leu) of *SERPINC1* leading to spontaneous antithrombin polymerization *in vivo* has been shown to be associated with severe childhood thrombosis (Picard et al., 2003). The heterozygous variant is mainly associated with a high risk of venous thrombosis (Arruda et al., 1997; Navarro-Fernandez et al., 2016). However, most of these variants are rare and seldom replicate in other populations. rs2227589, a polymorphism of *SERPINC1* gene (NG_012462.1:g.5301G > A), was found to be associated with the risk of venous thrombosis in a Dutch population (Bezemer et al., 2008). The minor allele frequencies of rs2227589 is 0.10 (gnomAD) and 0.329 in the East Asian population. Previous study of normal Spanish Caucasian showed a functional effect of the rs2227589 on antithrombin levels (Anton et al., 2009). The findings regarding the association between rs2227589 and antithrombin levels were inconsistent among different studies (Segers et al., 2014; Bhakuni et al., 2015). Furthermore, the association between rs2227589 and the risk of VTE in Northern European, American, and Norwegian populations is inconsistent (Tregouet et al., 2009; Austin et al., 2011; Dahm et al., 2012; Jiang et al., 2017; Kajuna et al., 2018). The study of Jiang did not enroll a large sample size as in the study of Tregouet et al. (2009) and showed negative results with publication bias (Jiang et al., 2017). Similarly, another Swedish population study showed no significant association between rs2227589 and the risk of VTE (Bruzeliu et al., 2015). These findings imply that associations might differ between Western and Asian populations depending on ethnicity.

Given that the association between rs2227589 and the Chinese PTE population has not been examined, we performed this case-control study to investigate the association between variant rs2227589, antithrombin deficiency, and PTE risk in a Chinese population. Furthermore, a systematic review and a meta-analysis were carried out to evaluate the variant in antithrombin deficiency and VTE.

MATERIALS AND METHODS

Study Subjects

A total of 98 patients with PTE, 3 patients with familial VTE, and 199 matched controls were recruited from the Second Clinical Medical College of Shenzhen People's Hospital from December 2013 to September 2017. The three familial VTE subjects were enrolled because each had more than two siblings who were diagnosed with PTE. Strict criterion was conducted to exclude the multiple PTE effected risk factors in this association analysis,

such as cancer, diabetes, and so on. In this regard, only 101 cases were enrolled in our studies. The inclusion criteria for the PTE patients were the criteria released by the European Society of Cardiology (ESC) in 2014 (Konstantinides, 2014). Most of the recruited subjects are unprovoked, recurrent, or inherited acute PTE. The healthy controls were recruited as described previously (Yue et al., 2019). Informed consent was obtained from each patient. Research and ethics approval was obtained from the ethics committee of Shenzhen People's Hospital. All procedures involving human participants were performed in accordance with the 1964 Declaration of Helsinki. Demographic characteristics and medical histories were recorded, including age and history of family disease.

Plasma Antithrombin Anticoagulant Activity Assay in PTE Subjects

Blood samples were collected into trisodium citrate by venopuncture. Plasma samples were obtained by centrifugation at 2,000g for 15 min at 4°C and stored at -80°C until analysis. The antithrombin anticoagulant activity was determined by automatic chromogenic assay on the IL Coagulation systems according to the manufacturer's instructions (Cat.0020300400, Werfen, Bedford, MA). Briefly, the plasma was first incubated with Factor Xa reagent in the presence of an excess of heparin, followed by incubation with a synthetic chromogenic substrate. Quantification of the residual Factor Xa was performed by detecting the absorbance at 405 nm by ACLTOP 700 system (Werfen). The absorbance is inversely proportional to the antithrombin level in the test sample. The activation levels were calculated by the standard curve of each detection. The established normal range of plasma antithrombin levels was 83–128%, which was based on our hospital pathology reference from thousands of previously diagnosed clinical subjects.

DNA Preparation and Genotyping

Genomic DNA was extracted from whole blood using QIAamp DNA Blood kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The candidate single-nucleotide polymorphism (SNP) was genotyped by MassARRAY SNP genotyping platform (Sequenom, San Diego, CA) among PTE case and control subjects. The following primers for polymerase chain reaction (PCR) amplification were used: forward, 5'-ACGTTGGATGGAAAGGCCTTACCCCAAGAG-3' and reverse, 5'-ACGTTGGATGTCTCCCTGGTAGTTACAGTC-3'. The genotyping assay extension primers for SNP rs2227589 were 5'-GGAGAGCACTTGAAATGAT-3'. All primers were designed by Sequenom's MassARRAY Designer software. A specific primer extension was performed to detect single-base polymorphisms in the amplified DNA.

Systematic Review Analysis

A systematic literature search was performed using PubMed, Science Direct, ISI Web of Knowledge, Google Scholar, and the CNKI Database, and the cutoff date was defined as January 2019. A meta-analysis was performed in accordance with the Preferred Reporting Items for Systematic Reviews and

Meta-Analyses (PRISMA) (Moher et al., 2009). Search terms using the Boolean operators at PubMed included but were not limited to the following: "venous thrombosis" (MeSH Terms) OR VTE (All Fields) AND *SERPINC1* (MeSH Terms) OR rs2227589 (All Fields). Only case-control studies regarding an association between rs2227589 and VTE were considered eligible. The data extracted from the eligible studies included author, publication year, racial and genotype methods, and variants frequency. The control population genotype data of publications followed the Hardy-Weinberg equilibrium.

Eligibility Criteria and Data Validity Assessment

The eligibility criteria of studies are according to our previous publication (Yue et al., 2019). The criterion for including eligible studies was that the DVT or PTE should be diagnosed with standard criteria in the case group. Only case-control studies regarding an association between rs2227589 and VTE were considered eligible. For exclusion, reviews, conference abstracts, republished or duplicate studies, and meta-analyses were not included. The data extracted from the eligible studies included author, publication year, racial and genotype methods, and variants frequency. The control population genotype data of publications followed the Hardy-Weinberg equilibrium. The validity of data was evaluated by three independent reviews. The genotype results of two eligible studies were not complete or available in the publications. Finally, the genotype data were kindly shared by professor Pierre-Emmanuel Morange (Tregouet et al., 2009), but the other study results were not obtained from the authors upon request (Heit et al., 2017).

Statistical Analysis and Meta-Analysis

All statistical analyses were performed using software STATA 14.0 and SPSS 19 (IBM). The strength of association was evaluated by pooled odds ratio (OR) and 95% confidence interval (CI). Either a fixed-effects model (Mantel-Haenszel method) or a random-effects model was used for meta-analysis. The z-test was used to evaluate the significance of the pooled OR (Lewis, 2002). Heterogeneity was tested using Cochran's Q statistic (Higgins et al., 2003), and publication bias was tested using Egger's test (Egger et al., 1997; Sterne et al., 2001). Nonparametric trim-and-fill method was performed to assess the possible effect of funnel plot asymmetry and publication bias in the meta-analysis (Duval and Tweedie, 2000). Sensitivity analysis was conducted by sequentially omitting one study at a time. Chi square and Bonferroni (*post hoc*) tests were used for cross-tabulation and multiple comparisons, respectively. $P < 0.05$ was considered statistically significant.

RESULTS

Demographics of the Participants

After filtering with the inclusion criteria and exclusion criteria, a total of 101 patients were enrolled for plasma antithrombin anticoagulant activity level assay in this study. The general

TABLE 1 | Demographics of the participants and antithrombin anticoagulant activity in patients with pulmonary embolism.

Characteristic/ Genotype	Number (%) or mean \pm SEM	Antithrombin activity (Mean \pm SEM)	p value
Gender			
Male/female(n)	51/50		
Age (years)	59 \pm 1.7		
Total subjects	101	90 \pm 2.0%	83–128% (normal range)
Deficiency case	27	62.9 \pm 2.8%	P < 0.001
Normal case	84	99.9 \pm 1.4%	
rs2227589 genotype			
CC	40 (39.6%)	91.6 \pm 2.8%	P > 0.05
CT	45 (44.6%)	88.5 \pm 3.5%	
TT	16 (15.8%)	90.3 \pm 4.6%	

characteristics of PTE patients are presented in **Table 1**. The mean age of the patients with PTE was 59 \pm 1.7 years old. Antithrombin activity deficiency was present in 27 patients (26.7%) with venous thrombosis. The average level of active antithrombin was 90 \pm 2.0%. The details of antithrombin anticoagulant activity in patients with PTE were presented in **Supplemental Table S1**.

Association of Genetic Models Between rs2227589 and PTE Risk

Allele and genotype frequencies of *SERPINC1* polymorphism (rs2227589) in PTE patients and controls are summarized in **Figure 1** and **Table 2**. The genotypes of rs2227589 followed the

Hardy-Weinberg equilibrium among the controls. In our study, the CT genotype (heterozygote) was 44.6%, which was much higher than in the previously reported ethnic cohorts (around 20%). Furthermore, homozygote SNPs were also present in the PTE group at a very high frequency (>7%). The frequencies of the SNP homozygote and heterozygote were similar between the case and control groups. Carriers of *SERPINC1* polymorphism (rs2227589) were associated ($P = 0.026$) with an increased risk of PTE [OR: 2.31, 95% CI (1.09–4.89)], which was detected in the recessive model but not in the additive and dominant models (**Table 3**).

Association Assay Between Plasma Antithrombin Anticoagulant Activity and Frequency of rs2227589 Carriers

The correlation between rs2227589 and antithrombin anticoagulant activity of PTE was also evaluated. Among rs2227589 carriers, 16 (16/61) showed antithrombin deficiency compared with noncarriers (11/40) (**Table 1**). Only 4 subjects showed antithrombin deficiency among the 16 homozygote SNP carriers. There was no significant difference in antithrombin activity levels among the three rs2227589 genotype groups. Carriers of the TT alleles had slightly lower anticoagulant activity than did carriers of the CC alleles, but this was not statistically significant ($P = 0.45$). We also checked whether significant differences existed among the associated recessive genetic model groups (CC + CT vs. TT). The results showed that the T allele carriers had slightly lower antithrombin activity, but this was not statistically significant (**Table 3**).

Meta-Analysis for the Association Between rs2227589 and VTE Risk in Different Populations

The genotyping data of rs2227589 in 101 Chinese PTE patients of our case-control study were enrolled in the meta-analysis. In pooled systematic analysis, 10 case-control cohorts from six studies were enrolled. The total numbers of VTE-affected patients and controls were 5,518 and 8,935, respectively. The characteristics, genotype distribution, and allelic frequencies of the eligible studies are displayed in **Table 2**. The meta-analysis was carried out using the additive, recessive, and dominant genetic models. The heterogeneity value was less than 50% ($P > 0.05$), indicating that our study had low heterogeneity and that a fixed model could be used in the analysis (**Table 4**). The association effect distribution is presented as forest plots (**Figure 2**, **Supplementary Figure S1**). The association effect of pooled OR with the additive model was 1.09 (95% CI 1.01–1.18, $P = 0.029$) and the dominant model was 1.10 (95% CI 1.01–1.20, $P = 0.034$). Systematic analysis of previous cohort studies showed a significant association between rs2227589 and the risk of VTE under additive and dominant genetic models. Subgroups analysis of Caucasians showed consistent results of the association (**Table 4**). Furthermore, potential publication bias was examined by funnel and Galbraith plots (**Figure 3**; **Supplemental Figure S2**). The results showed that there was

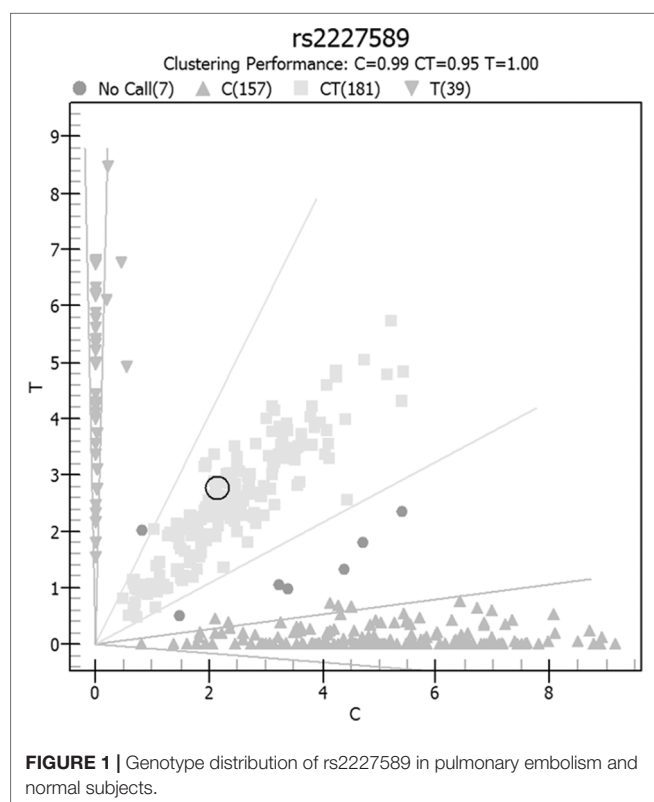
**FIGURE 1 |** Genotype distribution of rs2227589 in pulmonary embolism and normal subjects.

TABLE 2 | Genotype distribution of *SERPINC1* polymorphism (rs2227589) among different studies.

Author	Years	Ethnicity	Source	Case			Control			Methods
				CC	CT	TT	CC	CT	TT	
Bezemer	2008	Netherlands	Mega2	1,001	278	15	2325	483	28	PCR
Tregouet	2009	Northern European	GWAS	317	89	5	983	235	10	Beadchip
Tregouet	2009	Northern European	MARTHA	922	228	12	621	149	12	Taqman
Tregouet	2009	Northern European	FARIVE	478	111	5	467	119	1	Taqman
Austin	2011	White America	White	426	113	5	530	127	7	Taqman
Austin	2011	Black America	Black	469	58	2	499	71	5	Taqman
El-Galaly, TC	2013	Danish	White	494	105	8	1,434	279	17	Taqman
Rovite, V	2014	Latvia	Latvia	139	38	0	179	53	3	Taqman
Kajuna, I	2018	Latvia	Latvia	82	17	0	71	27	1	Taqman
Yue, YY	2019	Asian	Chinese	40	45	16	84	100	15	Sequenom

TABLE 3 | OR, 95% CIs, and antithrombin activity levels in three genetic models of rs2227589 among patients with pulmonary embolism.

Genetic models		Pooled effect		Antithrombin activity	
Models	Allele	OR (95% CI)	P	Mean \pm SEM	P
Additive	T vs. C	1.27 (0.89–1.81)	0.18	88.9 \pm 2.9 vs. 89.9 \pm 2.3	0.79
Dominant	TT + TC vs. CC	1.11 (0.68–1.81)	0.67	89.0 \pm 2.9 vs. 91.6 \pm 2.8	0.25
Recessive	TT vs. TC + CC	2.31 (1.09–4.89)	0.026	86.8 \pm 3.7 vs. 90.6 \pm 2.3	0.5

TABLE 4 | The results of pooled OR, 95% CIs, and heterogeneity by meta-analysis.

Genetic models		Pooled effect		Z	Heterogeneity	
Models	Allele	OR (95% CI)	P _Z		I ² (%)	P _H
Additive	T vs. C	1.09 (1.01–1.18)	0.029	2.18	44.3	0.06
	Subgroups (Caucasian)	1.10 (1.01–1.20)	0.023	2.27	44.9	0.08
Dominant	TT + TC vs. CC	1.10 (1.01–1.20)	0.034	2.21	40.1	0.09
	Subgroups (Caucasian)	1.12 (1.02–1.22)	0.017	2.39	45.1	0.08
Recessive	TT vs. TC + CC	1.17 (0.85–1.61)	0.328	0.98	16.4	0.29
	Subgroups (Caucasian)	1.06 (0.74–1.52)	0.741	0.33	0.0	0.54

no sharp asymmetry between the two models, which indicated no obvious potential bias among the enrolled publications. But Galbraith assay showed potential bias of additive and dominant models, which indicated that the bias may be caused by the small sample size or ethnic differences of enrolled cohorts. Thus, trim-and-fill method analysis was applied to correct publication bias. The results showed the pooled effect unchanged of the two genetic models and continued to be statistically significant, which indicated that the effect of bias is very slight with reliable conclusions (**Supplemental Figure S3**). Sensitivity analyses showed that the pooled ORs fluctuated among confidence intervals (**Figure 4**), which suggested good reliability among our results and methods to provide an effective evaluation.

DISCUSSION

Because of controversial results from GWAS and case-control studies of rs2227589, a prediction of the risk factors associated with developing VTE is unreliable. In this study, we explored the association between the variant rs2227589 with antithrombin deficiency and risk of PTE in a Chinese population. Our case-control study found a significant association between *SERPINC1* rs2227589 polymorphism and increased risk of PTE in the recessive model. Pooled systematic analysis of all cohorts showed a significant association in the additive and dominant genetic models. The association between rs2227589 and repeated episodes of VTE demonstrated a genetic risk with ethnic differences.

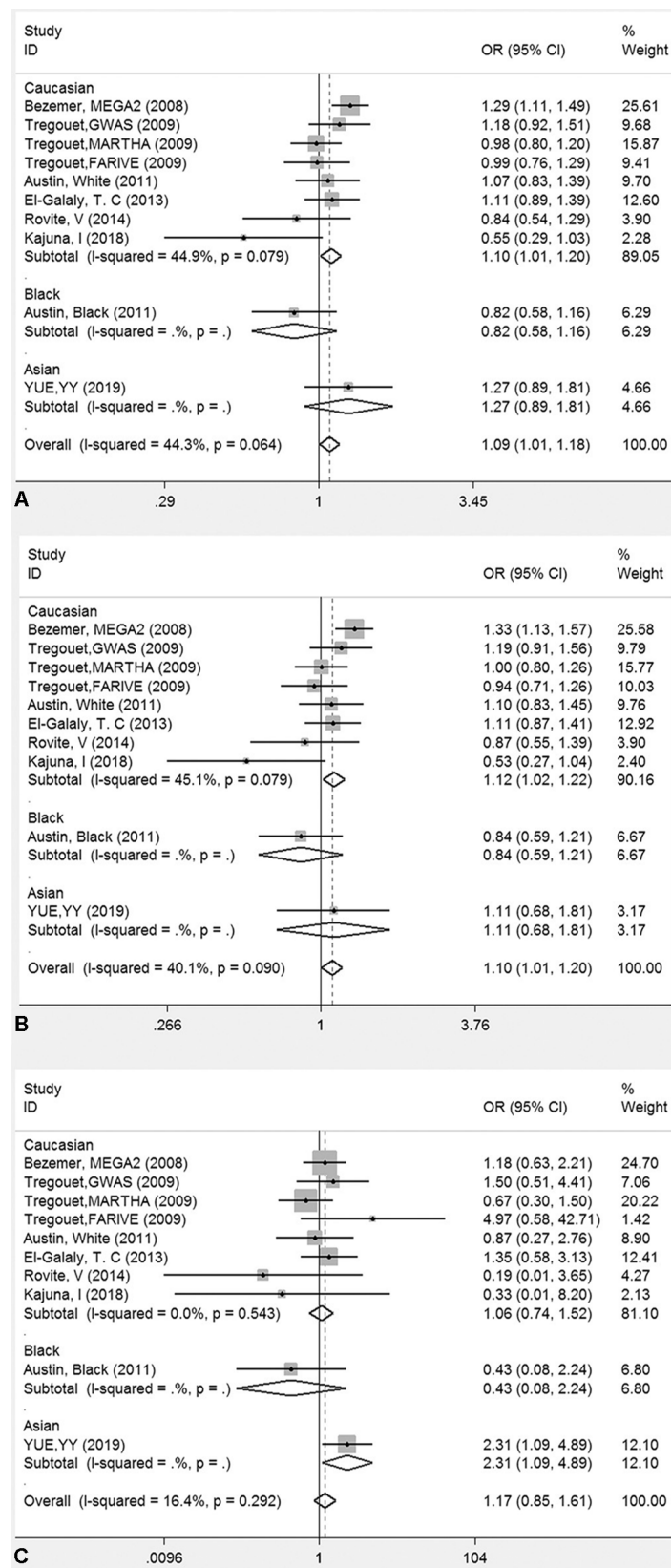
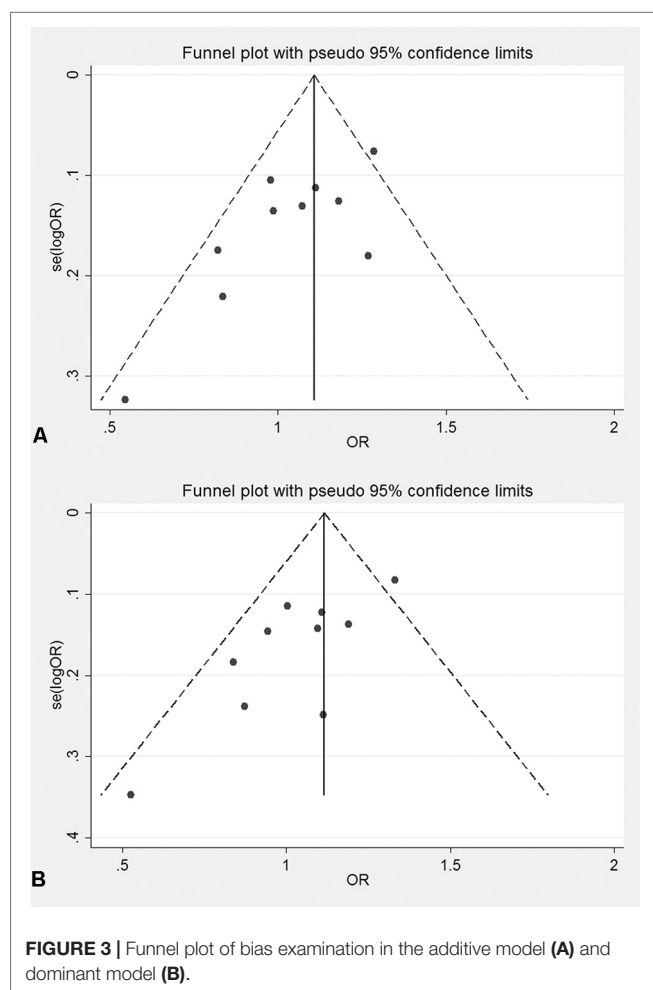


FIGURE 2 | Forest plots for the association between rs2227589 and the risk of venous thromboembolism among different populations **(A)**, additive model; **(B)** dominant model; **(C)** recessive model).



Inherited antithrombin deficiency is an autosomal dominant thrombotic disorder associated with potential risk factors for the development of DVT. Antithrombin is a plasma serine protease inhibitor that can progressively inactivate thrombin, FIIa, and FXa anticoagulation functions. A functional study of rare variations in *PROC*, *PROS1*, and *SERPINC1* genes or other coagulation factors suggests that rare variants may cause inherited deficiencies in the anticoagulant system. Previous studies in an Asian population reported that the incidence of antithrombin deficiency in VTE was around 5.9%–9.61% (Liu et al., 1994; Zheng et al., 2009). Most VTE subjects with antithrombin deficiency have been found to carry genetic variants of *SERPINC1* (Zeng et al., 2015; Mulder et al., 2017). However, the functional mechanism of variant rs2227589 underlying the levels of antithrombin in plasma remains largely unknown (Anton et al., 2009).

Our studies showed that the incidence of antithrombin deficiency is 26.7% in PTE. In our study, the definition of antithrombin deficiency was based on our established antithrombin deficiency normal range (83–128%). The time point of the sample collection, the definition of antithrombin deficiency, and the different selection of patients are the several factors that may affect the frequency of antithrombin

deficiency (Bucciarelli et al., 2012; Di Minno et al., 2014). The high prevalence of antithrombin deficiency in our study may be caused by these factors. The minor allele frequencies of the variant rs2227589 is 0.329 (gnomAD) in the East Asian population but not reported in the Chinese VTE population before. Our genotype data showed that the frequency of rs2227589 was also two times higher than that found in other studies. However, correlation analysis and comparison study of plasma antithrombin activity levels and genotype showed no significant association. The limitation of our study was the relatively small sample size because of the limited sources for us to recruit more patients. More patients recruited from multicenters may consolidate our findings and draw a more reliable conclusion. A study in a Spanish Caucasian population showed that healthy carriers of the rs2227589 SNP T allele had slightly but significantly lower anticoagulant activity with low antithrombin levels (Anton et al., 2009). Moreover, rs2227589 with haplotype in the *SERPINC1* gene of *FV Leiden* carriers showed significantly decreased antithrombin levels (Segers et al., 2014). However, another familial study showed that both rs2227589 carriers and noncarriers have low antithrombin levels (Bhakuni et al., 2015). Thus, variations located in the coding regions of *SERPINC1* but not the rs2227589 (C > T) located in the intron may cause antithrombin deficiency. Thus, whether there is an association between rs2227589 and antithrombin levels requires further investigation. The functional consequences of *SERPINC1* rs2227589 polymorphism might be directly caused by the regulatory effects of this genetic change or caused by other genetic variants linked to the rs2227589 polymorphism. Nevertheless, our study showed consistency with previous studies in that T allele carriers of VTE had slightly lower antithrombin levels, although this was not significant.

Our meta-analysis showed a significant association between rs2227589 and the risk of VTE in additive and dominant genetic models. Indeed, in our initial recruitment, we have recruited 225 patients. Our study showed that the genotyping distributions of rs2227589 in Chinese VTE and PTE groups are similar. When we included all of the 225 patients for meta-analysis, we found that the outcomes of meta-analysis (see **Supplemental Tables 2, 3** and figures for the reanalyzed data) were consistent with the ones using 101 patients for analysis, suggesting the reliability of our findings. Genome-wide and familial case association studies have shown that numerous common or rare variants contribute to the risk of VTE (Smith et al., 2009). For common variants, a GWAS study already identified polymorphisms associated with the risk of DVT, such as *CYP4V2* (rs13146272), *SERPINC1* (rs2227589), and *GP6* (rs1613662) (Bezemer et al., 2008; Tregouet et al., 2009). Our study showed that carriers with low frequent TT genotype had a much higher risk of VTE based on the recessive model. The first GWAS reported by Bezemer et al. showed that carriers with the low-frequency T allele have a modest thrombotic risk in LETS, MEGA-1, and MEGA-2 case-control studies (Bezemer et al., 2008). This finding was later confirmed in the study by Austin et al. (2011). However,

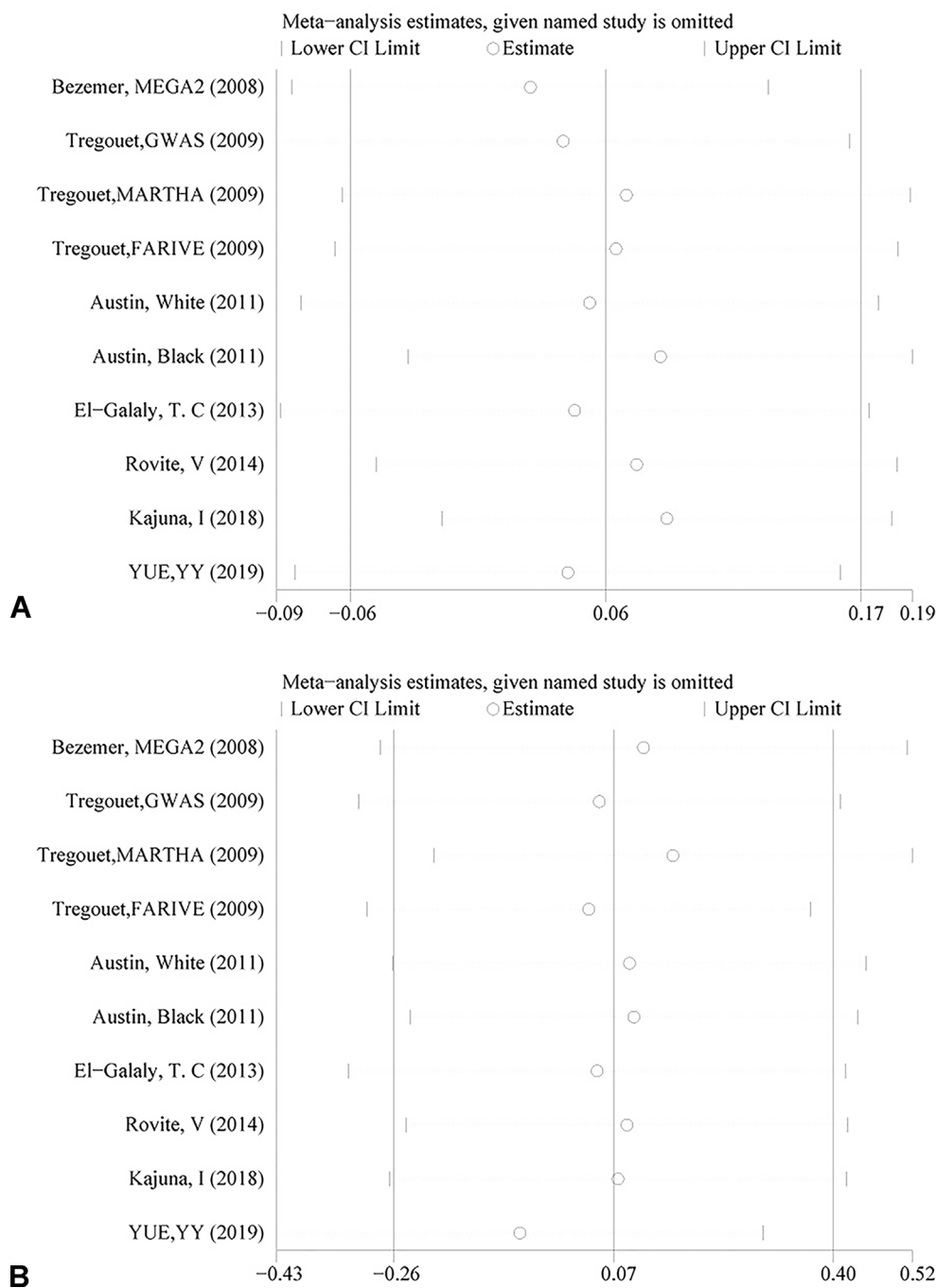


FIGURE 4 | Sensitivity analysis of rs2227589 and the risk of venous thromboembolism (VTE) in the additive model (A) and dominant model (B).

other studies failed to find any association between rs2227589 (*SERPINC1*) and VTE (Tregouet et al., 2009; Dahm et al., 2012; El-Galaly et al., 2013; Bruzelius et al., 2015; Heit et al., 2017). Thus, whether rs2227589 plays any role in the risk of VTE remains controversial.

In conclusion, our study showed that variant rs2227589 is associated with an increased risk of PTE in a Chinese population. However, there appears to be no correlation

with antithrombin deficiency in PTE. Further validation by employing a larger sample size and multiple-center studies may provide a more reliable conclusion of the association. Pooled systematic analysis showed a significant association between rs2227589 and the risk of VTE in the additive and dominant genetic models of multiple populations. As the contribution of rs2227589 to the risk of VTE may vary with different PTE ethnicities, further investigations into the genetic risk factors

and phenotype interactions would be needed to improve the prognosis, prevention, and genetic counseling for populations at high risk of VTE.

ETHICS STATEMENT

The project was approved by the ethics committee of the Ethics Committee of Shenzhen People's Hospital. All procedures performed in studies involving human participants were in accordance with the 1964 Declaration of Helsinki ethical standards. Written informed consent was obtained from the patients for the publication of the patient's identifiable information.

AUTHOR CONTRIBUTIONS

YY and YF prepared the project proposal and study design. YY and QS analyzed all of the genotyping data and conducted the statistical analysis. LX and SL conducted sample collection. QH, MW, and MH conducted antithrombin activity detection experiment. MY assisted with the preparation and revision of the manuscript. All of the authors have read and approved the final manuscript.

REFERENCES

- Anton, A. I., Teruel, R., Corral, J., Minano, A., Martinez-Martinez, I., Ordonez, A., et al. (2009). Functional consequences of the prothrombotic SERPINC1 rs2227589 polymorphism on antithrombin levels. *Haematologica* 94 (4), 589–592. doi: 10.3324/haematol.2008.000604
- Arruda, V. R., von Zuben, P. M., Chiapari, L. C., Annichino-Bizzacchi, J. M., and Costa, F. F. (1997). The mutation Ala677→Val in the methylene tetrahydrofolate reductase gene: a risk factor for arterial disease and venous thrombosis. *Thromb. Haemost.* 77 (5), 818–821. doi: 10.1055/s-0038-1656059
- Austin, H., De Staercke, C., Lally, C., Bezemer, I. D., Rosendaal, F. R., and Hooper, W. C. (2011). New gene variants associated with venous thrombosis: a replication study in White and Black Americans. *J. Thromb. Haemost.* 9 (3), 489–495. doi: 10.1111/j.1538-7836.2011.04185.x
- Beckman, M. G., Hooper, W. C., Critchley, S. E., and Ortel, T. L. (2010). Venous thromboembolism: a public health concern. *Am. J. Prev. Med.* 38 (4 Suppl), S495–S501. doi: 10.1016/j.amepre.2009.12.017
- Bezemer, I. D., Bare, L. A., Doggen, C. J., Arellano, A. R., Tong, C., Rowland, C. M., et al. (2008). Gene variants associated with deep vein thrombosis. *JAMA* 299 (11), 1306–1314. doi: 10.1001/jama.299.11.1306
- Bhakuni, T., Sharma, A., Rashid, Q., Kapil, C., Saxena, R., Mahapatra, M., et al. (2015). Antithrombin III deficiency in Indian patients with deep vein thrombosis: identification of first India based AT variants including a novel point mutation (T280A) that leads to aggregation. *PLoS One* 10 (3), e0121889. doi: 10.1371/journal.pone.0121889
- Bruzeli, M., Bottai, M., Sabater-Lleal, M., Strawbridge, R. J., Bergendal, A., Silveira, A., et al. (2015). Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J. Thromb. Haemost.* 13 (2), 219–227. doi: 10.1111/jth.12808
- Bucciarelli, P., Passamonti, S. M., Biguzzi, E., Gianniello, F., Franchi, F., Mannucci, P. M., et al. (2012). Low borderline plasma levels of antithrombin, protein C and protein S are risk factors for venous thromboembolism. *J. Thromb. Haemost.* 10 (9), 1783–1791. doi: 10.1111/j.1538-7836.2012.04858.x
- Corral, J., Hernandez-Espinosa, D., Soria, J. M., Gonzalez-Conejero, R., Ordonez, A., Gonzalez-Porras, J. R., et al. (2007). Antithrombin Cambridge II

FUNDING

The study was supported by the National Natural Science Foundation of China (21807072), the Guangdong Provincial Natural Science Foundation (2018A030310674), the Shenzhen Science and Technology Project (JCYJ20170413093032806, JSGG20170414104216477), and the Guangdong Provincial Science and Technology Project (2017A020214016).

ACKNOWLEDGMENTS

We would like to express our thanks to the laboratory support from the Key Laboratory of Shenzhen Respiratory Diseases and Shenzhen Public Service Platform on Tumor Precision Medicine and Molecular Diagnosis. Thanks to the abstract publication of this study in the special supplement of journal - *Respirology* by the 23rd Congress of the Asian Pacific Society of Respirology (APSR 2018).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00844/full#supplementary-material>

- (A384S): an underestimated genetic risk factor for venous thrombosis. *Blood* 109 (10), 4258–4263. doi: 10.1182/blood-2006-08-040774
- Dahm, A. E., Bezemer, I. D., Bergrem, A., Jacobsen, A. F., Jacobsen, E. M., Skretting, G., et al. (2012). Candidate gene polymorphisms and the risk for pregnancy-related venous thrombosis. *Br. J. Haematol.* 157 (6), 753–761. doi: 10.1111/j.1365-2141.2012.09121.x
- Di Minno, M. N., Dentali, F., Lupoli, R., and Ageno, W. (2014). Mild antithrombin deficiency and risk of recurrent venous thromboembolism: a prospective cohort study. *Circulation* 129 (4), 497–503. doi: 10.1161/CIRCULATIONAHA.113.003756
- Duval, S., and Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56 (2), 455–463. doi: 10.1111/j.0006-341X.2000.00455.x
- Egger, M., Davey Smith, G., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315 (7109), 629–634. doi: 10.1136/bmj.315.7109.629
- El-Galaly, T. C., Severinsen, M. T., Overvad, K., Steffensen, R., Vistisen, A. K., Tjønneland, A., et al. (2013). Single nucleotide polymorphisms and the risk of venous thrombosis: results from a Danish case-cohort study. *Br. J. Haematol.* 160 (6), 838–841. doi: 10.1111/bjh.12132
- Garcia de Frutos, P., Fuentes-Prior, P., Hurtado, B., and Sala, N. (2007). Molecular basis of protein S deficiency. *Thromb. Haemost.* 98 (3), 543–556. doi: 10.1160/TH07-03-0199
- Heit, J. A. (2015). Epidemiology of venous thromboembolism. *Nat. Rev. Cardiol.* 12 (8), 464–474. doi: 10.1038/nrcardio.2015.83
- Heit, J. A., Armasu, S. M., McCauley, B. M., Kullo, I. J., Sicotte, H., Pathak, J., et al. (2017). Identification of unique venous thromboembolism-susceptibility variants in African-Americans. *Thromb. Haemost.* 117 (4), 758–768. doi: 10.1160/TH16-08-0652
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ* 327 (7414), 557–560. doi: 10.1136/bmj.327.7414.557
- Holzhauser, S., Goldenberg, N. A., Junker, R., Heller, C., Stoll, M., Manner, D., et al. (2012). Inherited thrombophilia in children with venous thromboembolism and the familial risk of thromboembolism: an observational study. *Blood* 120 (7), 1510–1515. doi: 10.1182/blood-2012-01-405514

- Hotoleanu, C. (2017). Genetic Risk Factors in Venous Thromboembolism. *Adv. Exp. Med. Biol.* 906, 253–272. doi: 10.1007/5584_2016_120
- Jiang, J., Liu, K., Zou, J., Ma, H., Yang, H., Zhang, X., et al. (2017). Associations between polymorphisms in coagulation-related genes and venous thromboembolism: a meta-analysis with trial sequential analysis. *Medicine (Baltimore)* 96 (13), e6537. doi: 10.1097/MD.00000000000006537
- Kajuna, I., Stanevica, V., and Mikazane, H. (2018). Single nucleotide polymorphism in patients with vein thrombosis in the population of Latvia. *J. Surg.* 14 (2), 3. doi: 10.7438/1584-9341-14-2-3
- Kim, H. J., Seo, J. Y., Lee, K. O., Bang, S. H., Lee, S. T., Ki, C. S., et al. (2014). Distinct frequencies and mutation spectrums of genetic thrombophilia in Korea in comparison with other Asian countries both in patients with thromboembolism and in the general population. *Haematologica* 99 (3), 561–569. doi: 10.3324/haematol.2013.092023
- Konstantinides, S. V. (2014). 2014 ESC Guidelines on the diagnosis and management of acute pulmonary embolism. *Eur. Heart J.* 35 (45), 3145–3146. doi: 10.1093/eurheartj/ehu393
- Lee, E. J., Dykas, D. J., Leavitt, A. D., Camire, R. M., Ebberink, E., Garcia de Frutos, P., et al. (2017). Whole-exome sequencing in evaluation of patients with venous thromboembolism. *Blood Adv.* 1 (16), 1224–1237. doi: 10.1182/bloodadvances.2017005249
- Lewis, C. M. (2002). Genetic association studies: design, analysis and interpretation. *Brief Bioinform.* 3 (2), 146–153. doi: 10.1093/bib/3.2.146
- Liu, H. W., Kwong, Y. L., Bourke, C., Lam, C. K., Lie, A. K., Wei, D., et al. (1994). High incidence of thrombophilia detected in Chinese patients with venous thrombosis. *Thromb. Haemost.* 71 (4), 416–419. doi: 10.1055/s-0038-1642452
- Luxembourg, B., Pavlova, A., Geisen, C., Spannagl, M., Bergmann, F., Krause, M., et al. (2014). Impact of the type of SERPINC1 mutation and subtype of antithrombin deficiency on the thrombotic phenotype in hereditary antithrombin deficiency. *Thromb. Haemost.* 111 (2), 249–257. doi: 10.1160/TH13-05-0402
- Mili, F. D., Hooper, W. C., Lally, C., and Austin, H. (2011). The impact of co-morbid conditions on family history of venous thromboembolism in Whites and Blacks. *Thromb. Res.* 127 (4), 309–316. doi: 10.1016/j.thromres.2010.12.012
- Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6 (7), e1000097. doi: 10.1371/journal.pmed.1000097
- Mulder, R., Croles, F. N., Mulder, A. B., Huntington, J. A., Meijer, K., and Lukens, M. V. (2017). SERPINC1 gene mutations in antithrombin deficiency. *Br. J. Haematol.* 178 (2), 279–285. doi: 10.1111/bjh.14658
- Navarro-Fernandez, J., de la Morena-Barrio, M. E., Padilla, J., Minano, A., Bohdan, N., Aguila, S., et al. (2016). Antithrombin Dublin (p.Val30Glu): a relatively common variant with moderate thrombosis risk of causing transient antithrombin deficiency. *Thromb. Haemost.* 116 (1), 146–154. doi: 10.1160/TH15-11-0871
- Picard, V., Dautzenberg, M. D., Villoutreix, B. O., Orliaguet, G., Alhenc-Gelas, M., and Aiach, M. (2003). Antithrombin Phe229Leu: a new homozygous variant leading to spontaneous antithrombin polymerization in vivo associated with severe childhood thrombosis. *Blood* 102 (3), 919–925. doi: 10.1182/blood-2002-11-3391
- Rosenberg, R. D., and Bauer, K. A. (1987). Thrombosis in inherited deficiencies of antithrombin, protein C, and protein S. *Hum. Pathol.* 18 (3), 253–262. doi: 10.1016/S0046-8177(87)80008-4
- Rosendaal, F. R., and Reitsma, P. H. (2009). Genetics of venous thrombosis. *J. Thromb. Haemost.* 7 Suppl 1, 301–304. doi: 10.1111/j.1538-7836.2009.03394.x
- Segers, O., Simioni, P., Tormene, D., and Castoldi, E. (2014). Influence of single nucleotide polymorphisms on thrombin generation in factor V Leiden heterozygotes. *Thromb. Haemost.* 111 (3), 438–446. doi: 10.1160/TH13-05-0360
- Smith, N. L., Wiggins, K. L., Reiner, A. P., Lange, L. A., Cushman, M., Heckbert, S. R., et al. (2009). Replication of findings on the association of genetic variation in 24 hemostasis genes and risk of incident venous thrombosis. *J. Thromb. Haemost.* 7 (10), 1743–1746. doi: 10.1111/j.1538-7836.2009.03567.x
- Souto, J. C., Almasy, L., Borrell, M., Blanco-Vaca, F., Mateo, J., Soria, J. M., et al. (2000). Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. Genetic Analysis of Idiopathic Thrombophilia. *Am. J. Hum. Genet.* 67 (6), 1452–1459. doi: 10.1086/316903
- Sterne, J. A., Egger, M., and Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *BMJ* 323 (7304), 101–105. doi: 10.1136/bmj.323.7304.101
- Tregouet, D. A., Heath, S., Saut, N., Biron-Andreani, C., Schved, J. F., Pernod, G., et al. (2009). Common susceptibility alleles are unlikely to contribute as strongly as the FV and ABO loci to VTE risk: results from a GWAS approach. *Blood* 113 (21), 5298–5303. doi: 10.1182/blood-2008-11-190389
- Yue, Y., Sun, Q., Man, C., and Fu, Y. (2019). Association of the CYP4V2 polymorphism rs13146272 with venous thromboembolism in a Chinese population. *Clin. Exp. Med.* 19 (1), 159–166. doi: 10.1007/s10238-018-0529-y
- Zeng, W., Tang, L., Jian, X. R., Li, Y. Q., Guo, T., Wang, Q. Y., et al. (2015). Genetic analysis should be included in clinical practice when screening for antithrombin deficiency. *Thromb. Haemost.* 113 (2), 262–271. doi: 10.1160/TH14-05-0446
- Zhang, Z., Lei, J., Shao, X., Dong, F., Wang, J., Wang, D., et al. (2019). Trends in hospitalization and in-hospital mortality from VTE, 2007 to 2016, in China. *Chest* 155 (2), 342–353. doi: 10.1016/j.chest.2018.10.040
- Zheng, C. C., Wu, J. S., Ding, K. Y., Li, J., Hu, H. J., Ding, X. L., et al. (2009). [Clinical significance of anticoagulant proteins detection in patients with thrombotic events]. *Zhonghua Xue Ye Xue Za Zhi* 30 (4), 264–267. doi: 10.3760/cma.j.issn.0253-2727.2009.04.014
- Zhu, T., Ding, Q., Bai, X., Wang, X., Kaguelidou, F., Alberti, C., et al. (2011). Normal ranges and genetic variants of antithrombin, protein C and protein S in the general Chinese population. results of the Chinese hemostasis investigation on natural anticoagulants study I group. *Haematologica* 96 (7), 1033–1040. doi: 10.3324/haematol.2010.037515

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Yue, Sun, Xiao, Liu, Huang, Wang, Huo, Yang and Fu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Systematically Analyzing the Pathogenic Variations for Acute Intermittent Porphyrria

Yibao Fu¹, Jinmeng Jia¹, Lishu Yue¹, Ruiying Yang¹, Yongli Guo^{2,3,4*}, Xin Ni^{2,3,4*} and Tielu Shi^{1,5*}

¹ Center for Bioinformatics and Computational Biology, and the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, ² Big Data and Engineering Research Center, Beijing Key Laboratory for Pediatric Diseases of Otolaryngology, Head and Neck Surgery, MOE Key Laboratory of Major Diseases in Children, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China, ³ Biobank for Clinical Data and Samples in Pediatrics, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China, ⁴ Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, National Center for Children's Health, Capital Medical University, Beijing, China, ⁵ National Center for International Research of Biological Targeting Diagnosis and Therapy, Guangxi Key Laboratory of Biological Targeting Diagnosis and Therapy Research, Collaborative Innovation Center for Targeting Tumor Diagnosis and Therapy, Guangxi Medical University, Nanning, China

OPEN ACCESS

Edited by:

Mike Mikailov,
United States Food and Drug
Administration, United States

Reviewed by:

Olimpia Musumeci,
University of Messina, Italy
Nazareno Paolucci,
Johns Hopkins University,
United States

*Correspondence:

Yongli Guo
yongliguo@bch.com.cn
Xin Ni
nixin@bch.com.cn
Tielu Shi
tlshi@bio.ecnu.edu.cn

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 16 November 2018

Accepted: 09 August 2019

Published: 13 September 2019

Citation:

Fu Y, Jia J, Yue L, Yang R, Guo Y,
Ni X and Shi T (2019) Systematically
Analyzing the Pathogenic Variations
for Acute Intermittent Porphyrria.
Front. Pharmacol. 10:1018.
doi: 10.3389/fphar.2019.01018

The rare autosomal dominant disorder acute intermittent porphyria (AIP) is caused by the deficient activity of hydroxymethylbilane synthase (HMBS). The symptoms of AIP are acute neurovisceral attacks which are induced by the dysfunction of heme biosynthesis. To better interpret the underlying mechanism of clinical phenotypes, we collected 117 *HMBS* gene mutations from reported individuals with AIP and evaluated the mutations' impacts on the corresponding protein structure and function. We found that several mutations with most severe clinical symptoms are located at dipyrromethane cofactor (DPM) binding domain of HMBS. Mutations on these residues likely significantly influence the catalytic reaction. To infer new pathogenic mutations, we evaluated the pathogenicity for all the possible missense mutations of *HMBS* gene with different bioinformatic prediction algorithms, and identified 34 mutations with serious pathogenicity and low allele frequency. In addition, we found that gene *PPARA* may also play an important role in the mechanisms of AIP attacks. Our analysis about the distribution frequencies of the 23 variations revealed different distribution patterns among eight ethnic populations, which could help to explain the genetic basis that may contribute to population disparities in AIP prevalence. Our systematic analysis provides a better understanding for this disease and helps for the diagnosis and treatment of AIP.

Keywords: acute intermittent porphyria, *HMBS* gene, genotype and phenotype relationship, hypergeometric test, variation ethnic distribution difference, *PPARA* gene

INTRODUCTION

Acute intermittent porphyria (AIP) is a rare autosomal dominant disorder caused by the deficient activity of hydroxymethylbilane synthase (HMBS), which is also referred as porphobilinogen deaminase (PBGD), the third enzyme in the pathway of heme biosynthesis (Gill et al., 2009; Chen et al., 2018). Clinically, most of the symptomatic AIP patients are women (Pulgar et al.,

2019), especially in their reproductive age (Innala et al., 2010). It is commonly recognized that sex steroids play an important role for the clinical manifestations in porphyria in women and they act as inducers in heme biosynthesis (Innala et al., 2010). The clinical symptoms of AIP patients include acute recurrent abdominal pain, and often are accompanied by gastrointestinal disorders such as nausea, vomiting and constipation (Yang et al., 2015; Yang et al., 2016; Duque-Serrano et al., 2018). In addition, hypertension, tachycardia, hyponatremia and motor weakness are often present (Puy et al., 2010; Kong et al., 2013). Mental changes include anxiety, depression, agitation, etc. (To-Figueras et al., 2006; Duque-Serrano et al., 2018). Most of the clinical features of an attack are caused by the effects on the nervous system and the characteristics of the symptoms are not so obvious (Stein et al., 2017). This is one of the reasons why it can easily be misdiagnosed. In many cases, patients suffer from frequent convulsions and seizures. But patients with severe psychiatric symptoms such as psychosis, hallucinations and delirium are rare. Due to the similarity to other disease symptoms, it is a challenge for clinicians to diagnose patients with intermittent porphyria at their first attack. Porphyrin levels in urine and blood is a useful diagnostic index for patients with a suspected acute attack of AIP.

Heme is mostly synthesized in erythropoietic cells and liver cells, and plays an essential role in the synthesis of hemoproteins such as hemoglobin, cytochromes, myoglobin, catalase, and peroxidase, all of which are important for the transportation of oxygen and the oxidation–reduction reactions (Karim et al., 2015). HMBS is the third enzyme in the heme biosynthesis pathway, in which the synthesis of ALA (5-aminolaevulinic acid) is one of the most important controlling steps for heme formation. ALAS (5-aminolaevulinic acid synthase) is the enzyme that controls the first step in the heme biosynthesis and encoded by two different genes: *ALAS1* (ubiquitously expressed in every human cell) and *ALAS2* (only expressed in erythroid) (Puy et al., 2010). *ALAS1* acts as the rate limiting enzyme in the heme synthesis pathway in the liver and can be controlled by heme through negative feedback regulation loop, as heme down-regulates the transcription of *ALAS1* (Ajioka et al., 2006). The partial deficiency of HMBS activity can hinder the heme synthesis and the negative regulation mechanism could lead to the excessive accumulation of heme precursors such as ALA and PBG in tissues, which may trigger acute attacks (Tracy and Dyck, 2014).

Attacks of AIP can be induced by many different factors. These factors contribute to the attacks by inducing *ALAS1* transcription or the activity of ALA synthase, either directly or indirectly (Figure 1I) (Stein et al., 2017). Some triggers increase the demand for heme in the liver which de-represses the transcription of *ALAS1* thereafter (Ajioka et al., 2006). Others, like hormones estrogen and progesterone, increase the activity of ALA synthase (Figure 1IV), which partially explains why AIP often attacks during the luteal phase of the menstrual cycle (Junior et al., 2017). Up to now, the identified factors that induce the attacks include alcohol, smoking, nutritional factors, hormonal factors, and the usage of drugs etc. (Figures 1III, VI, III, V) (Handschin et al., 2005; Windebank et al., 2005; Tracy and Dyck, 2014; Karim et al., 2015). All of these factors affect the heme production pathway,

and then lead to the over-accumulation of toxic heme precursors (Figure 1).

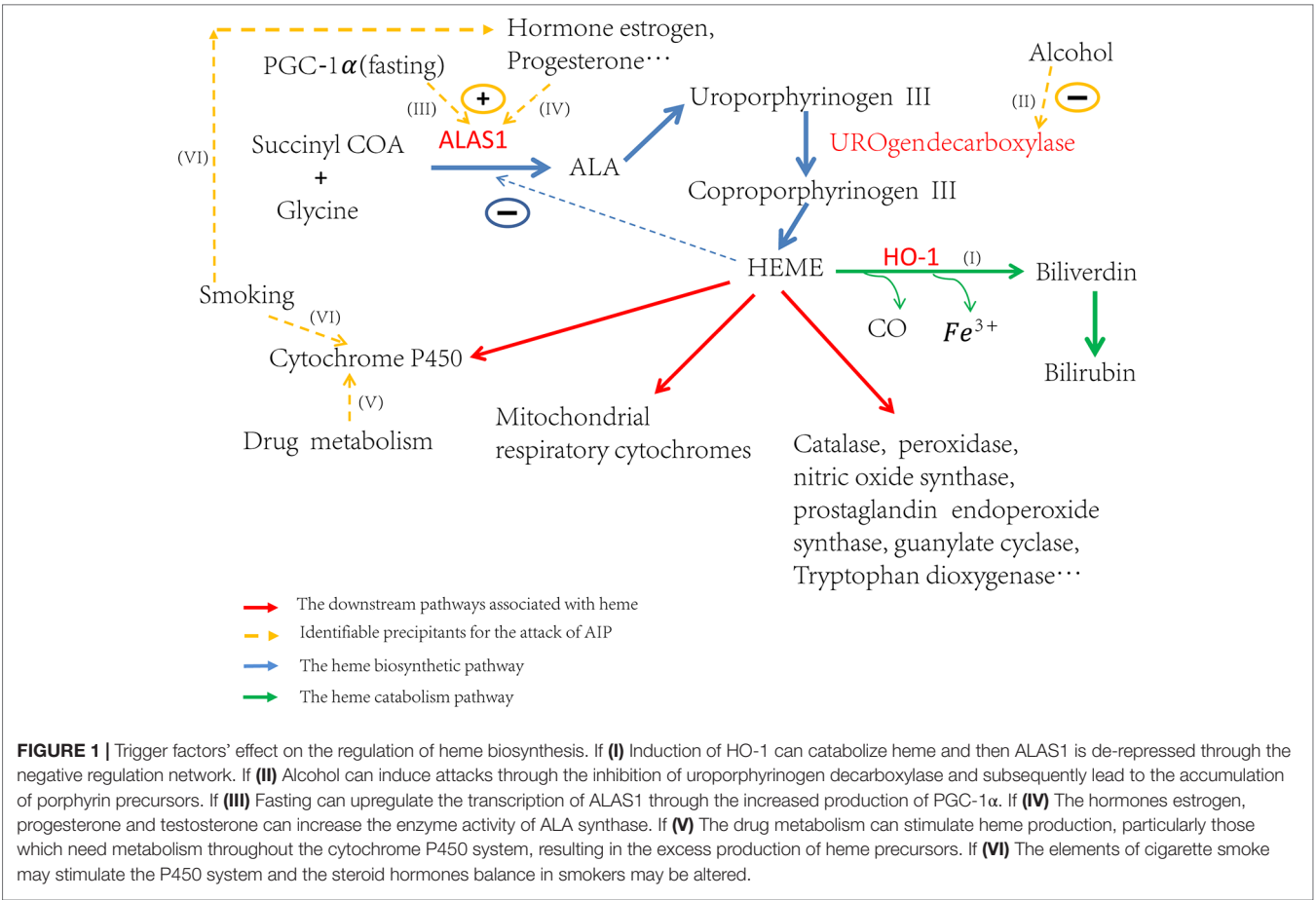
The mechanism of psychiatric symptoms in AIP is not fully understood. The current researches suggest that the accumulation of ALA is the prime culprit in the damage to nervous system. One of the hypotheses is the structural similarity between ALA and γ -aminobutyric acid (GABA), the accumulation of ALA may affect normal GABA function in the nervous system (Windebank and McDonald, 2005; Tracy and Dyck, 2014). The excess of ALA leads to the pronounced HO (heme oxygenase) activity, resulting in deregulation of the cholinergic system, increasing oxidative stress, affecting activity and expression of NOS (nitric oxide synthases) (Lavandera et al., 2016), decreasing GABAergic neurons' activity (Brennan and Cantrill, 1979) and increasing glutamate release (Satoh et al., 2008; Duque-Serrano et al., 2018).

With the rapid development of next generation sequencing (NGS) technology (Gill et al., 2009; Shen, 2018), more and more disease related mutations have been detected. Up to now, the Human Gene Mutation Database (HGMD) (<http://www.hgmd.cf.ac.uk/ac/>) has collected a total of 421 different mutations on gene *HMBS*, including missense/nonsense, splicing, small deletions, small insertions, small indels, and gross deletions, etc. Patients with different mutations present various severity of clinical symptoms. However, many mutations' effect on the protein structure and their relationship with phenotype are not fully explored. For example, there are cases that patients' family members with the same mutation are asymptomatic (Li et al., 2015). Also, patients with the same mutation exhibit clinical manifestation of various severity. Therefore, to obtain a comprehensive landscape of the effects of different mutations, we conducted analysis on the effects of mutations based on the second structure and the three-dimensional structure of corresponding proteins and tried to infer the relationships between phenotypes and genotypes. Considering the complexity of metabolic processes in human body, we supposed that other genes which interact with genes or proteins in the heme synthesis pathway may also play roles in the attack of AIP. To identify new genes that potentially contribute to this disease, we conducted both protein–protein interaction network and pathway enrichment analyses. In addition, we also explored the genetic difference among different ethnic groups based on reported patients' cases for this disease. We analyzed the risk allelic frequencies among eight different populations world-wide. Our research results provide overview picture for the distribution of AIP mutations and shed light on better understanding of AIP disease.

MATERIALS AND METHODS

Analysis of the Relationship Between Phenotypes and Genotypes

We collected the clinical symptoms and corresponding mutation information of 117 patients from related resources and literature (Table S1). All of those clinical and genetic data have been inputted into eRAM and PedAM systems (Jia et al., 2018a; Jia et al., 2018b). Based on the different clinical symptoms, we classified patients into three categories: severe, moderate, and



mild. Our classification criteria include frequency of attacks, blood pressure, heart rate, mental condition, blood tests result, serum sodium concentration, the extent of abdominal pain, vomiting, and nausea, etc. (Zanella et al., 2007). Then we analyzed the corresponding missense mutations in each phenotype class.

In order to understand the effect of these missense mutations on HMBS protein structure, we mapped those mutations to the crystal structure (PDB: 5M7F, Resolution: 2.78-Å) (Pluta et al., 2018) to study how the mutations affect the structure of the protein and then evaluated the relationship between mutations and clinical symptoms. For the visualization of the three-dimensional structure of protein, we used a web tool (<http://www.sbg.bio.ic.ac.uk/~ezmol/>). We also studied the interactions between different amino acid residues using the Residue Interaction Network Generator (RING) (<http://protein.bio.unipd.it/ring/>) and interpreted the interrelation between different residues, including 5M7F (Human porphobilinogen deaminase in complex with DPM cofactor) and 5M6R (Human porphobilinogen deaminase in complex with reaction intermediate).

To further explore the mutation' effect on the HMBS protein structure, we performed multi-sequence alignment for those homologous protein sequences of different species extracted from the UniGene database (<https://www.ncbi.nlm.nih.gov/>

unigene/?term=HMBS) with bioEdit software and inspected the conservation of amino acids (Mahdavi et al., 2018).

To identify the new pathogenic variations, we obtained the FASTA file of gene *HMBS* from Ensembl database (GRCh37:CM000673.1), and then listed all the missense variations with the method of enumeration (Table S2). The analysis of the pathogenic effects of these missense variations on *HMBS* gene was conducted by using five prediction algorithms: SIFT, PolyPhen 2 HDIV, PolyPhen 2 Hvar, CADD, GERP ++ (Table 1). At first, 579 missense variations which met the selection criteria were identified. The cutoff values for selecting the possible pathogenic mutations were set larger than the average of predicted score of collected mutations to ensure the

TABLE 1 | Criteria for selecting predicted Pathogenic SNPs.

Protein prediction algorithm	Cutoff-Value	Interpretation
SIFT	D	D:Deleterious(sift< = 0.05)
PolyPhen 2 HDIV	D	D:Probably damaging(> = 0.957)
PolyPhen 2 Hvar	D	D:Probably damaging(> = 0.909)
CADD	Score >25	10 means 10% percentile highest scores; 20 means 1% percentile highest scores; 30 means 0.1% percentile highest scores
GERP++	Score >5	High scores are more deleterious

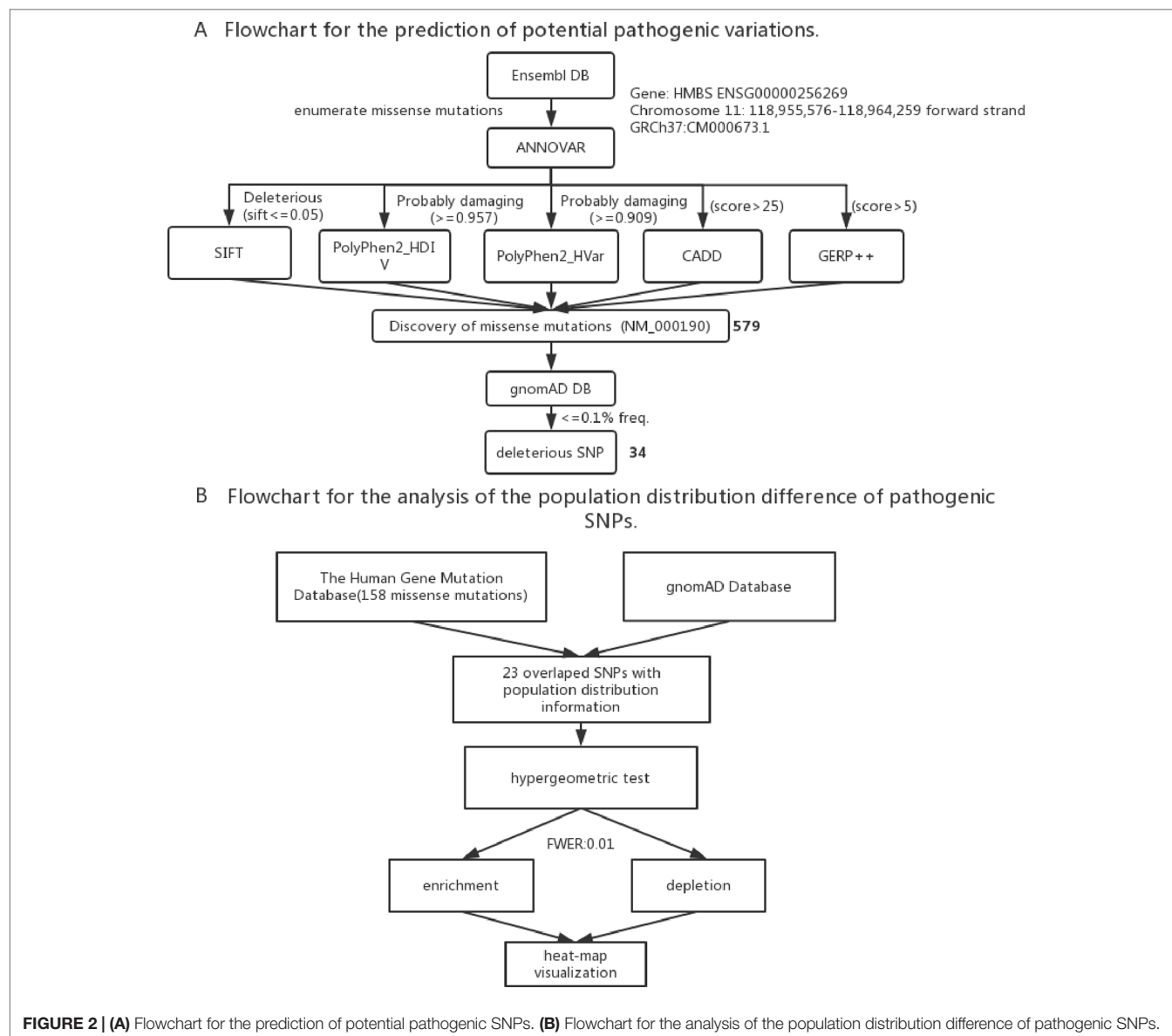


FIGURE 2 | (A) Flowchart for the prediction of potential pathogenic SNPs. **(B)** Flowchart for the analysis of the population distribution difference of pathogenic SNPs.

reliability of prediction results. Next, we selected those missense variations with frequency less than 0.1% based on the gnomAD database (<http://gnomad.broadinstitute.org/>) and obtained 34 deleterious variations (**Figure 2A**) (**Table S3**).

Prediction of New Associated Genes

As the control step of heme biosynthesis pathway, ALAS1 plays a rate-limiting role which is regulated by heme through the negative feedback loop. So we tried to predict new AIP-related pathogenic genes from the perspective of interactions between ALAS1 and other genes. STRING (<https://string-db.org/cgi/input.pl>) is a database for known and predicted protein-protein interactions and presents the physical and functional associations between proteins. inBio Map™ (<https://www.intomics.com/inbio/map.html>)

is another protein-protein interaction platform known for its high coverage and high quality (Li et al., 2017). We used the two databases to identify the possible interacting genes for ALAS1.

To validate our results and explore the underlying mechanisms of those predicted genes, we conducted the protein functional analysis and pathway enrichment analysis using Reactome database (<https://www.reactome.org>), which is a relational database of signaling and metabolic molecules (Fabregat et al., 2018).

To check the gene expression pattern of those interacting proteins, we used the Genotype-Tissue Expression (GTEx) database (<https://gtexportal.org/home/>). Furthermore, to validate the predictive associated gene, we studied their protein or gene function based on UniProt database (<https://www.uniprot.org/>) and published literature.

Distribution Difference of Variations Among Populations

To study the population distribution difference of AIP — associated variations in allele frequency, we used the published 158 missense mutations in *HMBS* gene from HGMD and obtained their population distribution information from gnomAD database, the ethnic groups include South Asian, European (Non-Finnish), African, East Asian, Ashkenazi Jewish, European (Finnish), Latino and other. In the end, there were 23 variations with population distribution information available (Table S4).

To assess if the risk allele of an AIP — associated variations is enriched or depleted (separated) significantly in each of the 8 populations, we performed 16 (2×8) hypergeometric tests for each variation (Mao et al., 2017). If the p-value of enrichment is less than the p-value of depletion, then it is over-represented, but may not significantly over-represented. If the p-value of enrichment is greater than the p-value of depletion, then it is depleted. To control a family-wise error rate (FWER) of 0.01, we used a raw p-value of $0.01/368 = 2.717\text{E-}5$ as a cutoff.

To visualize variation enrichment/depletion patterns in 8 populations, we first transformed the hypergeometric testing p-values by \log_{10} . After that, we used Seaborn package of Python to generate hierarchical clustering heat-map based on enrichment/depletion p-values (\log_{10} based) of risk variations in different populations (Figure 2B). If an AIP-associated variation is enriched in a population, we used the negative of \log_{10} of the enrichment p-value to represent the variation in the cluster heat-map. In contrast, if a variation is depleted in a population, the value of \log_{10} of the depletion p-value was used.

RESULTS

The Relationship Between Genotypes and Phenotypes

Previous studies experimentally characterized two novel mutations by comparing them with wild-type (wt) *HMBS* (Bustad et al., 2013). Our collected clinical data show that patients with mutation R116W and R173W have severe clinical symptoms, which is consistent with previous study that R116W mutation leads to the protein defection in conformational stability while R173W impacts both enzyme kinetics and conformational stability (Bustad et al., 2013). In the residue interaction network of 5M6R and 5M7E, R173 interacts with many other residues, including S146, I166, G168, N169, K176, L177, I186, and L188. Besides, it also interacts with the reaction intermediate (7J8) and dipyrrromethane (DPM) cofactor (Figure 3B). R173 locates at α -helices structure in domain 2 (Figure 3A), mutations on this residue will generate severe defection to the catalytic proteins, which is very similar to the effect of mutation R116W. In addition, residue R116 locates at the hinge-bending region of domains 1 and 2, mutations on it are believed to affect the binding ability of its substrate.

In addition, mutations R149X, Q217H, G218R, A219P and A330P can also lead to severe clinical manifestations. In the residue interaction network, residues R149, Q217, G218 and A219 all have direct interactions with both the reaction intermediate and DPM cofactor (Figures 3C, S1). Residue A330 locates at α -helices structure in domain 3. Although A330 has no direct interaction with the reaction intermediate, the nonsynonymous mutations on it could disrupt the secondary structure of protein and affect other five residues. Mapping the missense mutation to the crystal structure of human *HMBS* enzyme showed that residues Q217, G218, and A219 are all in close proximity to the attachment site of dipyrrromethane (DPM) cofactor (Figure 3A). The substitution of these residues may affect the cofactor's binding which impact the catalytic reaction. Multiple sequence alignment revealed that all these four residues are strongly conserved among many vertebrate species (Figure S2). Taken together, it is reasonable to hypothesize that missense mutations on these residues can lead to severe clinical manifestations.

The corresponding clinical symptoms for mutations R26C, D99N, R167W, G168X, Q194X, and G221D are relatively moderate. Although they interact with the reaction intermediate and DPM cofactor directly in the residue interactive network, their locations on the protein 3D structure reveal that they are relatively away from the active site (Figure S3), and only a few other residues interact with these residues. Therefore, mutations on these residues most likely lead to relatively moderate defection compared to the normal activity of *HMBS*. Patients with mutations T35M, G111R, and L238P present a mild clinical phenotype. Those amino acids locate far from the active site based on the three-dimensional structure and have no direct interaction with the reaction intermediate and DPM cofactor. Besides, they have small number of interacting residues in the residues interacting network. Mutations on these residues may have less effect on the protein structure and these mutations should have relatively low pathogenic impact, which can partially explain why patients carrying those mutations only present mild clinical symptoms.

To identify other potential pathogenic mutations, we conducted the prediction for potential deleterious variations (Figure 2A). First, according to the cutoff-value of five different bioinformatics algorithms, 579 missense mutations were selected as the most deleterious mutations. Then we checked the allele frequencies of all the selected mutations in gnomAD database and excluded those mutations whose frequency are more than 0.1%. As the result, we obtained 34 missense mutations as the potential pathogenic variations. Residues R22, R26, M56, S75, E80, E82, E86, N88, E89, L97, R116, A122, S141, V142, V143, T145, A152, P159, R167, T172, D178, A189, A219, G260, V265, V282, G287, G317, A331, and G346 may have potentially important roles for the normal activity of *HMBS*. Remarkably, mutations on residues R26, R116, R167, and R219 have been reported to be pathogenic and present relatively severe clinical manifestations, which provide the evidence to support our prediction results.

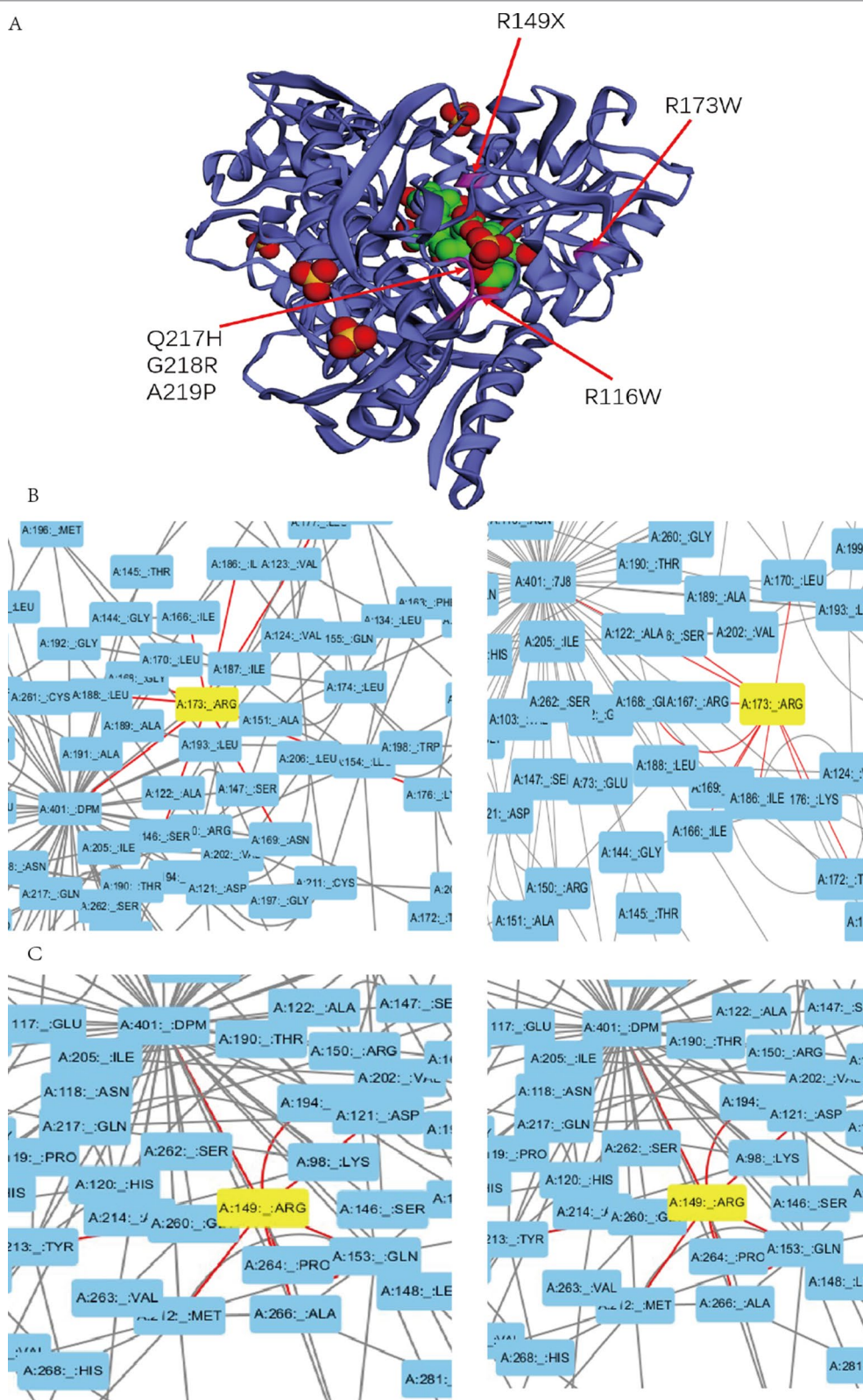


FIGURE 3 | (A) Mutations with severe clinical symptoms are near the binding site for DPM (dipyrromethane) cofactor at active site on the crystal structure of human HMBS enzyme (5M7F). **(B) (C)** Residue A173 and R149 interacts with the DPM cofactor (5M7F) and reaction intermediate (5M6R) at atomic level in the HMBS protein structure by the Residue Interaction Network Generator (RING) software.

In addition, taking residues L97, A122, A145, A189, G260, and V265 as examples, the RING analysis result showed that all these five mutations had connections with DPM cofactor (5M7F) and reaction intermediate (5M6R). Those mutations could affect the normal binding of catalytic cofactor and reaction intermediate, which result in the abnormality of the reaction. Overall, these new identified variations could be the candidates that contribute to the disease.

Prediction for Pathogenic Candidate Genes

There are 10 possible interactive partners for ALAS1 protein from STRING database and 16 from inBio Map™ database. Based on the protein functional analysis, we noticed that gene *PPARA* (peroxisome proliferator-activated receptor alpha) had obvious function overlap with ALAS1.

Heme is an iron porphyrin compound which plays as the prosthetic group of hemoglobin, myoglobin, cytochrome, peroxidase, and catalase. Most drug metabolisms are closely associated with the cytochrome P450 enzymes, so the drugs, particularly those ones metabolized through the cytochrome P450 system, can increase hepatic heme turnover (**Figure IV**) and affect the ALAS1 through the negative feedback regulation loop, thus leading to the excess production of heme precursors. Protein functional analysis results showed that gene *PPARA* had a direct regulation of the transcription of cytochrome P450 gene *CYP2C8* (Thomas et al., 2015) and hepatic cytochrome P450 3A4 (*CYP3A4*) (Thomas et al., 2013). Therefore, mutations on *PPARA* can affect the cytochrome P450 system, and consequently may have an influence on the heme biosynthesis pathway indirectly, resulting in acute of intermittent porphyria.

Pathway enrichment analysis with Reactome showed that *ALAS1* shared three same pathways with *PPARA*, they are metabolism of lipids and lipoproteins, fatty acid triacylglycerol and ketone body metabolism and *PPARA* activates gene expression respectively. One of the clinical symptoms of AIP is hyperlipidemia (HP: 0003077), which is closely associated with the abnormal metabolism of lipids. And it is well known that the activation of the *PPARα* can reduces hyperlipidemia (Kimura et al., 2013). The relationships between the phenotypes of hyperlipidemia and the regulation mechanism of lipid metabolism indicates that gene *PPARA* is likely to be associated with AIP.

Heme is synthesized mainly in erythropoietic cells (80%) and liver parenchymal cells (15%). ALAS1 acts as the rate limiting enzyme in the production of heme in liver and can be regulated by heme. Whereas, in erythroid cells, ALAS2 acts as the catalyzing enzyme in the first step of heme biosynthesis and the synthetic rate is limited by iron availability (Puy et al., 2010). The cytochrome P450 gene *CYP2C8* and *CYP3A4* are highly expressed in liver, which show the same expression pattern as gene *PPARA* (**Figure 4**). Gene *PPARA* can affect the drug metabolism by the direction regulation of transcription of *CYP2C8* and *CYP3A4*, and thus affect the heme biosynthesis pathway. These functional associations between *PPARA* and *ALAS1* suggest that gene *PPARA* may play an

important role in the mechanisms of AIP attacks and could act as the possible pathogenic gene.

The Distribution Difference of Variations Among Populations

We extracted the variation frequency of 23 *HMBS* gene mutations among eight different ethnic groups from gnomAD and assessed if a given variation was significantly enriched or depleted in different ethnic groups by hypergeometric test. The heat-map (**Figure 5**) demonstrated that missense mutation R167W was enriched in Finnish people, D319N was significantly enriched in African and depleted in the rest of regions. Mutation D178N was enriched and R281H was depleted in South Asian compared with the global average, whereas the distribution of the two variations was in the opposite situation in Non-Finnish European. Besides, East Asian, Ashkenazi Jewish, European (Finnish) and Latino exhibit similar allele enrichment/depletion patterns and South Asian, Non-Finnish European and African share overall similar distribution patterns for the 23 mutations.

The distribution frequency of the 23 variations reveals their different patterns among 8 populations. The possible explanation for these results is that the allele abundance may be shaped by different environmental factors and evolution patterns. Because of the small amount of variations used in our analysis, the distribution pattern may not be so obvious. More pathogenic variations could provide better view for the pattern difference among populations.

DISCUSSION

In this study, we have explored the triggers and their mechanisms for acute attacks in AIP. Currently, many attacks have no clearly identifiable triggers. For those suspected patients, close attention should be paid to avoid these precipitants in their daily life. Although the integral and exact mechanisms of nervous system dysfunction are not fully understood, current works have offered basic guidelines to analyze the pathogenesis of variants and provided useful information for the diagnosis and the treatment of the disease.

Based on the collected mutation pattern and clinical information, we analyzed the relationship between genotypes and phenotypes. Previous research has made the conformational stability and activity analysis of several *HMBS* mutations, they interpreted the molecular mechanisms of *HMBS* mutations and inferred the connection between genotypes and phenotypes to certain extent (Bustad et al., 2013). By mapping the mutations to the protein crystal structure and conducting the residues interaction analysis, we have obtained more comprehensive view about how the mutations affect the protein function. Further experiment of molecular biology can verify our interpretation about the specific influence of those mutations on the protein structure. We also used different bioinformatics prediction algorithms to evaluate the pathogenicity of enumerated possible missense mutations. Some of the new predicted deleterious mutations may affect the binding site for the dipyrromethane

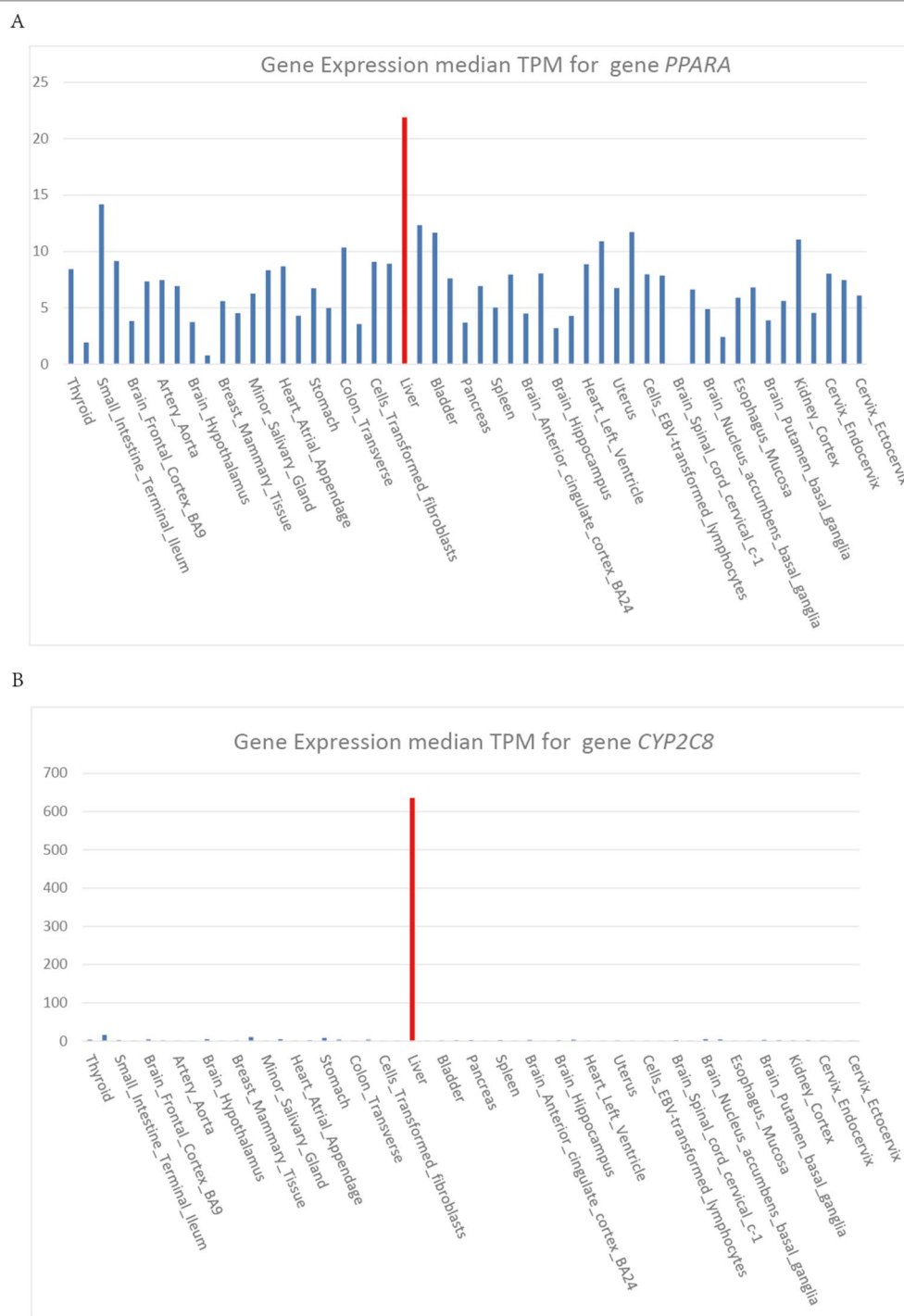


FIGURE 4 | (A) (B) Gene expression median TPM for gene *PPARA* and *CYP2C8*.

cofactor at the active site, which provides new resource to better understand the mutation spectra of the disease.

By protein functional analysis and pathway enrichment analysis, we found that gene *PPARA* was correlated with the attacks of AIP through the direct regulation of the transcription of cytochrome P450 gene *CYP2C8*, which may affect the turnover of heme in the

liver. Besides, in the *PPARA* activating gene expression pathway, we also noticed that *NRF1* and *PPARGC1B* (PGC-1 β) could regulate the expression of *ALAS1*. Experiment on mouse has demonstrated that PGC-1 β (*PPARGC1B*) binds *NRF1* and co-activates genes regulated by *NRF1* (Ivanova et al., 2013). The dysfunction of *NRF1* and *PPARGC1B* may be associated with this disease. Also, gene

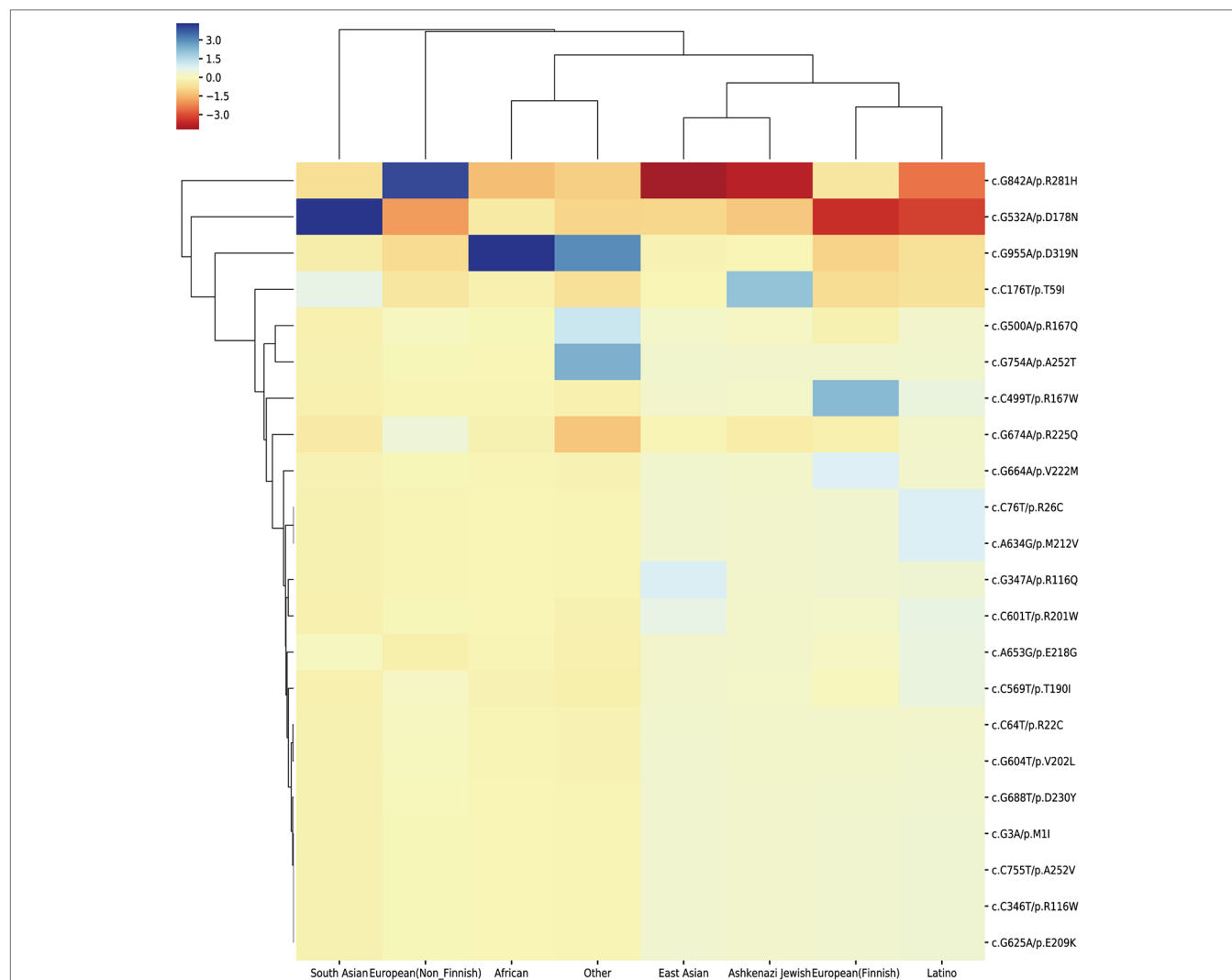


FIGURE 5 | Cluster heat-map shows the significance of the 23 pathogenic variations enriched or depleted in each population compared with the worldwide average. Each row represents a variation, and each column represents a population. A unit in the cluster heat map is colored based on the \log_{10} of p-value.

ESRRA can affect the normal expression of *ALAS1*, indicating that the normal heme biosynthesis pathway may be affected and gene *ESRRA* may be the potential new pathogenic gene which is related to the attacks of AIP.

However, one limitation of our study is that we have only focused on those mutations located around the binding region for DPM at the active site. Other mutations located away from the region could also play an important role in the three-dimensional structure of protein and affect the activity of enzyme. Another limitation of our work is that we only have studied the missense mutations in the coding region of *HMBS* gene. We do not consider other mutation types (frameshift, deletion, insertion and nonsense) and mutations in the non-coding region. In addition, larger amount of variations with population distribution frequency should be collected to have a better view for the allelic distribution difference patterns. Nevertheless, our systematic analysis provides a better

understanding for this disease and help for the diagnosis and treatment of AIP.

In the past few decades, the diagnosis and therapy of acute porphyria have been paid a lot attention by researchers and clinicians around the world (Jia and Shi, 2017). European Union Directorate-General (EU DG) for Health & Consumers have established an effective network (European Porphyria Network) of specialist porphyria centers throughout the EU. Currently, EPNET (European Porphyria Network, <http://porphyria.eu/en/content/acute-porphyria>) consists of 33 EU specialist centers from 21 European and candidate countries. They also have associate members from Australia, Brazil, New Zealand, South Africa, and the USA. They collected information for people with AIP, variegate porphyria or hereditary coproporphyria, and their families, which is really helpful for the diagnosis and treatment of patients. However, for a long time, public health agencies and the whole society have paid less attention to rare diseases in China (Li et al., 2017;

Ni and Shi, 2017). So, the systematic research on acute porphyria is progressing slowly. To effectively diagnose and treat AIP disease, a systematic collection of patients' information is urgently needed. Our study should provide more useful information to the study of AIP.

AUTHOR CONTRIBUTIONS

In this study, TS, YG and XN designed the study. YF, LY, RY, and JJ conducted the data collection and data analysis. YF, JJ, TS, YG and XN interpreted data. YF and JJ drafted the manuscript. TS, XN, and YG revised and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the China Human Proteome Project (Grant No.2014DFB30010, 2014DFB30030), National High Technology Research and Development Program of China (863 project) (2015AA020108), National Natural Science Foundation of China (31671377, 81472369 and 81502144), Clinical Application Research Funds of Capital Beijing (Z171100001017051), Beihang University & Capital Medical University Advanced Innovation Center for Big Data-Based Precision Medicine Plan (BHME-201801) and Shanghai 111 Project (B14019).

REFERENCES

- Ajioka, R. S., Phillips, J. D., and Kushner, J. P. (2006). Biosynthesis of heme in mammals. *Biochim. Biophys. Acta* 1763, 723–736. doi: 10.1016/j.bbamcr.2006.05.005
- Brennan, M. J., and Cantrill, R. C. (1979). Delta-aminolaevulinic acid is a potent agonist for GABA autoreceptors. *Nature* 280, 514–515. doi: 10.1038/280514a0
- Bustad, H. J., Vorland, M., Ronneseth, E., Sandberg, S., Martinez, A., and Toska, K. (2013). Conformational stability and activity analysis of two hydroxymethylbilane synthase mutants, K132N and V215E, with different phenotypic association with acute intermittent porphyria. *Biosci. Rep.* 33, 617–626. doi: 10.1042/BSR20130045
- Chen, B., Solis-Villa, C., Erwin, A. L., Balwani, M., Nazrenko, I., Phillips, J. D., et al. (2018). Identification and characterization of 40 novel hydroxymethylbilane synthase mutations that cause acute intermittent porphyria. *J. Inher. Metab. Dis.* 42 (1), 186–194. doi: 10.1007/s10545-018-0163-6
- Duque-Serrano, L., Patarroyo-Rodriguez, L., Gotlib, D., and Molano-Eslava, J. C. (2018). Psychiatric aspects of acute porphyria: a comprehensive review. *Curr. Psychiatry Rep.* 20, 5. doi: 10.1007/s11920-018-0867-1
- Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., et al. (2018). The reactome pathway knowledgebase. *Nucleic Acids Res.* 46, D649–D655. doi: 10.1093/nar/gkx1132
- Gill, R., Kolstoe, S. E., Mohammed, F. A., D. B. A., Mosely, J. E., Sarwar, M., et al. (2009). Structure of human porphobilinogen deaminase at 2.8 Å: the molecular basis of acute intermittent porphyria. *J. Biochem. Acids* 120, 17–25. doi: 10.1042/BJ20082077
- Handschin, C., Lin, J., Rhee, J., Peyer, A. K., Chin, S., Wu, P. H., et al. (2005). Nutritional regulation of hepatic heme biosynthesis and porphyria through PGC-1α. *Biochem. Cell Biol.* 122, 505–515. doi: 10.1016/j.cell.2005.06.040
- Innala, E., Backstrom, T., Bixo, M., and Andersson, C. (2010). Evaluation of gonadotropin-releasing hormone agonist treatment for prevention of menstrual-related attacks in acute porphyria. *Acta Obstet. Gynecol. Scand.* 89, 95–100. doi: 10.3109/00016340903390729
- Ivanova, M. M., Radde, B. N., Son, J., Mehta, F. F., Chung, S. H., and Klinge, C. M. (2013). Estradiol and tamoxifen regulate NRF-1 and mitochondrial function

ACKNOWLEDGMENTS

We thank the supercomputer center of East China Normal University for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.01018/full#supplementary-material>

FIGURE S1 | (A) (B) (C) Residues' interaction with DPM cofactor and reaction intermediate in the residue interaction network.

FIGURE S2 | Multiple sequence alignment revealed that mutation residues are strongly conserved among most vertebrate species.

FIGURE S3 | The locations of residues R26, D99, R167, G168, Q194 and G221 on the crystal structure of human HMBS enzyme (5M7F).

TABLE S1 | One hundred seventeen collected AIP patients' clinical symptoms and mutation information.

TABLE S2 | All the enumerated missense mutations and their pathogenic evaluation.

TABLE S3 | Thirty four new potential deleterious mutations identified by our selection.

TABLE S4 | Twenty three deleterious mutations with population distribution frequency.

- in mouse mammary gland and uterus. *J. Mol. Endocrinol.* 51, 233–246. doi: 10.1530/JME-13-0051
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Jia, J., An, Z., Ming, Y., Guo, Y., Li, W., Li, X., et al. (2018a). PedAM: a database for pediatric disease annotation and medicine. *Nucleic Acids Res.* 46, D977–D983. doi: 10.1093/nar/gkx1049
- Jia, J. M., An, Z. X., Ming, Y., Guo, Y. L., Li, W., Liang, Y. X., et al. (2018b). eRAM: encyclopedia of rare disease annotations for precision medicine. *Nucleic Acids Res.* 46, D937–D943. doi: 10.1093/nar/gkx1062
- Junior, V. R. S., Lemos, V. M. V., Feitosa, I. M. L., Florencio, R., Correia, C. W. B., Fontes, L. B. C., et al. (2017). Effect of menstrual cycle on acute intermittent porphyria. *Child Neurol. Open* 4, 2329048X17736170. doi: 10.1177/2329048X17736170
- Karim, Z., Lyoumi, S., Nicolas, G., Deybach, J. C., Gouya, L., and Puy, H. (2015). Porphyrias: a 2015 update. *Clin. Res. Hepatol. Gastroenterol.* 39, 412–425. doi: 10.1016/j.clinre.2015.05.009
- Kimura, R., Takahashi, N., Lin, S., Goto, T., Murota, K., Nakata, R., et al. (2013). DHA attenuates postprandial hyperlipidemia via activating PPARα in intestinal epithelial cells. *J. Lipid. Res.* 54, 3258–3268. doi: 10.1194/jlr.M034942
- Kong, X. F., Han, Y., Li, X. H., Gao, D. Y., Zhang, X. X., and Gong, Q. M. (2013). Recurrent porphyria attacks in a Chinese patient with a heterozygous PBGD mutation. *Gene* 524, 401–402. doi: 10.1016/j.gene.2013.03.130
- Lavandera, J., Rodriguez, J., Ruspini, S., Meiss, R., Zuccoli, J. R., Martinez Mdel, C., et al. (2016). Pleiotropic effects of 5-aminolevulinic acid in mouse brain. *Biochem. Cell Biol.* 94, 297–305. doi: 10.1139/bcb-2015-0094
- Li, Y., Qu, H., Wang, H., Deng, H., and Liu, Z. (2015). Novel A219P mutation of hydroxymethylbilane synthase identified in a Chinese woman with acute intermittent porphyria and syndrome of inappropriate antidiuretic hormone. *Ann. Hum. Genet.* 79, 310–312. doi: 10.1111/ahg.12107
- Li, C., Zhang, J., Li, S., Han, T., Kuang, W., Zhou, Y., et al. (2017). Gene mutations and clinical phenotypes in Chinese children with Blau syndrome. *Sci. China Life Sci.* 60, 758–762. doi: 10.1007/s11427-017-9090-6

- Li, T., Wernersson, R., Hansen, R. B., Horn, H., Mercer, J., Slodkowicz, G., et al. (2017). A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* 14, 61–64. doi: 10.1038/nmeth.4083
- Mahdavi, M., Koulivand, L., Khorrami, M., Mirsafoie, M., and Kheirollahi, M. (2018). In silico analysis of SLC3A1 and SLC7A9 mutations in Iranian patients with Cystinuria. *Mol. Biol. Rep.* 45, 1165–1173. doi: 10.1007/s11033-018-4269-6
- Mao, L., Fang, Y., Campbell, M., and Southerland, W. M. (2017). Population differentiation in allele frequencies of obesity-associated SNPs. *BMC Genomics* 18, 861. doi: 10.1186/s12864-017-4262-9
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60, 681–685. doi: 10.1007/s11427-017-9100-1
- Pluta, P., Roversi, P., Bernardo-Seisdedos, G., Rojas, A. L., Cooper, J. B., Gu, S., et al. (2018). Structural basis of pyrrole polymerization in human porphobilinogen deaminase. *Biochim. Biophys. Acta Gen. Subj.* 1862, 1948–1955. doi: 10.1016/j.bbagen.2018.06.013
- Pulgar, V. M., Yasuda, M., Gan, L., Desnick, R. J., and Bonkovsky, H. L. (2019). Sex differences in vascular reactivity in mesenteric arteries from a mouse model of acute intermittent porphyria. *Mol. Genet. Metab.* doi: 10.1016/j.ymgme.2019.01.005
- Puy, H., Gouya, L., and Deybach, J. C. (2010). Porphyrias. *Lancet* 375, 924–937. doi: 10.1016/S0140-6736(09)61925-5
- Satoh, Y., Iwadate, R., Watanabe, Y., Kawai, H., Kudo, N., Kawashima, Y., et al. (2008). Manifestation of psychiatric behaviors in a mouse model of griseofulvin-induced hepatic porphyria. *J. Toxicol. Sci.* 33, 599–608. doi: 10.2131/jts.33.599
- Shen, Y. (2018). Next-generation sequencing based molecular testing is an equalizer for diagnostic service of rare genetic disorders in China. *Pediatr. Invest.* 2, 96–97. doi: 10.1002/ped4.12036
- Stein, P. E., Badminton, M. N., and Rees, D. C. (2017). Update review of the acute porphyrias. *Br. J. Haematol.* 176, 527–538. doi: 10.1111/bjh.14459
- Thomas, M., Burk, O., Klumpp, B., Kandel, B. A., Damm, G., Weiss, T. S., et al. (2013). Direct transcriptional regulation of human hepatic cytochrome P450 3A4 (CYP3A4) by peroxisome proliferator-activated receptor alpha (PPARalpha). *Mol. Pharmacol.* 83, 709–718. doi: 10.1124/mol.112.082503
- Thomas, M., Winter, S., Klumpp, B., Turpeinen, M., Klein, K., Schwab, M., et al. (2015). Peroxisome proliferator-activated receptor alpha, PPARalpha, directly regulates transcription of cytochrome P450 CYP2C8. *Front. Pharmacol.* 6, 261. doi: 10.3389/fphar.2015.00261
- To-Figueras, J., Badenas, C., Carrera, C., Munoz, C., Mila, M., Lecha, M., et al. (2006). Genetic and biochemical characterization of 16 acute intermittent porphyria cases with a high prevalence of the R173W mutation. *J. Inher. Metab. Dis.* 29, 580–585. doi: 10.1007/s10545-006-0344-6
- Tracy, J. A., and Dyck, P. J. (2014). Porphyria and its neurologic manifestations. *Handb. Clin. Neurol.* 120, 839–849. doi: 10.1016/B978-0-7020-4087-0.00056-5
- Windebank, A. J., and McDonald, E. S. J. P. N. (2005) Chapter 17 — neurotrophic factors in the peripheral nervous system. 377–386. doi: 10.1016/B978-0-7216-9491-7.50020-X
- Windebank, A. J., and Bonkovsky, H. L. (2005). “Chapter 80 — porphyric neuropathy,” in *Peripheral Neuropathy* (Elsevier), 1883–1892. doi: 10.1016/B978-0-7216-9491-7.50083-1
- Yang, J., Wang, H., Yin, K., Hua, B., Zhu, T., Zhao, Y., et al. (2015). A novel mutation in the porphobilinogen deaminase gene in an extended Chinese family with acute intermittent porphyria. *Gene* 565, 288–290. doi: 10.1016/j.gene.2015.04.027
- Yang, J., Chen, Q., Yang, H., Hua, B., Zhu, T., Zhao, Y., et al. (2016). Clinical and laboratory features of acute porphyria: a study of 36 subjects in a Chinese tertiary referral center. *Biomed. Res. Int.* 2016, 3927635. doi: 10.1155/2016/3927635
- Yu, D. (2018). Next-generation sequencing applied in pediatric diseases. *Pediatr. Invest.* 2, 82–82. doi: 10.1002/ped4.12045
- Zanella, A., Fermo, E., Bianchi, P., Chiarelli, L. R., and Valentini, G. (2007). Pyruvate kinase deficiency: the genotype-phenotype association. *Blood Rev.* 21, 217–231. doi: 10.1016/j.blre.2007.01.001

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor is currently editing co-organizing a Research Topic with one of the authors TS, and confirms the absence of any other collaboration.

Copyright © 2019 Fu, Jia, Yue, Yang, Guo, Ni and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Frequent Mutations of VHL Gene and the Clinical Phenotypes in the Largest Chinese Cohort With Von Hippel–Lindau Disease

Baoan Hong^{1,2,3,4,5,6†}, Kaifang Ma^{1,2,3,4†}, Jingcheng Zhou^{1,2,3,4}, Jiufeng Zhang^{1,2,3,4}, Jiangyi Wang^{1,2,3,4}, Shengjie Liu^{1,2,3,4}, Zhongyuan Zhang^{1,2,3,4}, Lin Cai^{1,2,3,4*}, Ning Zhang^{5,6} and Kan Gong^{1,2,3,4}

¹ Department of Urology, Peking University First Hospital, Beijing, China, ² Hereditary Kidney Cancer Research Center, Peking University First Hospital, Beijing, China, ³ Institute of Urology, Peking University, Beijing, China, ⁴ National Urological Cancer Center, Beijing, China, ⁵ Department of Urology, Beijing Cancer Hospital, Beijing, China, ⁶ Beijing Institute for Cancer Research, Beijing, China

OPEN ACCESS

Edited by:

Weida Tong,
National Center for Toxicological
Research (FDA),
United States

Reviewed by:

Asma Sultana,
King Saud University Medical City,
Saudi Arabia
Yanfeng Zhang,
HudsonAlpha Institute for
Biotechnology,
United States

*Correspondence:

Lin Cai
drcallin@163.com

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 05 September 2018

Accepted: 20 August 2019

Published: 18 September 2019

Citation:

Hong B, Ma K, Zhou J, Zhang J,
Wang J, Liu S, Zhang Z, Cai L,
Zhang N and Gong K (2019)
Frequent Mutations of VHL Gene and
the Clinical Phenotypes in the Largest
Chinese Cohort With Von Hippel–
Lindau Disease.
Front. Genet. 10:867.
doi: 10.3389/fgene.2019.00867

Von Hippel–Lindau (VHL) disease is a rare autosomal-dominant inherited tumor syndrome. We aimed to analyze the correlations between frequent *VHL* mutations and phenotypes in Chinese VHL families. We screened 540 patients from 187 unrelated Chinese VHL families for 19 frequent VHL mutations. The penetrance and mean age at onset for VHL-associated susceptible organs were calculated and compared. The overall survival of VHL patients was described with Kaplan–Meier curves. Among the 19 frequent germline mutations, there were four hotspot mutation sites (194, 481, 499, and 500). Missense mutations were the most common types of mutations (70.0%) followed by nonsense mutations (20.0%) and splicing mutations (10.0%). Due to the diversity of these mutations, the penetrance for each organ and the age at onset are distinct. Even in cases of similar mutations, variance in the penetrance and age at onset was observed. The mean age at death for the patients in this cohort was 42.4 ± 13.5 years, and variability was observed in the Kaplan–Meier curves. We present a precise summary of the phenotypes for the frequent *VHL* mutations in the largest Chinese VHL cohort, which provides valuable strategies for genetic counseling and clinical surveillance of VHL individuals.

Keywords: von Hippel–Lindau disease, VHL mutation, genotype–phenotype correlation, onset age, survival

INTRODUCTION

Von Hippel–Lindau (VHL) disease (OMIM no. 193300) is an autosomal-dominant familial neoplastic condition that is caused by germline mutations in the *VHL* gene located on chromosome 3p25–26. This gene comprises three exons: exon 1 spans nucleotides 1–340 (codons 1–113), exon 2 spans nucleotides 341–463 (codons 114–154), and exon 3 spans nucleotides 464–642 (codons 155–213) (Latif et al., 1993; Nielsen et al., 2016; Varshney et al., 2017). Patients with VHL syndrome inherit a single mutant *VHL* allele from a parent and develop the disease when the second wild-type copy is deactivated or lost. Incidence of the *VHL* mutation is approximately 1 in 36,000 live births, and it is greater than 90% penetrant by age 65 years (Kim et al., 2010; Gossage et al., 2015). Common VHL-associated clinical manifestations include central nervous system hemangioblastoma (CHB),

renal cell carcinoma or renal cyst (RCC), retinal angioma (RA), pancreatic tumor or cyst (PCT), pheochromocytoma and paragangliomas (PHEO), endolymphatic sac tumor, and epididymis or broad ligament cystadenoma (**Supplementary Table 1**; Lonser et al., 2003; Lonser et al., 2004; Butman et al., 2008; Chou et al., 2013; Launbjerg et al., 2017). Von Hippel–Lindau disease predisposes the affected individuals to the development of lesions in multiple systems with CHBs (25%–51%) and RCCs (13%–47%) as the major causes of mortality (Grubb et al., 2005; Wilding et al., 2012; Lonser et al., 2014). Individuals with a family history of VHL will be clinically diagnosed when he/she presents with VHL-associated tumors that include CHB, RA, or RCC (Chittiboina and Lonser, 2015). For patients who do not have a family history of VHL, two or more CHBs or RAs or one hemangioblastoma and a visceral tumor are required for a clinical diagnosis (Nielsen et al., 2016). Typically, genetic testing is the standard method to diagnose VHL disease. The detection of *VHL* mutations not only contributes to an early and precise diagnosis of at-risk individuals but also helps to elucidate the genotype–phenotype correlations within a given population.

A series of studies have reported genotype–phenotype correlations in VHL diseases from different research perspectives or within different ethnic backgrounds (Yoshida et al., 2000; Patocs et al., 2008; Gomy et al., 2010; Nordstrom-O'Brien et al., 2010; Vikkath et al., 2015). For example, a retrospective study that included 63 VHL patients from two large VHL kindreds (family 1: Y112H mutation and family 2: Y98H mutation) with pheochromocytoma/paraganglioma found that pheochromocytoma expressivity differed by genotype (Nielsen et al., 2011). Ong et al. (2007) evaluated the genotype–phenotype correlations in 573 VHL patients and confirmed that pheochromocytoma was linked to *VHL* missense mutations. Additionally, the age at onset for VHL syndrome was significantly earlier ($P = 0.001$) and the age-related risks of RA and RCC were higher ($P = 0.022$ and $P = 0.0008$, respectively) for individuals with nonsense or frameshift mutations compared to those with deletions. Importantly, the results of these studies provided valuable strategies for genetic counseling and clinical prophylactic surveillance for VHL family members.

Due to the rarity of VHL disease, studies on the correlations between the frequent mutations of the *VHL* gene and clinical phenotypes are relatively scarce, with the majority being case reports or studies involving a limited number of VHL patients or families. In clinical practice, there is an urgent need to identify the clinical symptoms and survival statistics for VHL patients based on their specific types of mutations. Therefore, an improved and precise understanding of specific genotype–phenotype correlations in VHL disease is essential for targeted monitoring and counseling. In this study, we screened for frequent mutations in the *VHL* gene across 187 unrelated Chinese VHL families and analyzed the genotype–phenotype correlations between frequent *VHL* mutations and clinical manifestations. This study improves our understanding of how frequent mutations of the *VHL* gene affect the age at onset for each susceptible organ and their impact on prognosis in a Chinese population and provides a more accurate resource for genetic counseling and the monitoring of VHL patients.

MATERIALS AND METHODS

Patient Selection

By May 31, 2018, 540 patients from 187 unrelated families had been diagnosed with VHL disease at the Peking University First Hospital, and all were screened in the present study. Any mutation site that appeared in two or more families was included in this study. Individuals who carried a *VHL* germline mutation or who met the clinical criteria were diagnosed with VHL disease (Wu et al., 2012). The evidence assessment for VHL disease includes a medical history, a physical examination by multidisciplinary teams, laboratory tests, and medical imaging of the abdomen, pelvis, brain, and spine (ultrasound, computed tomography, or magnetic resonance imaging). However, at least one patient from a family confirmed the presence of a *VHL* mutation through genetic testing.

Clinical characteristics including date of birth, age at death, mutation type, clinical symptoms, and the age at onset were collected from family members or through medical records. The age at onset was defined as the age at which VHL-related symptoms or signs first appear. A follow-up was performed on all patients and asymptomatic carriers in the VHL families to determine the age at onset for six major VHL-related lesions: CHB, RA, RCC, PCT, PHEO, and GS (genital system including the epididymis or broad ligament). Follow-ups were carried out until May 31, 2018, or until patient death following the first presentation of the VHL-related clinical symptoms or a positive genetic diagnosis of VHL disease. The life span of the patient from birth until death or until the end of the follow-up period was used in the survival analysis.

Genetic Analysis

Genetic testing was performed on at least one member from each family to confirm the diagnosis of VHL disease. Genomic DNA was extracted from peripheral blood samples that were obtained from members of the suspected families. Individuals who refused the genetic test or who died were excluded. The three exons and the flanking intronic sequences of the *VHL* gene were amplified by polymerase chain reaction, and direct sequencing was used to detect missense mutations, splicing mutations, and small indels. Large deletions and duplications were detected by multiplex ligation-dependent probe amplification (MLPA, P016-C2 kit, Amsterdam, the Netherlands) and verified by real-time quantitative polymerase chain reaction. The primers and reaction conditions used for the amplification were as previously described (Peng et al., 2017; Wang et al., 2018). The spectrum of *VHL* mutations was screened for frequent mutations and further analyzed to determine their correlations with phenotypes.

Statistical Analysis

For each type of mutation, the mean age at onset of VHL-associated susceptible organs (CHB, RA, RCC, PCT, PHEO, and GS) and the mean age at death were calculated as the mean \pm standard deviation. Statistical significance was determined by the Student *t* test or the LSD multiple-comparisons *t* test. The

overall survival of VHL patients was described with a Kaplan–Meier curve. Statistical analysis was performed using SPSS 20.0 software (IBM-SPSS, Inc., Chicago, IL, USA), and $P < 0.05$ was considered to be statistically significant.

RESULTS

Clinical Characteristics of Chinese VHL Patients and the Distribution of Frequent Germline Mutations

A total of 540 patients from 187 unrelated Chinese VHL families were included in our database, and 126 different types of *VHL* mutations were identified. Insertions, small deletions, and large deletions were detected in 61 families [61/187 (32.6%)]. Point mutations resulting in missense, nonsense, or splicing mutations were detected in 126 families [126/187 (67.4%)]. The distribution of germline point mutations was further analyzed and identified 19 germline mutation sites that appeared in two or more families. Of these 19 frequent germline mutations, 10 were located in exon 1, 3 were located in exon 2, 4 were located in exon 3, and 2 were located in intron 2 (Figure 1). The clinical characteristics and mutation frequencies for each of the relevant germline mutations in the 258 patients from 80 unrelated Chinese VHL families are shown in Table 1. The mean age of the 258 patients was 39.4 ± 15.5 years with a range of 3 to 74 years. Additionally, the mean age for each mutation group is listed in Table 1. As would be expected due to the consequences of frequent mutations, missense mutations were the most common mutation types in these families (70.0%) followed by nonsense mutations (20.0%) and splicing mutations (10.0%). Notably, there were four mutation hotspot sites at 194, 481, 499, and 500 that were separately distributed across 9, 10, 14, and 10 unrelated Chinese VHL families, respectively.

Correlations Between Frequent VHL Germline Mutations and Clinical Phenotypes in Chinese VHL Patients

In this study, we analyzed and compared the mean age at onset for six major VHL-related lesions (CHB, RA, RCC, PCT, PHEO, and GS) in 19 frequent germline mutations (Table 2). Due to diversity in the types of mutations, the penetrance for each organ and the age at onset are not the same. For CHB, both the c.481C > T p.Arg161stop (group 16) and c.486C > G p.Cys162Trp (group 17) mutations had a high penetrance of approximately 80.0% (12/15). The mean age at onset of CHB for the c.481C > T p.Arg161stop (group 16) mutation was 27.4 ± 9.4 years (range = 14–40 years), while for the c.486C > G p.Cys162Trp (group 17) mutation, it was 31.4 ± 10.0 years (range = 12–49 years). For RCC, the mean ages at onset for the c.269A > T p.Asn90Ile (group 8) and c.486C > G p.Cys162Trp (group 17) mutations were 41.5 ± 15.5 years and 41.8 ± 10.3 years, while the penetrance was 26.7% (4/15) and 53.3% (8/15), respectively. Variation in the penetrance and age at onset exists even when the types of mutations are similar. There were two types of missense mutations in group 1 located in the 194 mutation site, c.194C > T p.Ser65Leu and c.194C > G p.Ser65Trp, but the clinical phenotypes were different between these two mutational subgroups. Six major VHL-related lesions (CHB, RA, RCC, PCT, PHEO, and GS) were observed in the c.194C > T p.Ser65Leu mutational subgroup, while only three VHL lesions (CHB, RCC, and PCT) presented in the c.194C > G p.Ser65Trp mutational subgroup. The mean age at onset for the common VHL lesions in the c.194C > T p.Ser65Leu mutational subgroup was older than that of the c.194C > G p.Ser65Trp mutational subgroup (Figures 2A–D). Three mutations were related to codon 88 (Trp) in groups 5 and 6. Both c.262T > C p.Trp88Arg and c.263G > C p.Trp88Ser were missense mutations, while c.263G > A p.Trp88Stop resulted in a nonsense mutation. Additionally, the VHL lesions associated with these three mutations were not the

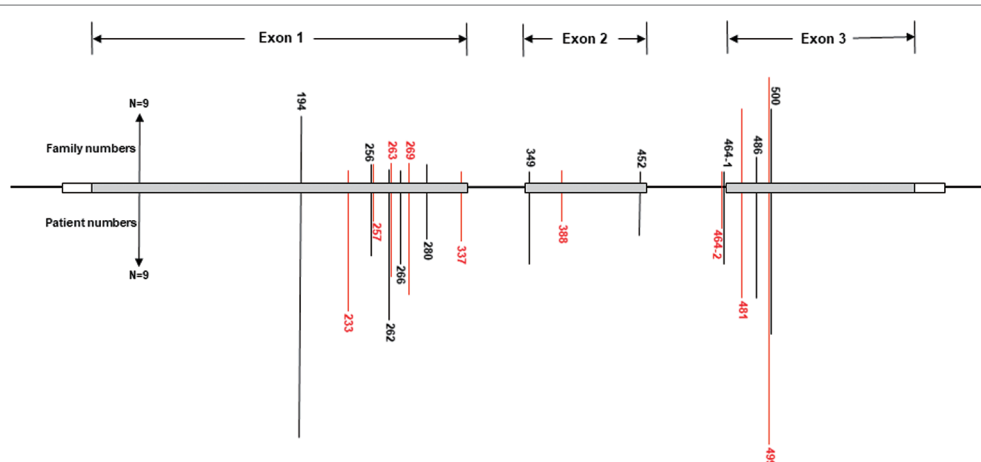


FIGURE 1 | Distribution of the frequent germline mutation sites in *VHL* gene for 258 patients from 80 unrelated Chinese families with VHL disease. Of the 19 frequent germline mutations, 10 were located in exon 1, 3 were located in exon 2, 4 were located in exon 3, and 2 were located in intron 2. Notably, there were four hotspot mutation sites (194, 481, 499, and 500). The bars on the upper part represent mutational family numbers, and those on the lower part represent mutational patient numbers.

TABLE 1 | Clinical characteristics and mutation frequency of relevant frequent germline mutations of 258 patients from 80 unrelated Chinese VHL families.

Mutation sites	Mean age	Exon	Mutation type	FAMILY NUMBERS		Patient numbers	
				n	%	n	%
194	39.6 ± 16.3	1	Missense	9	4.81	31	5.74
233	34.6 ± 19.5	1	Missense	2	1.07	19	3.52
256	35.0 ± 15.8	1	Missense	3	1.60	9	1.67
257	30.3 ± 8.0	1	Missense	3	1.60	4	0.74
262	45.1 ± 10.1	1	Missense	2	1.07	21	3.89
263	33.8 ± 17.6	1	Missense	3	1.60	13	2.41
			Nonsense				
266	36.6 ± 15.3	1	Missense	2	1.07	11	2.04
269	36.9 ± 15.6	1	Missense	3	1.60	15	2.78
280	42.0 ± 12.6	1	Nonsense	3	1.60	6	1.11
337	47.0 ± 18.4	1	Nonsense	2	1.07	6	1.11
349	44.2 ± 15.5	2	Missense	2	1.07	11	2.04
388	41.8 ± 10.1	2	Missense	2	1.07	4	0.74
452	39.0 ± 13.1	2	Missense	2	1.07	7	1.30
464-1	28.8 ± 19.6	Intron 2	Splicing	2	1.07	11	2.04
464-2	42.6 ± 17.9	Intron 2	Splicing	2	1.07	5	0.93
481	39.2 ± 10.4	3	Nonsense	10	5.35	15	2.78
486	38.3 ± 13.9	3	Missense	4	2.14	15	2.78
499	38.2 ± 17.4	3	Missense	14	7.49	31	5.74
500	47.8 ± 12.7	3	Missense	10	5.35	24	4.44

same. A comparison was made between the mean age at onset for CHB in these three subgroups and found that the c.262T > C p.Trp88Arg mutational group was older than that of the c.263G > C p.Trp88Ser mutational subgroup ($P = 0.0152$) and the c.263G > A p.Trp88Stop mutational subgroup ($P = 0.0232$) (Figure 2E). However, the CHB-associated age at onset for the c.263G > A p.Trp88Stop mutational subgroup was younger than the c.263G > C p.Trp88Ser mutational subgroup, but the difference was not significant ($P = 0.481$) (Figure 2E). In groups 18 and 19, the c.499C > T p.Arg167Trp and c.500G > A p.Arg167Gln mutations were located in codon 167 (Arg). There were differences in the penetrance and age at onset for six major VHL-related lesions (CHB, RA, RCC, PCT, PHEO, and GS) between these two mutation types, but the difference was not statistically significant ($P > 0.05$) (Figures 2F–I).

Frequent VHL Germline Mutations and Survival

Kaplan–Meier curves were used to describe the survival of patients with different VHL mutations, and the results are presented in Figure 3A. This analysis identified a variety of Kaplan–Meier curves for different frequent VHL germline mutations. The mutation sites 256 and 257, 262 and 263, and 499 and 500 were located in codons 86, 88, and 167, respectively, and grouped together for the analysis. Of the 258 patients from the 80 unrelated Chinese VHL families, 59 died of VHL-related diseases, such as CHB [71.2% (42/59)], RCC [25.4% (15/59)], and PCT-related complications [3.4% (2/59)]. The mean age at death for this cohort was 42.4 ± 13.5 years (range = 16–68 years). Additionally, we summarized the mean age at death for each of the frequent VHL germline mutation groups, and no deaths were observed in the groups with mutation sites 337, 388, and 464-2 (Figure 3B). Differences in the risk of VHL-related death

across the different frequent VHL germline mutations were not statistically significant.

DISCUSSION

A comprehensive understanding of the correlations between genotype and phenotype for hereditary diseases is critical to the clinical management and scientific analysis of their pathogenesis. Changes in specific genotypes can lead to alterations in protein expression patterns that result in their corresponding phenotypes. Elucidating these correlations may provide insight into the molecular pathogenesis of the individual manifestations of VHL syndrome. Screening for mutations in the VHL gene helps to clarify the diagnosis of asymptomatic first-degree relatives, thereby improving patient outcomes through early disease surveillance. To date, the studies on genotype–phenotype correlations have provided clinicians with tools that help predict the VHL disease processes in individual patients. Hence, it is important to increase the sample size and analyze the correlations between genotypes and phenotypes for different ethnic groups.

In this study, we analyzed the correlations between frequent mutations in the VHL gene and clinical phenotypes in the largest Chinese VHL cohort to date. In total, we screened 540 patients from 187 unrelated VHL families and identified 126 different VHL mutations. Furthermore, we identified 19 frequent mutations and four mutation hotspots and further investigated the genotype–phenotype correlations. Notably, patients or families with the VHL disease have a range of different phenotypes. A variety of factors may contribute to this diversity of phenotypes, including the type of VHL mutation, the site of the mutation, and ethnic background.

Different ethnic backgrounds are associated with diverse phenotypes. Several studies about Western and Japanese populations highlighted the differences in the spectrum of VHL

TABLE 2 | Frequent *VHL* germline mutations and related phenotypes in Chinese patients with VHL disease.

Group	NC	CHB		RA		RCC		PCT		PHEO		GS	
		Ratio (%)	OA (Mean ± SD)	Ratio (%)	OA (Mean ± SD)	Ratio (%)	OA (Mean ± SD)	Ratio (%)	OA (Mean ± SD)	Ratio (%)	OA (Mean ± SD)	Ratio (%)	OA (Mean ± SD)
1*	c.194C > T	62.5	34.1 ± 13.8	8.3	(30, 53)	37.5	45.4 ± 12.5	45.8	43.5 ± 13.5	4.2	(34)	16.7	25.0 ± 8.3
	p.Ser65Leu	(15/24)	(15–59)	(2/24)	—	(9/24)	(25–60)	(11/24)	(25–65)	(1/24)	—	(4/24)	(17–34)
	c.194C > G	71.4	30.2 ± 18.3	—	—	57.1	36.3 ± 13.6	42.9	32.3 ± 13.6	—	—	—	—
2	p.Ser65Trp	(5/7)	(17–61)	—	—	(4/7)	(24–48)	(3/7)	(24–48)	—	—	—	—
	c.233A > G	21.1	41.8 ± 10.4	—	—	15.8	38.3 ± 2.5	36.8	27.4 ± 6.9	10.5	(28, 36)	—	—
	p.Asn78Ser	(4/19)	(30–51)	—	—	(3/19)	(36–41)	(7/19)	(20–40)	(2/19)	—	—	—
3	c.256C > T	33.3	34.0 ± 2.6	22.2	(11, 12)	55.6	34.0 ± 4.1	33.3	33.3 ± 2.5	11.1	(31)	—	—
	p.Pro86Ser	(3/9)	(31–36)	(2/9)	—	(5/9)	(30–40)	(3/9)	(31–36)	(1/9)	—	—	—
	c.257C > T	75.0	26.3 ± 8.0	50.0	(26, 32)	75.0	28.0 ± 9.2	75.0	27.0 ± 10.5	—	—	25.0	(26)
4	p.Pro86Leu	(3/4)	(18–34)	(2/4)	—	(3/4)	(20–38)	(3/4)	(17–38)	—	—	(1/4)	—
	c.262T > C	52.4	41.1 ± 9.1	—	—	38.1	37.4 ± 12.9	4.8	(30)	4.8	(50)	—	—
	p.Trp88Arg	(11/21)	(29–62)	—	—	(8/21)	(23–65)	(1/21)	—	(1/21)	—	—	—
6*	c.263G > A	75.0	25.0 ± 12.3	—	—	50.0	(28, 34)	50.0	(28, 38)	25.0	(39)	50.0	(11, 32)
	p.Trp88Stop	(3/4)	(11–34)	—	—	(2/4)	—	(2/4)	—	(1/4)	—	(2/4)	—
	c.263G > C	77.8	29.9 ± 8.4	22.2	(25, 35)	11.1	(30)	22.2	(30, 31)	—	—	11.1	(24)
7	p.Trp88Ser	(7/9)	(20–44)	(2/9)	—	(1/9)	—	(2/9)	—	—	—	(1/9)	—
	c.266T > C	54.5	27.7 ± 7.5	45.5	29.6 ± 13.1	54.5	39.5 ± 9.3	36.4	25.8 ± 2.2	9.1	(45)	—	—
	p.Leu89Pro	(6/11)	(20–40)	(5/11)	(15–42)	(6/11)	(24–52)	(4/11)	(24–29)	(1/11)	—	—	—
8	c.269A > T	66.7	29.9 ± 13.4	13.3	(23, 30)	26.7	41.5 ± 15.5	33.3	41.4 ± 17.0	6.7	(38)	—	—
	p.Asn90Ile	(10/15)	(16–63)	(2/15)	—	(4/15)	(26–63)	(5/15)	(17–63)	(1/15)	—	—	—
	c.280G > T	66.7	34.5 ± 13.5	33.3	(17, 29)	50.0	35.0 ± 15.9	33.3	(29, 53)	16.7	(35)	33.3	(14, 20)
9	p.Glu94Stop	(4/6)	(23–54)	(2/6)	—	(3/6)	(23–53)	(2/6)	—	(1/6)	—	(2/6)	—
	c.337C > T	50.0	40.0 ± 22.5	50.0	23.0 ± 7.2	50.0	35.3 ± 11.3	33.3	(26, 27)	—	—	16.7	(18)
	p.Arg113Stop	(3/6)	(28–66)	(3/6)	(17–31)	(3/6)	(26–48)	(2/6)	—	—	—	(1/6)	—
11	c.349T > G	72.7	30.0 ± 14.6	45.4	23.8 ± 18.2	54.5	45.8 ± 8.1	72.7	43.4 ± 14.3	9.1	(18)	9.1	(44)
	p.Trp117Gly	(8/11)	(12–47)	(5/11)	(7–44)	(6/11)	(37–61)	(8/11)	(21–59)	(1/11)	—	(1/11)	—
	c.388G > C	50.0	(30, 30)	—	—	25.0	(31)	25.0	(31)	25.0	(30)	—	—
12	p.Val130Leu	(2/4)	—	—	—	(1/4)	—	(1/4)	—	(1/4)	—	—	—
	c.452T > G	71.4	34.2 ± 18.1	14.3	(26)	42.9	36.7 ± 11.7	28.6	(32, 38)	—	—	—	—
	p.Ile151Ser	(5/7)	(14–60)	(1/7)	—	(3/7)	(28–50)	(2/7)	—	—	—	—	—
14	Intron 2 *	54.5	33.0 ± 18.4	—	—	36.4	42.3 ± 16.5	36.4	42.3 ± 16.5	—	—	9.1	(10)
	c.464–1 G > C	(6/11)	(16–66)	—	—	(4/11)	(28–66)	(4/11)	(28–66)	—	—	(1/11)	—
	Intron 2 *	40.0	(24, 57)	20.0	(57)	60.0	40.3 ± 16.8	60.0	40.3 ± 16.8	20.0	(58)	—	—
15	c.464–2 A > G	(2/5)	—	(1/5)	—	(3/5)	(21–51)	(3/5)	(21–51)	(1/5)	—	—	—
	c.481C > T	80.0	27.4 ± 9.4	6.7	(28)	66.7	36.6 ± 8.2	60.0	32.3 ± 9.7	6.7	(41)	13.3	(30, 36)
	p.Arg161Stop	(12/15)	(14–40)	(1/15)	—	(10/15)	(23–47)	(9/15)	(14–46)	(1/15)	—	(2/15)	—
17	c.486C > G	80.0	31.4 ± 10.0	26.7	36.5 ± 16.3	53.3	41.8 ± 10.3	40.0	33.0 ± 12.9	6.7	(29)	6.7	(23)
	p.Cys162Trp	(12/15)	(12–49)	(4/15)	(17–54)	(8/15)	(29–56)	(6/15)	(17–56)	(1/15)	—	(1/15)	—
	c.499C > T	35.5	30.9 ± 13.8	12.9	23.5 ± 8.7	22.6	35.4 ± 11.8	29.0	36.9 ± 14.5	32.3	33.3 ± 14.3	3.2	(31)
18	p.Arg167Trp	(11/31)	(16–65)	(4/31)	(13–34)	(7/31)	(18–54)	(9/31)	(19–67)	(10/31)	(12–58)	(1/31)	—
	c.500G > A	62.5	35.6 ± 12.6	16.7	23.0 ± 6.8	41.7	38.3 ± 13.6	20.8	37.2 ± 14.5	37.5	36.3 ± 11.9	—	—
	p.Arg167Gln	(15/24)	(15–52)	(4/24)	(14–30)	(10/24)	(22–54)	(5/24)	(22–53)	(9/24)	(14–52)	—	—

*There are two mutational types in this site. *Splicing mutation.

—, no patient with this phenotype; NC, nucleotide change and consequence; OA, onset age (year); CHB, central nervous system hemangioblastoma; RA, retinal angioma; RCC, renal cell carcinoma or cyst; PCT, pancreatic tumor or cyst; PHEO, pheochromocytoma; GS, genital system (epididymis or broad ligament).

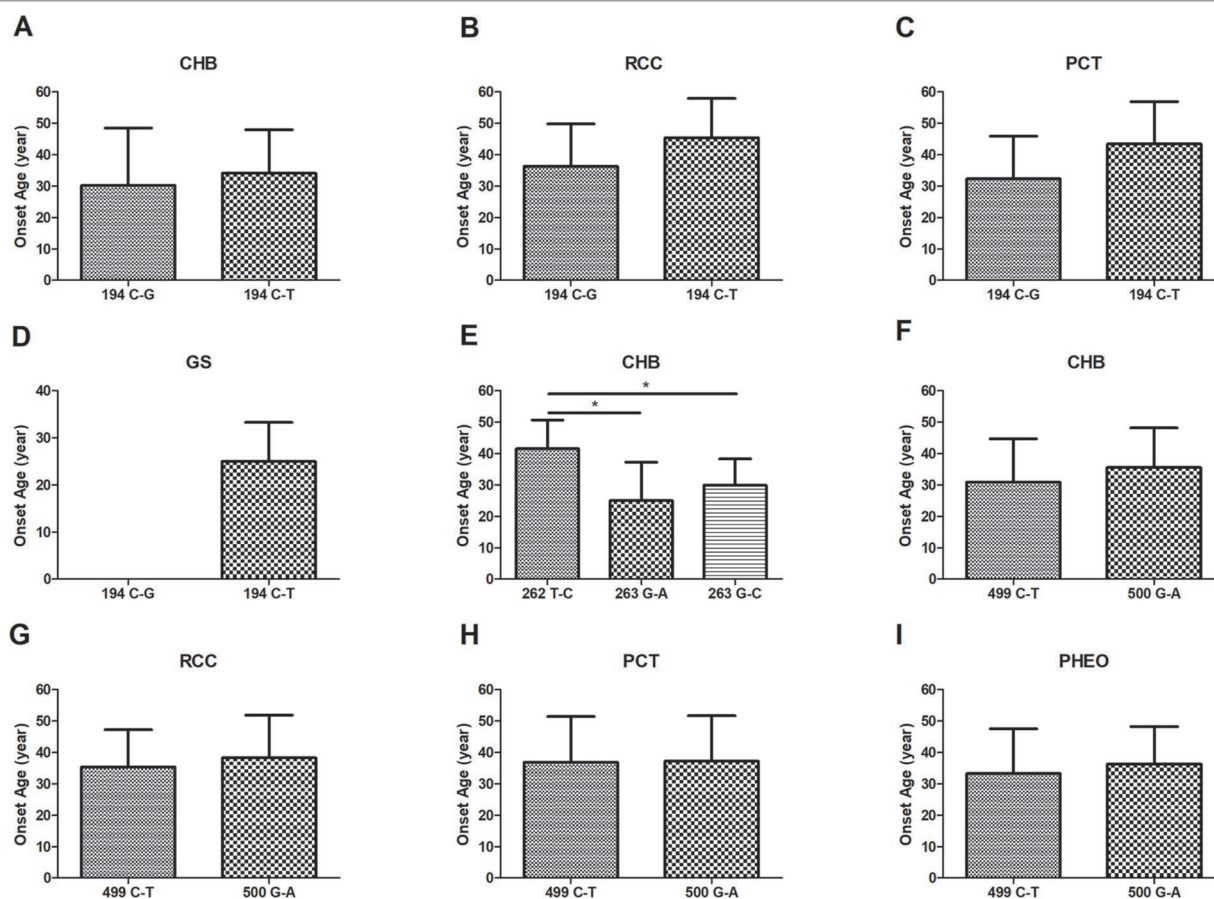


FIGURE 2 | The mean age at onset for the common VHL lesions in patients with different germline mutations. The mean age at onset for the common VHL lesions in the c.194C > T p.Ser65Leu mutational subgroup and the c.194C > G p.Ser65Trp mutational subgroup (A–D). Comparison of the mean age at onset for CHB among c.262T > C p.Trp88Arg, c.263G > A p.Trp88Stop, and c.263G > C p.Trp88Ser mutations (E). Differences in the mean age at onset for the four major VHL lesions between the c.499C > T p.Arg167Trp and c.500G > A p.Arg167Gln mutations (F–I, $P > 0.05$). * $P < 0.05$.

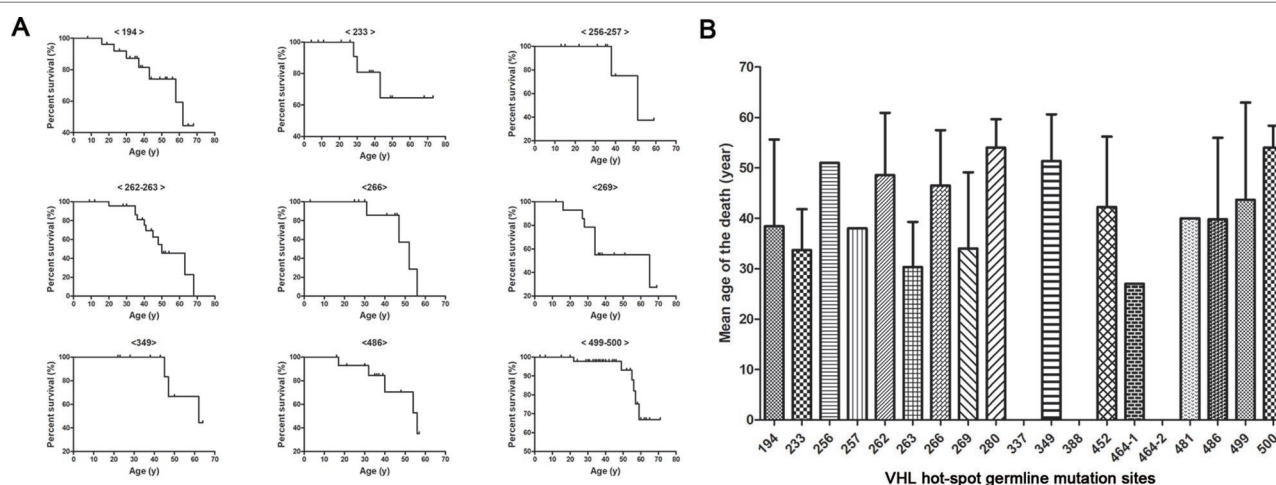


FIGURE 3 | Von Hippel-Lindau frequent germline mutations and survival. Kaplan-Meier curves were used to describe the survival of patients with different VHL mutations (A). Variability was observed in the mean age at death for the different frequent VHL germline mutations, but no death was observed in groups with mutation sites 337, 388 and 464-2 (B).

germline mutations (Maher et al., 1996; Patel et al., 2000). The mutation hotspots of the *VHL* gene that are already known include Leu178, Cys162, Arg167, Asn78, Pro86, and Tyr98 and have a frequency of approximately 3% to 17% (Stebbins et al., 1999). However, the common mutations are varied across different ethnic groups. Hwang et al. (2014) reported that Glu70Lys was a high-frequency *VHL* germline mutation in the Korean population, with nine unrelated patients [16.4% (9/55)] who had the same amino-acid alteration at codon 70 (Glu70Lys) and exhibited VHL type 1 phenotypes. However, in our cohort, the high-frequency mutations included Ser65 (4.81%), Arg161 (5.35%), and Arg167 (12.84%). Thus, the spectrum of *VHL* mutations varies in countries that have different ethnic backgrounds. Patients from different ethnic backgrounds that have the same *VHL* germline mutation may also develop distinct phenotypes. For example, Yoshida et al. (2000) found four mutations (Arg113Stop, Gln132Stop, Leu158Val, and Cys162Tyr) in Japanese families with the VHL type 2 phenotype, whereas Crossey et al. (1994) reported that these mutations were associated with the VHL type 1 phenotype in Western populations. Similarly, c.500G > A p.Arg167Gln is a hotspot mutation in many populations. Studies in Western populations showed that the mutation of c.500G > A p.Arg167Gln was associated with RCC and renal cysts, indicating that this mutation was associated with the VHL type 1 phenotype (Hes et al., 2007; Ciotti et al., 2009). However, according to our database, we found that this mutation was also related to PHEO (37.5%, 9 of 24), which indicated that the phenotype of the c.500G > A p.Arg167Gln mutation also differs in different ethnic backgrounds.

Different mutations may produce remarkably diverse phenotypes. For example, in group 1, two missense mutations occurred in the 194 mutation site, c.194C > T p.Ser65Leu and c.194C > G p.Ser65Trp. Intriguingly, the six major VHL-related lesions (CHB, RA, RCC, PCT, PHEO, and GS) were observed in the c.194C > T p.Ser65Leu mutational subgroup, while only three VHL lesions (CHB, RCC, and PCT) presented in the c.194C > G p.Ser65Trp mutational subgroup. Therefore, it indicates that missense mutations of different nucleotides in the same codon have potential effects on the clinical manifestation. The much lower number of patients in c.194C > G p.Ser65Trp mutational subgroup may limit the observation of the six major VHL-related lesions. This finding needs to be corroborated in larger cohorts. Bradley et al. (1999) also reported on the phenotypes of two distinct missense mutations in the same codon of the *VHL*

gene (c.334T > A p.Tyr112Asn and c.334T > C p.Tyr112His). Thirteen patients were found with the c.334T > A p.Tyr112Asn mutation, seven of whom had RCC, and one of these patients had a pheochromocytoma, which suggests that this type of mutation causes the VHL type 1 phenotype, as most of the patients presented with RCC. Conversely, the c.334T > C p.Tyr112His mutation was associated with the VHL type 2A phenotype, as every affected individual in two families (22 patients) had PHEO but did not have RCC. Thus, different amino-acid changes at the same position may have different effects on the stability of the VHL protein, resulting in distinct clinical phenotypes.

Family members with the same mutation in *VHL* can also display different phenotypes. Mete et al. (2014) evaluated the clinical presentation of 49 family members from three generations of a Turkish family and identified the *VHL* p.A149S mutation. All of the patients were diagnosed with VHL syndrome type 2B, while nine patients were diagnosed with a pheochromocytoma, and one patient was diagnosed with a lumbar spinal hemangioblastoma and a pancreatic neuroendocrine tumor without pheochromocytoma. In our study, group 5 represented the c.262T > C p.Trp88Arg mutation that included 21 patients from 2 families (Figure 4 and Table 3). Variability was observed in the mean age at onset for CHB and RCC, which were 41.1 ± 9.1 (range = 29–62) and 37.4 ± 12.9 (range = 23–65), respectively. Moreover, the penetrance of VHL lesions was also distinct. Taken together, this phenotypic variability suggests that other factors or molecular mechanisms may affect the phenotypes caused by specific mutations, such as environmental factors or telomere length (Ning et al., 2014; Wang et al., 2017).

A classification of VHL diseases was proposed that was based on the patient's preference for PHEO development. For example, the VHL type 1 phenotype has a low risk of PHEOs compared to the VHL type 2 phenotype, which is associated with PHEOs (Ong et al., 2007). The VHL type 2 phenotype is further classified into type 2A (hemangioblastoma and PHEO, but rarely RCC), type 2B (hemangioblastoma, PHEO, and RCC), and type 2C (PHEO only). However, during follow-up, the VHL disease shows characteristics of phenotypic variability, and new clinical manifestations may appear during the patient's lifetime. Recently, Liu et al. (2018) reported genotype-phenotype correlations in VHL disease based on the alteration of a HIF- α binding site in the VHL protein. Lee et al. (2016) analyzed the genotype-phenotype correlations

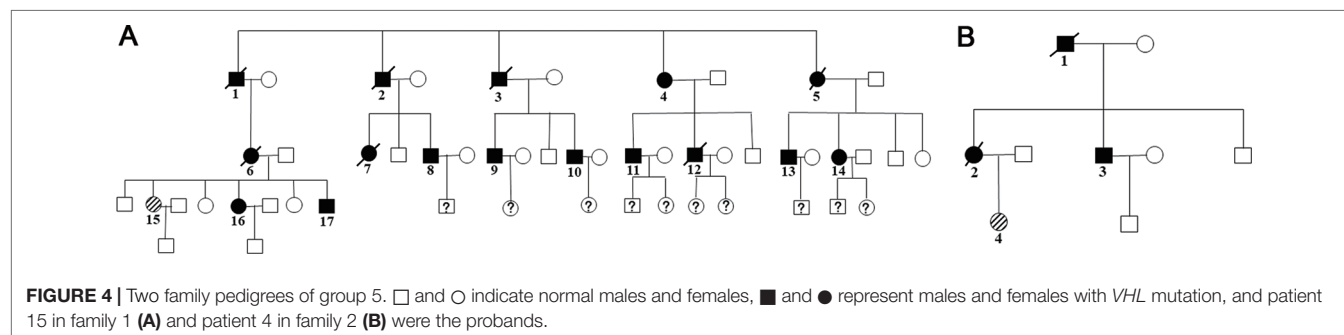


TABLE 3 | Characteristics of genotype and phenotype in group 5.

Family no.	Patient No.	Gender	Age	NC	Age at onset (y)					
					CHB	RA	RCC	PCT	PHEO	GS
1	1	Male	68*	c.262T > C	—	—	65	—	—	—
	2	Male	58*	p.Trp88Arg	—	—	—	—	—	—
	3	Male	63*		62	—	—	—	—	—
	4	Female	66		37	—	—	—	—	—
	5	Female	48*		—	—	—	—	—	—
	6	Female	35*		35	—	—	—	—	—
	7	Female	41*		41	—	—	—	—	—
	8	Male	48		40	—	—	—	—	—
	9	Male	51		35	—	42	—	—	—
	10	Male	44		29	—	—	—	—	—
	11	Male	45		—	—	35	—	—	—
	12	Male	40*		—	—	40	—	—	—
	13	Male	39		39		37			
	14	Female	41		37		37			
	15	Female	32		—	—	30	30	—	—
	(Proband)									
	16	Female	28		—	—	27	—	—	—
	17	Male	23		—	—	—	—	—	—
2	1	Male	58*		48	—	—	—	—	—
	2	Female	45*			—	—	PC	—	—
	3	Male	52		50	—	—	—	50	—
	4	Female	31		—	—	—	—	—	—
	(Proband)									

*Patient was dead. NC, nucleotide change and consequence; PC, pancreatic cancer.

of VHL syndrome in Korean families and concluded that missense mutations in the Hypoxia-inducible factor- α (HIF- α) binding site elevate the age-specific risk for CHB. The studies cited above linked mutations in the *VHL* gene, protein binding sites, and phenotypic diversity to provide insights into the genotype–phenotype correlations based on amino-acid changes in the HIF- α binding site. In this study, we provide a precise summary of the penetrance and overall survival for each of the frequent mutations in the *VHL* gene within the largest Chinese VHL cohort. Our findings provide a more precise and individualized dataset that can be used in genetic counseling and research of the disease pathogenesis.

The current study had several limitations. Von Hippel–Lindau disease is rare, the size of this cohort is relatively small, and the follow-up durations are not sufficiently long, which may influence the correlation analysis between the frequent mutations in the *VHL* gene and the clinical phenotypes. Prospective, large-scale, and long-term follow-up studies are needed to further validate these results. Ultimately, elucidating the genotype–phenotype correlations of VHL disease will help to predict the risk of developing the diverse range of VHL-related phenotypes and their prognoses for individuals with VHL.

ETHICS STATEMENT

The involvement of human participants in this study was based on the Declaration of Helsinki. The ethics of this study were reviewed and approved by the Institutional Ethics Committee of Peking University First Hospital. Informed consent to use clinical data was received from each patient or from their legal guardian.

AUTHOR CONTRIBUTIONS

Conceptualization: BH. Data curation: KM, JCZ, JFZ, JW, and SL. Formal analysis: BH. Funding acquisition: KG. Methodology: ZZ and LC. Project administration and supervision: KG and NZ. Original draft writing: BH.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant 81572506), the Special Health Development Research Project of Capital (grant 2016-2-4074), and the Fundamental Research Funds for the Central Universities (grant BMU2018JI002).

ACKNOWLEDGMENTS

We sincerely thank all the patients and the families in this study for their collaboration and support: Dingfang Bu, Medical Experiment Center, Peking University First Hospital, for his technical assistance and guidance for this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.00867/full#supplementary-material>

REFERENCES

- Bradley, J. F., Collins, D. L., Schimke, R. N., Parrott, H. N., and Rothberg, P. G. (1999). Two distinct phenotypes caused by two different missense mutations in the same codon of the VHL gene. *Am. J. Med. Genet.* 87, 163–167. doi: 10.1002/(SICI)1096-8628(19991119)87:2<163::AID-AJMG7>3.0.CO;2-A
- Butman, J. A., Linehan, W. M., and Lonser, R. R. (2008). Neurologic manifestations of von Hippel–Lindau disease. *JAMA* 300, 1334–1342. doi: 10.1001/jama.300.11.1334
- Chittiboina, P., and Lonser, R. R. (2015). Von Hippel–Lindau disease. *Handb. Clin. Neurol.* 132, 139–156. doi: 10.1016/B978-0-444-62702-5.00010-X
- Chou, A., Toon, C., Pickett, J., and Gill, A. J. (2013). Von Hippel–Lindau syndrome. *Front. Horm. Res.* 41, 30–49. doi: 10.1159/000345668
- Ciotti, P., Garuti, A., Gulli, R., Ballestrero, A., Bellone, E., and Mandich, P. (2009). Germline mutations in the von Hippel–Lindau gene in Italian patients. *Eur. J. Med. Genet.* 52, 311–314. doi: 10.1016/j.ejmg.2009.05.007
- Crossey, P. A., Foster, K., Richards, F. M., Phipps, M. E., Latif, F., Tory, K., et al. (1994). Molecular genetic investigations of the mechanism of tumorigenesis in von Hippel–Lindau disease: analysis of allele loss in VHL tumours. *Hum. Genet.* 93, 53–58. doi: 10.1007/BF00218913
- Gomy, I., Molfetta, G. A., de Andrade, B. E., Ferreira, C. A., Zanette, D. L., Casali-da-Rocha, J. C., et al. (2010). Clinical and molecular characterization of Brazilian families with von Hippel–Lindau disease: a need for delineating genotype–phenotype correlation. *Fam. Cancer* 9, 635–642. doi: 10.1007/s10689-010-9357-2
- Gossage, L., Eisen, T., and Maher, E. R. (2015). VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer* 15, 55–64. doi: 10.1038/nrc3844
- Grubb, R. R., Choyke, P. L., Pinto, P. A., Linehan, W. M., and Walther, M. M. (2005). Management of von Hippel–Lindau-associated kidney cancer. *Nat. Clin. Pract. Urol.* 2, 248–255. doi: 10.1038/ncpuro0179
- Hes, F. J., van der Luijt, R. B., Janssen, A. L., Zewald, R. A., de Jong, G. J., Lenders, J. W., et al. (2007). Frequency of Von Hippel–Lindau germline mutations in classic and non-classic Von Hippel–Lindau disease identified by DNA sequencing, Southern blot analysis and multiplex ligation-dependent probe amplification. *Clin. Genet.* 72, 122–129. doi: 10.1111/j.1399-0004.2007.00827.x
- Hwang, S., Ku, C. R., Lee, J. I., Hur, K. Y., Lee, M. S., Lee, C. H., et al. (2014). Germline mutation of Glu70Lys is highly frequent in Korean patients with von Hippel–Lindau (VHL) disease. *J. Hum. Genet.* 59, 488–493. doi: 10.1038/jhg.2014.61
- Kim, J. J., Rini, B. I., and Hansel, D. E. (2010). Von Hippel Lindau syndrome. *Adv. Exp. Med. Biol.* 685, 228–249. doi: 10.1007/978-1-4419-6448-9_22
- Latif, F., Tory, K., Gnarr, J., Yao, M., Duh, F. M., Orcutt, M. L., et al. (1993). Identification of the von Hippel–Lindau disease tumor suppressor gene. *Science* 260, 1317–1320. doi: 10.1126/science.8493574
- Launbjerg, K., Bache, I., Galanakis, M., Bisgaard, M. L., and Binderup, M. (2017). Von Hippel–Lindau development in children and adolescents. *Am. J. Med. Genet. A* 173, 2381–2394. doi: 10.1002/ajmg.a.38324
- Lee, J. S., Lee, J. H., Lee, K. E., Kim, J. H., Hong, J. M., Ra, E. K., et al. (2016). Genotype–phenotype analysis of von Hippel–Lindau syndrome in Korean families: HIF- α binding site missense mutations elevate age-specific risk for CNS hemangioblastoma. *BMC Med. Genet.* 17, 48. doi: 10.1186/s12881-016-0306-2
- Liu, S. J., Wang, J. Y., Peng, S. H., Li, T., Ning, X. H., Hong, B. A., et al. (2018). Genotype and phenotype correlation in von Hippel–Lindau disease based on alteration of the HIF- α binding site in VHL protein. *Genet. Med.* 20 (10), 1266–1273. doi: 10.1038/gim.2017.261
- Lonser, R. R., Butman, J. A., Huntuon, K., Asthagiri, A. R., Wu, T., Bakhtian, K. D., et al. (2014). Prospective natural history study of central nervous system hemangioblastomas in von Hippel–Lindau disease. *J. Neurosurg.* 120, 1055–1062. doi: 10.3171/2014.1.JNS131431
- Lonser, R. R., Glenn, G. M., Walther, M., Chew, E. Y., Libutti, S. K., Linehan, W. M., et al. (2003). Von Hippel–Lindau disease. *Lancet* 361, 2059–2067. doi: 10.1016/S0140-6736(03)13643-4
- Lonser, R. R., Kim, H. J., Butman, J. A., Vortmeyer, A. O., Choo, D. I., and Oldfield, E. H. (2004). Tumors of the endolymphatic sac in von Hippel–Lindau disease. *N. Engl. J. Med.* 350, 2481–2486. doi: 10.1056/NEJMoa040666
- Maher, E. R., Webster, A. R., Richards, F. M., Green, J. S., Crossey, P. A., Payne, S. J., et al. (1996). Phenotypic expression in von Hippel–Lindau disease: correlations with germline VHL gene mutations. *J. Med. Genet.* 33, 328–332. doi: 10.1136/jmg.33.4.328
- Mete, T., Berker, D., Yilmaz, E., Ozgen, G., Yalcin, Y., Tuna, M., et al. (2014). Clinical presentation of Von Hippel Lindau syndrome type 2B associated with VHL p.A149S mutation in a large Turkish family. *Endocrine* 45, 128–135. doi: 10.1007/s12020-013-9982-2
- Nielsen, S. M., Rhodes, L., Blanco, I., Chung, W. K., Eng, C., Maher, E. R., et al. (2016). Von Hippel–Lindau disease: genetics and role of genetic counseling in a multiple neoplasia syndrome. *J. Clin. Oncol.* 34, 2172–2181. doi: 10.1200/JCO.2015.65.6140
- Nielsen, S. M., Rubinstein, W. S., Thull, D. L., Armstrong, M. J., Feingold, E., Stang, M. T., et al. (2011). Genotype–phenotype correlations of pheochromocytoma in two large von Hippel–Lindau (VHL) type 2A kindreds with different missense mutations. *Am. J. Med. Genet. A* 155A, 168–173. doi: 10.1002/ajmg.a.33760
- Ning, X. H., Zhang, N., Li, T., Wu, P. J., Wang, X., Li, X. Y., et al. (2014). Telomere shortening is associated with genetic anticipation in Chinese Von Hippel–Lindau disease families. *Cancer Res.* 74, 3802–3809. doi: 10.1158/0008-5472.CAN-14-0024
- Nordstrom-O'Brien, M., van der Luijt, R. B., van Rooijen, E., van den Ouweland, A. M., Majoor-Krakauer, D. F., Lolkema, M. P., et al. (2010). Genetic analysis of von Hippel–Lindau disease. *Hum. Mutat.* 31, 521–537. doi: 10.1002/humu.21219
- Ong, K. R., Woodward, E. R., Killick, P., Lim, C., Macdonald, F., and Maher, E. R. (2007). Genotype–phenotype correlations in von Hippel–Lindau disease. *Hum. Mutat.* 28, 143–149. doi: 10.1002/humu.20385
- Patel, R. J., Appukuttan, B., Ott, S., Wang, X., and Stout, J. T. (2000). DNA-based diagnosis of the von Hippel–Lindau syndrome. *Am. J. Ophthalmol.* 129, 258–260. doi: 10.1016/S0002-9394(99)00328-1
- Patocs, A., Gergics, P., Balogh, K., Toth, M., Fazakas, F., Liko, I., et al. (2008). Ser80Ile mutation and a concurrent Pro25Leu variant of the VHL gene in an extended Hungarian von Hippel–Lindau family. *BMC Med. Genet.* 9, 29. doi: 10.1186/1471-2350-9-29
- Peng, S., Shepard, M. J., Wang, J., Li, T., Ning, X., Cai, L., et al. (2017). Genotype–phenotype correlations in Chinese von Hippel–Lindau disease patients. *Oncotarget* 8, 38456–38465. doi: 10.18632/oncotarget.16594
- Stebbins, C. E., Kaelin, W. J., and Pavletich, N. P. (1999). Structure of the VHL–ElonginC–ElonginB complex: implications for VHL tumor suppressor function. *Science* 284, 455–461. doi: 10.1126/science.284.5413.455
- Varshney, N., Kebede, A. A., Owusu-Dapaah, H., Lather, J., Kaushik, M., and Bhullar, J. S. (2017). A review of von Hippel–Lindau syndrome. *J. Kidney Cancer VHL* 4, 20–29. doi: 10.15586/jkcvhl.2017.88
- Vikkath, N., Valiyaveedan, S., Nampoothiri, S., Radhakrishnan, N., Pillai, G. S., Nair, V., et al. (2015). Genotype–phenotype analysis of von Hippel–Lindau syndrome in fifteen Indian families. *Fam. Cancer* 14, 585–594. doi: 10.1007/s10689-015-9806-z
- Wang, J. Y., Peng, S. H., Li, T., Ning, X. H., Liu, S. J., Hong, B. A., et al. (2018). Risk factors for survival in patients with von Hippel–Lindau disease. *J. Med. Genet.* 55, 322–328. doi: 10.1136/jmedgenet-2017-104995
- Wang, J. Y., Peng, S. H., Ning, X. H., Li, T., Liu, S. J., Liu, J. Y., et al. (2017). Shorter telomere length increases age-related tumor risks in von Hippel–Lindau disease patients. *Cancer Med.* 6, 2131–2141. doi: 10.1002/cam4.1134
- Wilding, A., Ingham, S. L., Lalloo, F., Clancy, T., Huson, S. M., Moran, A., et al. (2012). Life expectancy in hereditary cancer predisposing diseases: an observational study. *J. Med. Genet.* 49, 264–269. doi: 10.1136/jmedgenet-2011-100562
- Wu, P., Zhang, N., Wang, X., Ning, X., Li, T., Bu, D., et al. (2012). Family history of von Hippel–Lindau disease was uncommon in Chinese patients: suggesting the higher frequency of *de novo* mutations in VHL gene in these patients. *J. Hum. Genet.* 57, 238–243. doi: 10.1038/jhg.2012.10
- Yoshida, M., Ashida, S., Kondo, K., Kobayashi, K., Kanno, H., Shinohara, N., et al. (2000). Germ-line mutation analysis in patients with von Hippel–Lindau disease in Japan: an extended study of 77 families. *Jpn. J. Cancer Res.* 91, 204–212. doi: 10.1111/j.1349-7006.2000.tb00933.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hong, Ma, Zhou, Zhang, Wang, Liu, Zhang, Cai, Zhang and Gong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders

Alireza Shahryari^{1,2,3,4*}, Marie Saghaeian Jazi^{3,5}, Saeed Mohammadi³, Hadi Razavi Nikoo⁶, Zahra Nazari⁷, Elaheh Sadat Hosseini⁸, Ingo Bartscher^{1,2}, Seyed Javad Mowla^{4*} and Heiko Lickert^{1,2*}

¹ Institute of Diabetes and Regeneration Research, Helmholtz Zentrum München, Neuherberg, Germany, ² Institute of Stem Cell Research, Helmholtz Zentrum München, Neuherberg, Germany, ³ Stem Cell Research Center, Golestan University of Medical Sciences, Gorgan, Iran, ⁴ Department of Molecular Genetics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran, ⁵ Metabolic Disorders Research Center, Golestan University of Medical Sciences, Gorgan, Iran, ⁶ Infectious Disease Research Center, Golestan University of Medical Sciences, Gorgan, Iran, ⁷ Department of Biology, School of Basic Sciences, Golestan University, Gorgan, Iran, ⁸ Department of Nanobiotechnology, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

Ting Li,
University of Arkansas at Little Rock,
United States
Dongying Li,
Oak Ridge Associated Universities,
United States

*Correspondence:

Alireza Shahryari
alireza.shahryari@helmholtz-
muenchen.de
Seyed Javad Mowla
sjmowla@modares.ac.ir
Heiko Lickert
heiko.lickert@helmholtz-muenchen.de

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 26 January 2019

Accepted: 20 August 2019

Published: 25 September 2019

Citation:

Shahryari A, Saghaeian Jazi M,
Mohammadi S, Razavi Nikoo H,
Nazari Z, Hosseini ES, Bartscher I,
Mowla SJ and Lickert H (2019)
Development and Clinical
Translation of Approved
Gene Therapy Products
for Genetic Disorders.
Front. Genet. 10:868.
doi: 10.3389/fgene.2019.00868

The field of gene therapy is striving more than ever to define a path to the clinic and the market. Twenty gene therapy products have already been approved and over two thousand human gene therapy clinical trials have been reported worldwide. These advances raise great hope to treat devastating rare and inherited diseases as well as incurable illnesses. Understanding of the precise pathomechanisms of diseases as well as the development of efficient and specific gene targeting and delivery tools are revolutionizing the global market. Currently, human cancers and monogenic disorders are indications number one. The elevated prevalence of genetic disorders and cancers, clear gene manipulation guidelines and increasing financial support for gene therapy in clinical trials are major trends. Gene therapy is presently starting to become commercially profitable as a number of gene and cell-based gene therapy products have entered the market and the clinic. This article reviews the history and development of twenty approved human gene and cell-based gene therapy products that have been approved up-to-now in clinic and markets of mainly North America, Europe and Asia.

Keywords: gene therapy, cell-based gene therapy, drug, genetic disease, clinic

INTRODUCTION

In medicine, gene therapy is defined as therapeutic strategy that transfers DNA to a patient's cells to correct a defective gene or a gene product in order to treat diseases that are not curable with conventional drugs (Kumar et al., 2016). Direct *in vivo* administration of manipulated viral vehicle for gene delivery and *ex vivo* genetically engineered stem cells are the two principal approaches in advanced clinical gene therapy (Dunbar et al., 2018).

Over the last three decades, clinical gene therapy faced numerous obstacles and a great deal of failures, but it has now accomplished a huge progress in modern medicine and is finding its path into the clinic and the market (Corrigan-Curay et al., 2015), (Friedmann, 2007). In 2017, Luxurna, the first human gene therapy drug for an inherited retinal dystrophy, was approved by Food and Drug Administration (FDA) and entered the US market (Dias et al., 2017). In the same year, Kymriah and Yescarta, two cell-based gene therapies for the treatment of acute lymphoblastic leukemia (ALL),

were also approved by FDA (Butera, 2018; Vormittag et al., 2018). Various outstanding gene and cell-based gene therapies for both rare and common genetic disorders as well as life-threatening diseases, such as cancers and degenerative diseases, are in the evaluation phase prior to their translation into the clinic in the near future (Ehrke-Schulz et al., 2017; Colella et al., 2018). 2017 marks an important year of gene therapy and is considered as a launch point for a new era of modern gene therapy.

In the present review, we summarize the history of development, mechanism-of-action (MOA), target indications as well as primary clinical trials of the twenty so-far approved human gene and cell-based gene therapy products. Additionally, their limitation, safety, manufacturing, dosage and sales are discussed (Figure 1, Table 1).

HUMAN GENE THERAPY PRODUCTS

Vitravene (Fomivirsen)

Vitravene, also called as Fomivirsen is an antisense oligonucleotide (ASO) designed as a therapeutic strategy for cytomegalovirus (CMV) retinitis in HIV-positive patients who have not an option for CMV retinitis treatment (Azad et al., 1993; Anderson et al., 1996). Fomivirsen is the first ever gene-silencing antisense therapy approved for marketing by the FDA. This drug was developed through collaboration between Isis Pharmaceuticals and Novartis Ophthalmics and was approved by FDA in August of 1998 and 1 year later by EMEA (European Agency for the Evaluation of Medicinal Products) to treat cytomegalovirus retinitis (de Smet et al., 1999).

CMV retinitis, a serious viral eye infection of the retina in patients affected by AIDS, is estimated to affect 30% of these patients. Risk of AIDS-related CMV retinitis is almost always directly correlated with lower CD4 counts (less than 200 cells/ μ L) and lower peripheral-blood absolute CD4 lymphocyte counts (50 cells/ μ L or less) (Gallant et al., 1992; Studies of Ocular Complications of AIDS Research Group, 1992).

Fomivirsen is a 21-base phosphorothioate oligodeoxynucleotide which has a CpG motif near its 5' terminal part. It specifically targets the IE-2 mRNA molecule which is encoding a protein required for CMV replication (Anderson et al., 1996; Mulamba et al., 1998). The recommended dose of Fomivirsen is 330 mg intravitreal injection on day 1 and day 15 of treatment for induction, then it is continued 330 mg every 4 weeks. The most frequently observed adverse effects include ocular inflammation and increased intraocular pressure. Half-life clearance of the drug is approximately 55 h in humans vitreous body (Group, 2002a; Uwaydat and Li, 2002).

In clinical studies by Vitravene Study Group, lesion activity regressed in 80% of participants and also it became completely inactive in 55% of participants during Fomivirsen therapy. Different studies indicate that Fomivirsen can successfully ameliorate the symptoms of CMV retinitis (Group, 2002a; Group, 2002b; Group, 2002c; Uwaydat and Li, 2002).

The development of highly active anti-retroviral therapy (HAART) significantly decreased the CMV retinitis incidence by 55–95%. Therefore marketing of Fomivirsen stopped in Europe and the USA in 2002 and 2006 respectively, as a consequence of the low demand. According to the Novartis Ophthalmics,

demand for Vitravene was less than 100 units per year (Deayton et al., 2000; Varani et al., 2000; Kempen et al., 2003).

Gendicine (rAd-p53)

Gendicine gene therapy drug is harboring Tp53 gene which has been developed to treat head and neck squamous cell carcinoma (HNSCC). This recombinant adenovirus was developed by Shenzhen SiBionoGeneTech and was approved by China Food and Drug Administration (CFDA) on October 16, 2003; and found its way to the commercial market in 2004 (Pearson et al., 2004; Peng, 2005).

Inactivation of different tumor suppressors including: TP53, NOTCH1, CDKN2A, PIK3CA and FBXW7 have been reported in HNSCC. Cells harboring those inactivated proteins could divide without control resulting in cell immortalization and increased risk of malignant transformation (Decker and Goldstein, 1982; Nikoo et al., 2017).

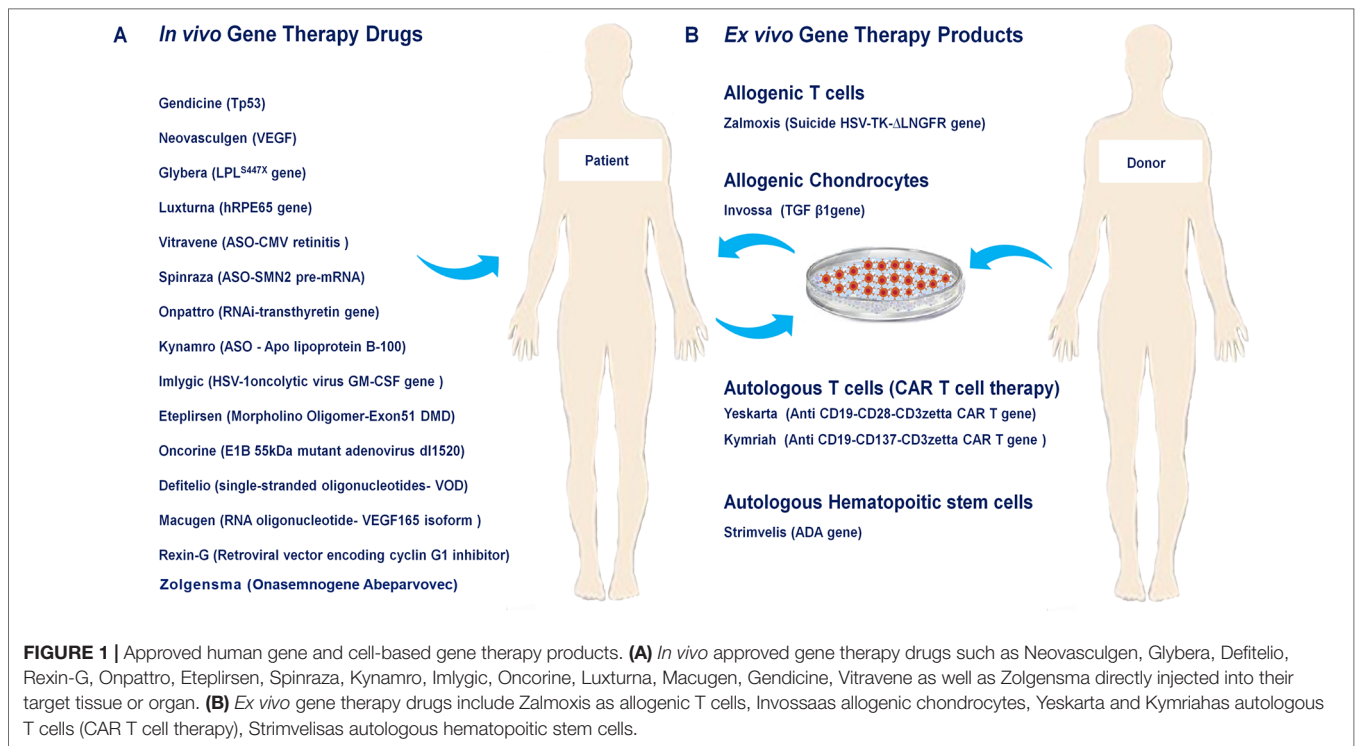
In Gendicine drug, E1 region of human serotype 5 adenovirus (Ad5) has been replaced by human wild-type Tp53. The expression of Tp53 in cancer cells stimulates antitumor properties by initiating apoptotic pathways, suppressing DNA repair and anti-apoptotic events as well as seizing the survival pathways. (Wilson, 2005). The vector is produced in HEK293 cells by co-transfection of the Tp53 expression cassette shuttle vector with an Ad5 genome recombinant plasmid. The cassette contains the Rous sarcoma virus promoter, the wild type human Tp53 gene and a bovine poly-A signal. Upon intratumor injection at a concentration of 1.0×10^{12} viral vector particles per vial, Gendicine binds to the coxsakievirus-adenovirus receptor and enters the tumor cells *via* receptor-mediated endocytosis, expressing ectopic Tp53 gene. The most common side effect with Gendicine is self-limiting fever of 37.5°C to 39.5°C which occurs usually 2 to 4 h after administration lasting for approximately 2 to 6 h (Chen et al., 2014; Li et al., 2015; Zhang et al., 2018).

The initial clinical trial of Gendicine drug was done in four hospitals of Beijing city between 1998 and 2003 years (Han et al., 2003; Wilson, 2005). Also, from 2003 to 2012, totally 16 human clinical studies were carried out for treatment of advanced stages and grades of head and neck cancer, malignant glioma, ovarian cancer, and hepatic cell carcinoma. Treatment with Gendicine resulted in a better overall response and higher survival rate compared to control groups (Chen et al., 2003; Zhang et al., 2003).

In a clinical study with patients suffering from nasopharyngeal cancer, administration of Gendicine in combination with radiotherapy resulted in higher survival rates compared to control groups (Zhang et al., 2005). A clinical trial reported that administration of Gendicine and chemotherapeutic drugs can significantly improve survival rate more than either chemotherapy-only or gene therapy-only groups in patients of advanced oral squamous cell carcinoma (Lu et al., 2004; Zhang et al., 2005).

Macugen (Pegaptanib)

Pegaptanib was developed by Eyetech Pharmaceuticals and PfizerInc with brand name of Macugen. It is a polynucleotide aptamer targeting vascular endothelial growth factor (VEGF165 isoform) for neovascular age-related macular degeneration (AMD) treatment. Pegaptanib was the first anti-angiogenic agent



approved by the USA FDA in December 2004 and was the only therapy for treatment of AMD. It is also the first therapeutic aptamer with RNA structure achieving FDA market approval (Gragoudas et al., 2004).

AMD is the most common cause of severe vision loss and blindness among the aged individuals in the developed world. It is characterized by deterioration of the central part of the retina (Wong et al., 2014). Due to abnormal growth of blood vessels it accounts for 90% of severe vision losses (Fine et al., 2000). It has been suggested that VEGF plays a prominent role in growth and permeability of new vessels in AMD. Anti-VEGF agents are molecular therapies which attempt to block angiogenesis as well as vessel permeability (Friedman et al., 2004; Wong et al., 2008; Bressler, 2009; Solomon et al., 2014).

Pegaptanib is a 28-mer RNA oligonucleotide covalently linked to 2 branched 20-kD polyethylene glycol chains. It specifically binds to VEGF165 isoform at the heparin binding site; thus preventing its binding to VEGF receptors that are located on the vascular endothelial cells surface. VEGF165 has been implicated in pathological ocular neovascularization by increasing vascular permeability and inflammation (Ruckman et al., 1998; Robinson and Stringer, 2001). The recommended dose is 0.3 mg/90 μl of Pegaptanib administered once every 6 weeks by intravitreal injection into the eye. Based on preclinical data, Pegaptanib is metabolized by endo- and exonucleases and is not influenced by the cytochrome P450 system (Drolet et al., 2000; Vinos, 2003).

In two clinical trials involving totally 1186 participants, efficacy of Pegaptanib was determined by the ability of patients to lose less than 15 letters of visual acuity from baseline without

dose-response. The result demonstrated that Pegaptanib is an effective therapy for AMD (Gragoudas et al., 2004; D'Amico, 2005). Moreover, a clinical trial assessed the side effects and the efficacy of Pegaptanib in the treatment of 23 participants suffering from neovascular AMD with previous history of arterial thromboembolic events (ATEs). Pegaptanib did not reveal any systemic or ocular side effects nor did it lead to any recurrent ATEs (Battaglia Parodi et al., 2018).

By 2010, sales of Pegaptanib declined due to better visual outcomes obtained through aflibercept, ranibizumab and bevacizumab drugs which are attested to be less expensive and appeared to be more effective than Pegaptanib (Brown et al., 2009; Sarwar et al., 2016). Nevertheless, Pegaptanib still holds a relatively small market share for the treatment of AMD. Currently, Pegaptanib is marketed by Bausch and Lomb and it approximately costs \$765 per dose.

Oncorine (rAd5-H101)

As the first oncolytic virus approved by CFDA, recombinant human adenovirus type 5 (rAd5-H101) was commercially marketed under the brand name of Oncorine in November 2005 and was manufactured by Shanghai Sunway Biotech. It was initially licensed for the treatment of patients with last-stage refractory nasopharyngeal cancer in combination with chemotherapy following phase III of the clinical trial (Liang, 2018).

Loss of Tp53 gene function is linked with resistance to chemotherapy and reduced survival rate of patients affected by non-small cell cancers of the breast, colon, lung, head and neck as well as ovaries. Therefore TP53 is considered as promising target for gene therapy of NSCC derived cancers. The E1B-55 KD

gene has been totally depleted in Oncorine adenovirus which is responsible for p53 inactivation. Selectively, Oncorine propagates in P53-deficient cancer cells while the adenovirus which lacks the E1b-55KD fails to replicate in normal cells. Following cancer cell lysis, adenoviruses releases and infects neighboring cells initiating a cascade of Oncorine-mediated cell cytotoxicity (Lee et al., 2017).

The first Oncorine clinical trial was carried out on 37 participants with recurrent head and neck carcinoma testing intratumoral and peritumoral injection. The findings were in favor of highly selective tissue destruction, significant tumor regression and no toxicity evidence to injected normal peritumoral tissues (Nemunaitis et al., 2000). Oncorine in combination with chemotherapy in participants with late stages and high grades of different cancers, is administered by intratumor injection. The findings exhibited potential anti-tumor property to refractory aggressive tumors in combined with chemotherapy with minimal toxicity and accepted tolerance of participants (Lu et al., 2004). A phase III clinical study was carried out on 160 participants with head and neck or esophagus squamous cell cancers. Intratumoral injection of Oncorine was followed with cisplatin plus 5-fluorouracil (PF) or adriamycin plus 5-fluorouracil (AF) regimen versus PF or AF regimen alone. The result demonstrated a distinct efficacy and relatively safety of the drug (Xia et al., 2004).

Following the successful phase III clinical trial, the process of commercialization of Oncorine was launched and its efficacy and safety was examined on other cancer types including colon cancer and non-small cell lung cancer (NSCLC) (Reid et al., 2005). Oncorine is not basically associated with major adverse effects except for self-limiting reaction to local injection, mild to moderate grade fever and influenza-like symptoms.

Rexin-G (Mx-dnG1)

Rexin-G is a retroviral vehicle harboring a cytotoxic cyclin G1 construct and is considered the world's first tumor targeting injectable gene therapy vector approved by FDA for metastatic pancreatic cancer. The drug was officially approved by Philippine FDA in December 2007 resulting in progression of clinical studies to Phase III trials in the USA (Gordon and Hall, 2010).

Rexin-G has a hybrid LTR promoter to express cyclin G1. The vector also comprises a neomycin resistance gene which is combined by the SV40 early promoter and is used for vector titer determination. Finally, the Rexin-G is produced by transient co-transfection of three separate vectors in 293T cells (Gordon et al., 2004; Gordon et al., 2006). Rexin-G triggers cell death and apoptosis (by suppressing the cell cycle in G1 phase) in cancer cells. Moreover, it is associated with neovasculature in preclinical studies (Gordon et al., 2006; Chawla et al., 2010).

The result from phase I/II trial of Rexin-G drug in participants with gemcitabine-resistant metastatic pancreatic cancer demonstrated it was well tolerated and safe. Also, elevated survival rate in patients was observed. Progressive clinical development of Rexin-G demonstrated the potential safety and efficacy of this gene therapy product for metastatic solid tumors

that are resistant to standard chemotherapy (Gordon et al., 2006; Chawla et al., 2010).

Neovasculgen (PI-VEGF165)

In 2010, Human Stem Cell Institute of Russia developed Neovasculgen (PI-VEGF165), a plasmid DNA encoding VEGF 165 under the control of a CMV promoter for treatment of atherosclerotic Peripheral Arterial Disease (PAD). The drug was listed in Vital and Essential Drugs (EUVED) of Russian Ministry of Health in 2012 and was then distributed in the Russian market (Deev et al., 2015).

VEGF is a kind of angiogenic effector which triggers cellular proliferation and endothelial migration, angiogenesis as well as enhanced endothelial renovation. These events happen by inducing rapid secretion of nitric oxide and prostacyclin molecules from endothelium stimulating a vasculoprotective effect which is confirmed in some preclinical and clinical studies (Baumgartner et al., 1998; Mäkinen et al., 2002). Neovasculgen recombinant DNA is composed of a transcription start site, the encoding VEGF165 isoform, a polyadenylation signal, a splicing signal and SV40 transcription terminator (Bondar et al., 2015).

The main and only phase 2b/3 multicenter clinical study of Neovasculgen was conducted on 75 patients with PAD. The intramuscular administration of the drug resulted in an increase of pain-free walking distance as well as a significant increase in ankle-brachial index (ABI) and blood flow velocity (BFV). Thus, it was introduced as an effective therapeutic strategy of medium to severe claudication or limping due to chronic lower limb ischemia. (Deev et al., 2015).

Moreover, an international post marketing surveillance study confirmed the safety and efficacy of Neovasculgen in 210 patients with PAD suggesting the absence of no adverse effects (Deev et al., 2017). Authorization of the drug has recently begun in Ukraine, European Medicines Agency (EMA). However, FDA have not yet evaluated or validated the drug probably due to low penetrance of the disease. Neovasculgen would cost nearly \$6,600 per treatment course.

Glybera (Alipogenetiparvovec)

Alipogenetiparvovec, marketed as Glybera, is a gene therapy drug for Lipoprotein Lipase Deficiency (LPLD) treatment. It was developed by Amsterdam Molecular Therapeutics (AMT) in April 2012. In October 2012, the European Commission (EC) approved UniQure as a marketing authorization of Glybera for treating LPLD. Glybera is the first licensed gene therapy product for an inherited disorder in Europe (Bryant et al., 2013). However, given the lack of an appropriate relationship between supply and demand, it was declared a halt to Glybera marketing authorization in Europe in April 2017 (Hampson et al., 2017).

Familial LPLD is an autosomal recessive genetic disorder caused by loss-of-function mutations in LPL gene encoded lipoprotein lipase enzyme. Lack of the enzyme results in LPLD leading to improper digestion of certain fats and massive accumulation of fatty droplets (Faustinella et al., 1991). Glybera

contains LPL gene variant cassette of LPL^{S447X} in a viral vector. The vector consisted of a protein shell part derived from adeno-associated virus serotype I (AAV1), the CMV promoter, a woodchuck hepatitis virus post-transcriptional regulatory element flanked by AAV2 derived inverted terminal repeats (ITRs). Furthermore, recombinant baculovirus technology was used for the Glybera production (Gaudet et al., 2012). Each vial of Glybera comprises of $3 \times 1,012$ genomic copies of alipogene tiparvovec (AAV1-LPL^{S447X}) in 1 ml of a phosphate-based formulation buffer containing 5% sucrose (Bennett et al., 2016). Increased level of creatine kinase in the blood is observed following Glybera injection and it also increased the risk of bleeding and muscle disease in patients with immunodeficiency (Ferreira et al., 2014).

In a clinical trial, 22 subjects were recruited and injected with Glybera; 7 cases demonstrated a decreasing in median plasma triglyceride (TG) by at least 40% over 3 and 12 weeks involvement. The duration of Glybera efficacy did not rise by the immune suppression (Fine et al., 2000; Gaudet et al., 2013). However, in the high-dose cohort, VLDL, total cholesterol and TG content in the VLDL segments were increased at 12th and 52nd weeks post injection. Accordingly, there was a consistent rise in plasma LPL activity despite lack of prolonged effect on total plasma TG and chylomicron metabolism modifications (Andrew E Libby, 2013). In another clinical trial, there was a single group assignment which enrolled five patients with LPLD. TG concentrations decreased in 12 weeks upon treatment. Also, reduction in chylomicrons was observed (Robert Hettle et al., 2017). EC approval for Glybera was based on the results obtained from three Phase III clinical trials conducted in Canada and the Netherlands. Results from 27 patients with LPLD demonstrated that Glybera was well approved in all three clinical trials and no crucial safety signals were noticed. The one-time Glybera administration reduced the frequency of acute pancreatitis (Gaudet et al., 2016).

Glybera, one the highest-priced drug in the world, over \$1.2 million per patient, has been withdrawn from the market because it has turned out to be a commercial loser. Some reporters declared that the drug has only been paid for once since its launch in 2017 (Friedman et al., 2004).

Kynamro (Mipomersen)

Mipomersen, with a market name of Kynamro, is useful as an adjunct therapy for homozygous familial hypercholesterolemia (HoFH) (Raal et al., 2010; McGowan et al., 2012; Stein et al., 2012; Thomas et al., 2013). Mipomersen was developed by Ionis Pharmaceuticals as a novel ASO inhibitor for the cure of HoFH (Crooke et al., 2005). It was rejected by EMA in 2012 due to cardiovascular and liver adverse effects (Mahley, 2001). However, in January 2013, the USA FDA granted approval on its marketing as an orphan drug for the management of HoFH (FDA).

FH is an autosomal dominant genetic disorder caused by mutations in genes of low-density lipoprotein receptor (LDL-R), apolipoprotein B (ApoB) and pro-protein convertase subtilisin/

kexin type 9 (PCSK9). *APOB* gene encodes APOB protein that it is critical for LDL production and delivery (Goldberg et al., 2011; Raal and Santos, 2012; Hovingh et al., 2013).

Mipomersen is an ASO that interfere with the synthesis of ApoB. Mipomersen contains 20 nucleotides that binds to the coding region of the ApoB mRNA in a sequence-specific manner. Thus, it resulted in RNase H-mediated disruption of the mRNA molecule, thereby reducing synthesis of ApoB in the hepatocytes. Mipomersen decreases LDL-C, ApoB, total cholesterol (TC) and non-high-density lipoprotein cholesterol (non-HDL-C) in HoFH patients. Furthermore, Mipomersen causes a dose dependent decreasing of ApoB mRNA in the hepatocytes which is correlated with the reduction of ApoB-containing lipid particles in blood (Raal et al., 2010; Ricotta and Frishman, 2012).

Mipomersen is typically administered at a dose of 200 mg subcutaneously once per week. Around 85% of the drug is bound by Albumin in plasma. Half-life time of the injected drug is approximately 2 to 5 h. In addition, the half-life of the drug in plasma and tissue is approximately 1 to 2 months (Levin et al., 2007). Mipomersen is metabolized and processed initially by tissue endonucleases to generate shorter oligonucleotides available for further metabolism by exonucleases. Due to the risk of hepatotoxicity, Mipomersen is used with caution when prescribed with other LDL-lowering or hepatotoxic medications (Rosie et al., 2009; Crooke and Geary, 2013).

Raal et al. conducted four different phase III trials in various populations of FH patients. The outcomes demonstrated that Mipomersen invariably reduced plasma Lp(a) levels by the 28th weeks by an average of 26.4% compared with the placebo groups (Santos et al., 2015). In phase III clinical trials, the most (84%) commonly-reported adverse effects were injection site reactions including erythema and pruritus, influenza-like symptoms (30%) such as fatigue, pyrexia and chills as well as nausea (14%) (Raal et al., 2010; Stein et al., 2012). In a phase III trial, cardiac events were observed with a high significant frequency in the Mipomersen group in comparison to the placebo (McGowan et al., 2012).

Due to the liver toxicity risk, Mipomersen is only used for patients under the constricted program called Kynamro™ Risk Evaluation and Mitigation Strategy. Single-use 1-ml vial of Mipomersen with a concentration of 200 mg/ml is available (Santos et al., 2015). The average price for 1 week of therapy with the drug is \$6910. Nevertheless, due to adverse events, serious liver toxicity and reactions at injection site, a large proportion of patients discontinued the drug within 2 years.

Imlygic (Talimogenelherparepvec)

Imlygic or Talimogenelherparepvec is a genetically manipulated oncolytic herpes simplex virus type 1 (HSV) that is developed to be used against multiple solid tumors such as unresectable cutaneous, subcutaneous and nodal lesions of melanoma (Chakradhar, 2017). Imlygic was created by BioVex Inc. under the brand name of OncoVEX^{GM-CSF}. The drug was approved by

the USA FDA in October 2015 for targeting melanoma. It was subsequently approved in Europe and Australia in 2016 (Reach, 2015; Printz, 2016; Chakradhar, 2017).

Imlygic is an advanced-generation of double-manipulated HSV-1 oncolytic virus with deletions in the $\gamma 34.5$ and $\alpha 47$ segments, which has been replaced by the human granulocyte-macrophage colony-stimulating factor (GM-CSF) gene (Kohlhapp and Kaufman, 2016). The $\gamma 34.5$ region deletion is mainly accounted for cancer-selective proliferation and reduction of pathogenicity. The $\gamma 34.5$ gene plays a role in inhibiting protein synthesis of host cell upon viral infection. Thus suppressing $\gamma 34.5$ seizes the virus replication in healthy cells. In cancer cells, $\gamma 34.5$ -deficient HSV-1 can still propagate (Johnson et al., 2015; Kohlhapp and Kaufman, 2016). The $\alpha 47$ gene functions to antagonize the host cell transporter associated with antigen presentation. Consequently, the depletion of the gene results in reducing of MHC class I regulation and expression, which enhances the activities of antitumor immune responses. Moreover, two copies of the human GM-CSF gene are incorporated into the virus under the control of the CMV promoter providing high levels of gene expression. Local GM-CSF production by Imlygic is responsible for stimulating the immune system responses (Bommareddy et al., 2017).

Imlygic drug is sold as a sterile, preservative-free solution for intralesional injection, developed at a formal concentration of either 10^6 or 10^8 plaque-forming units (PFU)/ml. Patients receive the drug on days 1 and 15 of each 28-day period for 24 weeks intervention. Each dose should be injected into cutaneous, subcutaneous, and/or nodal lesions that are visible, palpable or detectable by ultrasound guidance. (Fukuhara et al., 2016).

Harboring a herpes virus, Imlygic could be reactive at a later time point, making herpes infections such as cold sores. Imlygic could cause more extensive medical conditions and side effects in patients with impaired immune system (e.g. HIV infected). (Puzanov et al., 2014; Fukuhara et al., 2016).

In several clinical trials, it has been observed that intralesional administration of Imlygic improves immunological response and causes regression in injected lesions (Hu et al., 2006). An analysis of phase II clinical trial outcomes revealed that Imlygic exerted a clear oncolytic effect on injected tumors as well as a secondary immune response as anti-tumor effect on non-injected lesions. Thus, it paved the path for initiation of the phase III of the drug (Senzer et al., 2009; Kaufman et al., 2010). In phase III trial, Imlygic significantly improved durable response rate versus GM-CSF in 436 patients with unresectable advanced-stage melanoma. (Andtbacka et al., 2015).

Exondys 51 (Eteplirsen)

Eteplirsen, was developed by Sarepta Therapeutics under the trade name of Exondys 51. This drug is a 30-mer Phosphomorpholide Morpholino Oligomer (PMO) designed to cause depletion of exon 51 of dystrophin gene. Expression of functional dystrophin protein in patients with duchene muscular dystrophy (DMD) who have mutated DMD gene is amenable by skipping exon 51 (Kole and Krieg, 2015; Stein, 2016). This group of patients cover approximately 13% of all DMD cases, making exon 51 a suitable target for gene targeting (Lim et al., 2017). In September 2016, USA FDA approved

Exondys 51 in an accelerated procedure based on the production of dystrophin in skeletal muscle found in some cases treated with the drug. (U.S. Food and Drug Administration, 2016; Andre et al., 2017)

DMD is a severe X-linked genetic disorder causing a degenerative muscle atrophy and early death. The worldwide incidence is estimated with 1/5000 male births (Mendell and Lloyd-Puryear, 2013; Moat et al., 2013). The disease results from absence of the membrane-associated protein dystrophin, which renders a structural base connecting the cytoskeletal actin in muscle fibers to the extracellular matrix environment (Ervasti, 2007). The DMD gene encoding for dystrophin consists of 79 exons spread over 2.4 Mb region. Exon deletions that occurs more commonly in exons 47 to 63, derange the reading frame of the dystrophin mRNA molecule. This could result in losing protein synthesis in the striated muscle (Kinali et al., 2009).

Eteplirsen targets exon 51 in the dystrophin immature RNA (hnRNA) in DMD patients who carry a deletion between the terminus of exon 50 and the beginning of exon 52 comprising deletion of exons 45–50, 47–50, 48–50, 49–50, 50, 52 or 52–63. Then, the splicing machinery excludes the problematic exon from the final transcript, resulting in production of a shortened functional dystrophin protein (Kinali et al., 2009; Stein, 2016). Lack of charge and no interaction with nucleases have turned Eteplirsen into a highly stable and safe therapeutic drug. Lower protein interaction provides the drug insufficiently activate innate immune response (Summerton and Weller, 1997; Moulton, 2016).

Exondys51 was evaluated in two phase I/II, four phases II and two phase III clinical studies (Cirak et al., 2011; Burki et al., 2015; Geary et al., 2015; Mendell et al., 2016). Results from a confirmatory phase III trial comprising two 80-patient cohorts were required by the FDA to obtain the final approval. After 1 year of treatment, a significant average increase in dystrophin protein levels was seen at 0.22–0.32% of normal levels status. Taken together, clinical studies provided evidence that Exondys51 causes exon skipping, muscle cell penetration and induction of novel dystrophin synthesis. While Exondys51 is now available to applicants, further clinical trials are still required by the FDA to support the clinical benefit of the drug; 2 ml vial (50 mg/ml) of Exondys51 cost around \$1,678 and the average price per patient of the drug stands at \$300,000 annually. (Andre et al., 2017)

Spinraza (Nusinersen)

Nusinersen, commercialized under the name of Spinraza by Biogen, was the first ever medication approved for treatment of spinal muscular atrophy (SMA). Nusinersen was approved by USA FDA in December 2016 and by EMA in May 2017 (Garber, 2016).

SMA is a rare but one of the most common autosomal recessive disorder indicated by progressive disruption of motor neurons of the anterior horn of the spinal cord (Prakash, 2017). Deficiency in the survival motor neuron (SMN) protein is the molecular basis of the disorder (Lefebvre et al., 1995). As an evolutionarily-conserved protein, SMN is encoded by SMN1 and SMN2 genes and is critical for transcriptional regulation, telomerase regeneration and cellular trafficking of motoneurons (Singh et al., 2009). Deletion mutations (particularly in exon 7 and 8) in telomeric copies of the SMN1 gene have been observed in approximately 95% of SMA patients (Sun et al.,

2005). According to the well-defined underlying mechanism of SMA, several genetic-based therapeutic procedures have been defined which primarily aimed to increase the accessibility of SMN protein in motor neurons (D'ydevalle and Sumner, 2015). These approaches included SMN1 gene replacement (Lowes et al., 2017), SMN2 alternative splicing modulation (Zanetta et al., 2014), SMN2 gene activation by previously approved drugs such as Salbutamol (Mercuri et al., 2016), Butyrates (Chang et al., 2001) and Valproic acid (Brichta et al., 2003), SMN stabilization (Mattis et al., 2009), neuroprotection (Kato et al., 2009), as well as stem cell-based therapies (Mercuri and Bertini, 2012).

Nusinersen (Spinraza) is an ASO which targets intron 7 on the SMN2 hnRNA (Castro and Iannaccone, 2014; Communes Internationales Des Substances, 2016), modulating alternative splicing by increasing inclusion of exon 7 in the final processed RNA. This results in higher levels of functional SMN protein in central nervous system (CNS) (Zanetta et al., 2014; Corey, 2017).

Nusinersen is administered intrathecal while performing lumbar punctures under direct supervision of a healthcare professional (Chiriboga, 2017). Following the administration of Nusinersen, it is distributed from the site of injection to motor neurons, vascular endothelial cells as well as glial cells in the CNS tissue (Finkel et al., 2016). The safety and/or tolerability profile of Nusinersen was acceptable in patients with SMA participating in several clinical trials (Finkel et al., 2016; Bertini et al., 2017; Finkel et al., 2017; Mercuri et al., 2017). Yet, the most frequently observed adverse effects of the drug were respiratory complications and elevated urine protein levels (Finkel et al., 2017). Moreover, intrathecal administration limits the treatment to CNS which is crucial for motoneurons but does not target other disorders in the heart, liver, pancreas, intestine and lung organs in SMA patients (Shababi et al., 2014). As such, an optimal treatment to restore SMN protein in peripheral tissues is also needed.

Several therapeutic phase I, II, and III clinical trials demonstrated promising findings and significant improvements in motor milestones (Hoy, 2017). Nusinersen was initially approved on 23 December, 2016 for treatment of the SMA in pediatric and adult patients in USA and is currently in the market. However, Spinraza is belongs to one of the most expensive drugs in the world with the price of \$125000 per injection (Maharshi and Hasan, 2017). Although Spinraza expenses is supported by some health insurance providers in USA and France, Germany, Iceland, Italy and Japan, this expensive drug is not funded in other territories.

Defitelio (Defibrotide)

Defibrotide, commercially known as Defitelio, is manufactured by Jazz Pharmaceuticals plc. Defitelio is a DNA derivative anticoagulant used for patients with hepatic sinusoidal obstruction syndrome/veno-occlusive disease (SOS/VOD) with renal or pulmonary dysfunction following the cytoreductive treatment prior to hematopoietic stem-cell transplantation (HSCT). The efficacy data coming from investigating 528 hepatic VOD participants with renal or pulmonary dysfunction following HSCT, supported approval of Defibrotide by USA FDA

in March 2016 (Richardson et al., 2016). It was also evaluated and approved by EMA in May 2017.

SOS/VOD is an lethal indication of the conditioning regimens for HSCT which is distinguished by harmful hepatomegaly, hyperbilirubinemia, quick weight gain and gathering of ascitic fluid within the stomach (Bearman, 1995). VOD/SOS may also arise in patients treated by chemotherapy or calicheamicin-antibody drug conjugates (Mohty et al., 2015).

Defibrotide as the only approved gene therapy drug accessible for SOS/VOD patients with multi-organ damage (MOD) following HSCT has been associated with promising effects in the United States and the European Union. Defibrotide is a combination of primarily single-stranded oligo DNAs obtained from porcine mucosa tissue by controlled depolymerization with aptameric function on the vascular endothelial cells. It has antithrombotic, thrombolytic, anti-inflammatory and anti-ischemic properties (Francischetti et al., 2012). Defibrotide is considered an adenosine receptor agonist as it has affinity for receptors A1 and A2 on the plasma membranes of the vascular endothelium (Bianchi et al., 1993). It was originally introduced as therapeutic strategy for thrombophlebitis, and as prophylaxis of deep vein thrombosis (DVT) in Italy (Richardson et al., 2013).

Results of early clinical studies demonstrated promising response rates of 36–76% of Defibrotide in SOS/VOD with MOD (Richardson et al., 1998; Chopra et al., 2000; Corbacioglu et al., 2004; Richardson et al., 2010). Based on the result of a phase II clinical trial, 25 mg/kg/day was considered as the Defibrotide dosage with a base duration of 21 days (Richardson et al., 2010). Moreover, in a phase III trial, 356 pediatric participants at high risk of developing SOS/COD post-HSCT were involved to evaluate the prophylactic effects of Defibrotide. The result revealed a significant decreasing in SOS/VOD indication onset by Day +30 post-HSCT for the Defibrotide prophylaxis group in compared with control group (Corbacioglu et al., 2012).

However, several adverse side effects and reactions including coagulopathy, cerebral hemorrhage, hypotension, pulmonary hemorrhage, gastrointestinal hemorrhage and vomiting have been reported. Defibrotide was permitted marketing authorization by the EMA in October 2013 and FDA in March 2016 as the first approved therapeutic method for severe hepatic SOS/VOD post-HSCT indicated in adults and pediatric patients over 1 month of age. Defibrotide is an expensive DNA drug with a wholesale price of approximately \$825 per 200mg or 2.5milliliter vial (daily price for the medicine is \$7425 based on the recommended dose) (Chalandon et al., 2004; Strouse et al., 2016; Veenstra et al., 2017).

Luxturna (Voretigene Neparvovec)

Voretigene Neparvovec-rzyl (AAV2-hRPE65v2), also called Luxturna which is developed and now available on the market by Spark Therapeutics. It is the first USA FDA-approved gene therapy drug for an inherited disease. Approval from FDA and EMA were granted on 19 December, 2017 and 23 November, 2018, respectively. Luxturna is applied intraocularly and is an orphan drug designated for the cure of inherited retinal dystrophy caused by bi-allelic RPE65 mutations (Ramlogan-Steel et al., 2018). This form of inherited retinal dystrophies (IRD), leads to clinical phenotypes of leber congenital

amaurosis type 2 (LCA2) and retinitis pigmentosa type 20 (RP20). The most common form of IRD is retinitis pigmentosa (RP) with the reported impact of 1 in ~4000 individuals. Both LCA2 and RP20 are inherited in an autosomal recessive way. Due to biallelic mutations in the RPE65 gene, its isomerase deficiency destroys the ability of retinal pigment epithelium (RPE) cells to react to the light. Finally, the accumulation of toxic precursors resulted in RPE cells death, progressive visual exacerbation and total blindness (Chung et al., 2018; Miraldi Utz et al., 2018).

Luxturna is applied by a subretinal injection following a vitrectomy where AAV2 targets RPE cells and brings in a normal copy of the RPE65 gene to compensate for the biallelic mutation. Resulting RPE65 protein acting as isomerohydrolase transforms the trans-retinyl esters to 11-cis-retinal, which is the natural ligand and chromophore of the opsins of rod and cones photoreceptors (Russell et al., 2018). In absence of functional RPE65, the opsins in not able to record light or transduce it into electrical responses to induce vision. Functional RPE65 protein resulted in the restoration of the visual cycle by regeneration of 11-cis-retinal (a critical visual pigment component) (Utsav Patel, 2018).

Safety and efficacy of Luxturna were examined in two clinical studies. Phase I trial was a dose-exploration safety study and phase III trial was an efficiency controlled study. 41 participants with mild to advanced vision loss at the time of the first administration took part in the clinical trial program. A statistically significant and clinically meaningful difference was observed in phase III clinical trial between patients ($n = 21$) and control groups ($n = 10$) in 1 year involvement. Over 100-fold improvement was revealed in the original patients group after 1 year. Evaluation by means of the bilateral multi-luminance mobility testing (MLMT) over the follow-up interval of minimum 1 year from the time of administration, determined that 93% (27 of 29) of all phase III trial patients took improvement in their vision function. The luxturna administration does not cause harmful immune responses. Spark announced a list price of \$850,000 per patient, \$425,000 per eye depending on the treatment (Russell et al., 2017; Dias et al., 2018; Russell et al., 2018).

Patisiran (Onpattro)

With the brand name of Onpattro, Patisiran is the only FDA-approved RNA interference (RNAi) drug targeting polyneuropathy caused by hereditary transthyretin-mediated amyloidosis (hATTR) (Adams et al., 2018). The FDA approved this targeted RNA-based drug on August 10, 2018. Alnylam Pharmaceuticals, Inc. (Nasdaq), the leading RNAi therapeutics company, developed this lipid complex drug to treat familial amyloid polyneuropathy (FAP) in adults (Wood, 2018).

FAP, also known as hereditary transthyretin amyloidosis is caused by mutations in the gene encoding transthyretin (TTR), and is an autosomal dominant, progressive, multi systemic and life-threatening disease. In hereditary transthyretin amyloidosis, both wild and mutant-type transthyretin accumulate as amyloid in peripheral nerves, heart, kidney, and gastrointestinal tract giving rise to polyneuropathy and cardiomyopathy. Neuropathic alterations results in intense sensorimotor disruption with failure

of daily life activities and ambulation (Escolano-Lozano et al., 2017; Plante-Bordeneuve, 2018).

Patisiran is a lipid nanoparticle containing an RNAi targeting the transthyretin mRNA. Once Patisiran enters the cell, transthyretin mRNA is cleaved by the RNAi leading to a decrease in circulating transthyretin protein. This reduces the amyloid accumulations which are linked to transthyretin-mediated amyloidosis (Butler et al., 2016). Administered Patisiran targets primarily the liver. Nucleases cut Patisiran to nucleotides of various lengths (Suhr et al., 2015). Vitamin A deficiency is reported to be a main risk of Patisiran use by patients (Kerschen and Planté-Bordeneuve, 2016).

In an early clinical trial, abnormal transthyretin protein production rapidly diminished dose-dependently (Adams et al., 2016). Due to the phase II clinical trial results, Onpattro reduced the level of abnormal transthyretin protein by over 80%. The results also indicated that Patisiran improved neurological symptoms in all 27 patients for more than 24 months (Suhr et al., 2015). The outcomes of a phase III clinical trial revealed that Onpattro treatment lowered the abnormal transthyretin protein levels and improved FAP-related symptoms as well as the quality of life as compared with placebo groups (Adams et al., 2017).

Zolgensma (Onasemnogene Apeparvovec)

Recently, AveXis a drugmaker owned by pharmaceutical giant Novartis, developed Onasemnogene Apeparvovec with the brand name of Zolgensma. It is the most recent authorization gene therapy drug by USA FDA (May 2019). It was previously well-known with compound name AVXS-101. Zolgensma is a proprietary gene therapy strategy for the cure of pediatric patients less than 2 years of age which have mutations in both alleles of *SMN1* gene. Zolgensma has been designed to render a healthy copy of *SMN* gene to seize disease progression through maintenance of *SMN* gene expression with a single, one-time intravenous infusion (Rao et al., 2018; Waldrop and Kolb, 2019; Malone et al., 2019; Mendell et al., 2017).

This drug is a non-replicating recombinant AAV9 containing a functional copy of human *SMN1* gene under the control of CMV enhancer/chicken- β -actin-hybrid promoter (CB) to express *SMN1* in motor neurons of SMA patients. The unique AAV9 capsid is capable to cross the blood-brain barrier allowing efficient CNS delivery by intravenous administration. The modification of the AAV ITR produces a self-complementary DNA molecule that forms a double-stranded transgene which enhances active transcription (Rao et al., 2018; Waldrop and Kolb, 2019).

The efficacy of Zolgensma in SMA patients with bi-allelic *SMN1* gene mutations was investigated in several clinical trials such as STRIVE (NCT03306277) and START (NCT02122952). Bi-allelic *SMN1* gene depletion, two wild type copies of the *SMN2* gene region, as well as lack of the c.859G > C mutation in exon 7 of *SMN2* gene was confirmed in all participants. 21 patients (mean age of 3.9 months) enrolled at the ongoing clinical trial of STRIVE. All the patients received 1.1×10^{11} [4] vg/kg of Zolgensma drug. Comparing results from the clinical study to accessible natural history data of participants

with infantile-onset SMA renders initial evidence of the effectiveness of Zolgensma. The next clinical trial involving 15 patients was named START and involved a low-dose cohort of 3 patients with the mean age of 6.3 months and 12 patients in a high-dose cohort with the mean age of 3.4 months. The low-dose cohort received approximately one-third of the dosage of drug as the high-dose cohort. Comparing results from low- and high-dose cohorts showed a dose-response relationship that provides the clinical use of Zolgensma. (AL-Zaidy et al., 2019; Dabbous et al., 2019)

Elevated liver enzyme of aminotransferases has been reported with Zolgensma therapy. For patients with impaired liver function it is recommended to examine hepatic aminotransferases [aspartate aminotransferase (AST) and alanine aminotransferase (ALT)], total bilirubin and prothrombin before drug infusion. Transient decrease in platelet counts (considered as thrombocytopenia criteria) were revealed at different time points after Zolgensma infusion. Thus, monitoring platelet counts before Zolgensma use and on a regular basis eventually is recommended. Also, transient elevated cardiac troponin-I levels were reported following drug infusion in clinical studies. However, the clinical value of these findings is still unclear (Malone et al., 2019).

The drug also carries a heavy price tag of more than \$2.125 million for a one-time treatment which it is the most expensive gene therapy on the market, yet relative cost-effective.

HUMAN CELL-BASED GENE THERAPY PRODUCTS

Strimvelis (GSK-2696273)

Adenosine deaminase (ADA) deficiency is considered as an autosomal recessive genetic disease causing the severe combined immune deficiency (SCID). ADA enzyme deficiency is the most widespread kind of the SCID accounting for 15% of all patients (Hershfield, 2009). Affected Patients suffer from different metabolic disorders and life-threatening opportunistic infections caused by severe immune deficiency and lymphopenia (Whitmore and Gaspar, 2016). Current therapeutic approaches consist of hematopoietic stem cell transplantation (HSCT), Enzyme replacement therapy (ERT) as well as gene therapy tool.

HSCT from a matched sibling donor will be curative by permanently increasing the overall survival (86%) (Hassan et al., 2012); however, such a donor is available for about 30% of the patients (Tiercy, 2016). In the case of matched unrelated (66%) or haploidentical (43%) donors, less survival will be achieved (Hassan et al., 2012). The other treatment choice is ERT with PEG-ADA which can offer disease relief to the patients but is not a curative method and requires multiple administrations (Booth and Gaspar, 2009). Gene therapy has provided promising treatments for patients affected by ADA-SCID. Gene therapy attempts for ADA-SCID treatment started in early 1990, when the first gene therapeutic trial was carried out at the NIH Clinical Center using gamma-retroviral mediated gene delivery to the autologous peripheral blood lymphocytes of a 4-year-old ADA-SCID affected girl (Ferrua and Aiuti, 2017).

In 2016, the EC approved the GlaxoSmithKline (GSK) stem cell based *ex vivo* gene therapy as a therapeutic option for ADA-SCID. The gene therapy product, also named Strimvelis (GSK2696273), was initially evolved by the San Raffaele Telethon Institute. The manufacturing procedure from HSC transduction to the medicine infusion requires expertise and high standards in product management and, for the time being, is only administered at the San Raffaele Hospital in Milan (Aiuti et al., 2017). Strimvelis consists of autologous hematopoietic stem/progenitor (CD34+) enriched cells transduced *ex vivo* with a retroviral delivery system to express the functional human adenosine deaminase (ADA) cDNA sequence which can replace the enzyme deficiency. At least 4 million purified CD34+ cells/per/kg are required to produce Strimvelis and it is indicated to treat ADA-SCID patients without matched related donor. According to the manufacturer's recommendation, the optimum dose range of Strimvelis (2-20 million CD34+ cells/per/kg) should be administered only once to achieve the best outcome (updated 08/06/2016). Prior to autologous infusion of the gene-transduced CD34+ cell, pre-conditioning with low dose Busulfan (an anti-neoplastic alkylating agent) is necessary (Cicalese et al., 2016). The CD34+ enriched cells will be transduced with retroviral vectors encoding the human ADA cDNA and similar to other retroviral-based gene therapy methods, the potential integration mediated mutagenesis needs to be considered (Cicalese et al., 2016).

In a phase I/II trial, Strimvelis drug was evaluated on totally 18 ADA-SCID children. Evidence of increased immunoglobulin production and increased T cell subtypes (CD3+, CD4+ and CD8+) indicated that Strimvelis infusion can promote both cellular and humoral immunity. Gene modified blood cells were stably present in circulation of treated patients in the post gene therapy phase. Increased ADA enzyme activity and consequent decline in dAXP (metabolite accumulated in ADA deficiency) demonstrated the gene therapy engraftment in treated patients (Cicalese et al., 2016). Following gene therapy, the most commonly observed adverse effects were upper respiratory infections, gastroenteritis as well as rhinitis among the others, with the highest incidence occurring over the time span between pre-treatment up to 3 months into the treatment. Fortunately, no leukemic transformation was observed in the following 13 years (Cicalese et al., 2018).

Zalmoxis (Allogenic T cells encoding LNGFR and HSV-TK)

The allogenic HSCT remains the main therapeutic option for the high risk hematopoietic malignancies. Meanwhile, in the case of haploidentical HSCT, it may end up in failure due to Graft Versus Host Disease (GVHD) (Lorentino et al., 2017). To prevent GVHD in haploidentical HSCT, one option is T cell depletion prior to transplantation which may lead to poor survival due to a severe delay in immune reconstitution (Ciceri et al., 2009). The MolMed product, called Zalmoxis, made it possible to overcome this limitation. The partially-matched donor T cells will be modulated genetically to express HSV-TK (thymidine kinase

enzyme) as an inducible suicide gene. In the case of GVHD initiation, the engineered transplanted T cells can be targeted and killed *via* administration of the pro-drug Ganciclovir (GCV), which is activated to a toxic triphosphate form by HSV-TK enzyme (Denny, 2003).

Zalmoxis contains genetically-modified allogeneic T cells using a retroviral delivery system expressing a shortened human low affinity nerve growth factor receptor (Δ LNGFR) and HSV-TK Mut2 to transduce the allogeneic T immune cells. Δ LNGFR expression cassette was used as the selection marker of the transduced manipulated T cells and the HSV-TK Mut2 expression provides the suicide gene induction if necessary. Infusion of the genetically manipulated donor T cells to HSCT (T cell depleted) transplant patients can simply reconstitute the immunity to protect from infections and confront cancer cells; however, donor cells can potentially target the host cells leading to GVHD. In this case, suicide gene induction by GCV administration can kill the donor T cells expressing HSV-TK and GVHD control (Mohty et al., 2016; Mullard, 2017).

The producer recommended the dose of $1 \pm 0.2 \times 10^7$ cells/kg for the infusion, following 21–49 days after transplantation, without GVHD. The infusion should be repeated monthly up to 4 months to reach the ≥ 100 T cell count/ μ l. The Zalmoxis is not allowed to be administrated in participants younger than 18 years age or in the case of T cell $\geq 100/\mu$ l in circulation (updated 5/09/2016).

The application of the Zalmoxis as adjuvant treatment of leukemia patients in T cell depleted haploidentical stem-cell transplantation was evaluated in a phase I–II clinical trial (NCT00423124). Briefly, totally 28 participants were infused serially with $0.9\text{--}40 \times 10^6$ cells/kg of genetically modified purified donor T cells expressing TK after transplantation monthly up to a maximal point of four infusions. As many as 22 participants obtained immune reconstitution (CD3 + counts ≥ 100 cells/ μ l) 23 days (13–42) following the infusion, while 10 ones evolved acute GVHD (grade I–IV) while one was reported to have chronic GVHD, which was managed by suicide gene induction. The most common adverse effect of TK cell infusion was acute GVHD. Altogether; these findings indicated that Zalmoxis can enhance the immune system reactivation following T cell depleted HSCT, with controllable GVHD (Ciceri et al., 2009).

Engraftment of the TK positive allogeneic donor T cell can activate the host thymopoiesis with elevated systemic IL-7 (early maturation of T-cell cytokine), providing a high number of newly generated TK-negative naive lymphocytes. This finding suggests that even after suicide gene induction and TK-positive T cell elimination to control GVHD, long-term immune reconstitution will reduce the risk of infection after HSCT (no infection was recorded in treated patient of TK007 after 166 days) (Vago et al., 2012). Consistently, results published from the phase III clinical trials (TK008, NCT00914628) carried out in Europe and United States revealed that infusion of genetically engineered allogeneic T cells (MM-TK) (1×10^7 /kg up to four monthly infusions) can increase the 1-year-disease-free survival by 22% in patients (30% vs 52%) (Bonini et al., 2015).

Analysis of 36 individuals treated with Zalmoxis in different trials (22 patients from TK007 trial and 14 patients from the ongoing phase III TK008 trial) and 127 control patients, demonstrated 1 year overall survival (OS) (40% vs 51%, $p = 0.03$) in those patients who survived relapse-free 3 weeks post transplantation. Nevertheless, the relapse probability and leukemia free survival were not significantly different (updated 5/09/2016).

Zalmoxis can provide promising curative improvements for HSCT patients when the matched donor is not available. It can benefit the patients in different aspects including post-transplant GvHD control, Graft versus Leukemia (GvL) improvement and relapse decrease as well as more important long term immune reconstitution leading to reduced infection probability and mortality.

Kymriah (Tisagenlecleucel)

In 2012, Novartis drugmaker, in collaboration with the University of Pennsylvania, began to study chimeric antigen receptor T-cell (CAR T) therapies for acute lymphoblastic leukemia (ALL) treatment. The first approved CAR T-cell-based gene therapy by United States FDA (August 2017) was kymriah (tisagenlecleucel, Novartis Pharmaceuticals, co), which is used for the treatment of children and young adult patients (up to 25 years old) with relapsed B-cell ALL. ALL is the most common malignant tumor diagnosed in children (Belson et al., 2007) and the second most common acute leukemia in adults (Terwilliger and Abdul-Hay, 2017). Despite the advances in common therapeutics of chemotherapy or stem cell transplantation, the relapsed or refractory leukemia still remains a clinical challenge (Kell, 2016).

CAR is an engineered receptor exposing extracellular cancer-specific epitopes (scfv region) linked to the transmembrane and intracellular TCR derived and stimulatory domains. Four generations of CAR T-cells have been developed up to now with different cancer-killing efficiencies and cytokine releasing abilities (Smith et al., 2016). The scfv antibody domain connects to the target antigen in an MHC independent manner leading to CAR clustering and activation of T-cell through intracellular region composing the TCR-derived CD3 ζ chain, without (first generation) or with co-stimulatory domains of the CD28 (generations 3 and 4), OX40/4-1BB (TNF/NGF family, generation 3). The 4th generation of CART proves more effective in solid tumors elimination, also known as TRUCK cells which are induced for secretion of cytokines (IL-12) in target tissue for augmentation of the t cell response and host innate immunity activation for elimination of antigen negative cancer cells (Chmielewski et al., 2014; Chmielewski and Abken, 2015). The clinical efficacy depends on the durable persistence of CAR T-cells assured by co-stimulatory signals protecting the CAR T-cells from 'activation induced cell death' (long term). More importantly, the activated CAR T-cells render target-specific memory cells which it induces inhibiting tumor relapse. The CAR T-cell has been clinically used in leukemia and lymphoma (Hartmann et al., 2017); this is while, clinical trials have shown that it is also applicable to kill solid tumors (Rodriguez et al., 2017). While enjoying approved clinical

efficacy, CAR T-cells therapy causes side effects including on-target off-organ toxicity, cytokine release syndrome (CRS), neurotoxicity, auto-reactivity and other common reactions associated with antibody therapeutic like fever, nausea, and hypotension (Abken, 2015).

Kymriah is composed of autologous T cells suspension that are genetically manipulated with a lentiviral delivery system to produce a CAR comprising of a murine single-chain antibody fragment (scFv) specific for CD19 joint to an intracellular cytoplasmic domain for 4-1BB (CD137) and CD3 zeta with a CD8 transmembrane hinge. The autologous T cells is transduced with a self-inactivating lentiviral vector pseudo-typed with a VSV-G envelope derived from HIV-1 genome in ex-vivo condition. The vector will be integrated into the genome of transduced cells and will result in production of the tisagenlecleucel CAR under the regulation of a constitutively active promoter. After binding to target cells (CD19 expressing cells), the activated tisagenlecleucel CAR will initiate the antitumor activity through CD3 domain. The intracellular 4-1BB costimulatory domain will augment the antitumor reaction and also ensures durable persistence of the CAR T-cells (Updated: 05/07/2018; U.S. Food and Drug Administration, 2017).

As part of safety concerns to avoid replication competent retrovirus (RCR) during Kymriah manufacturing, all needed HIV-1 helper sequences and the pseudo-typed VSV-G envelope sequences are distributed among different constructs with minimum sequence homology, which ensures limited RCR. Moreover, the RCR formation will be checked in the peripheral blood samples of treated patients using VSV-G qPCR (U.S. Food and Drug Administration, 2017). The recommended dosage depends on body weight: $0.2-5 \times 10^6$ CAR-positive viable T cells per kg for $<50\text{kg}$ individuals or $0.1-2.5 \times 10^8$ per kg for $>50\text{kg}$ cases. Similar to other CAR T products, it should be infused after completion of the lymphodepleting chemotherapy (Updated: 05/07/2018).

In a phase II trial, the efficacy of Kymriah was evaluated in 63 pediatric or young patients with relapsed B-cell ALL. The results showed that 81% of patients represent overall remission with no minimal residual disease. The remission was durable with overall survival of 90% for 6 months and 76% after 1 year. The Kymriah was persisted in patients' blood as long as 20 months. The cytokine release syndrome was observed in 77% of patients (Maude et al., 2018). Kymriah was successful in treatment of adult patients with diffused large B-cell Lymphoma (DLBCL), the dominant form of lymphoma. In this global phase II trial, as many as 99 patients (22-76 years old) were infused with Kymriah, formerly CTL019, which resulted in 95% of complete remission in 3 months, sustained at 6 months. The CTL019 was persisted in blood of patients for up to 367 days and CRS was observed in 58% of treated patients (Schuster et al., 2017).

Yescarta (AxicabtageneCiloleucel)

In October 2017, Yescarta (AxicabtageneCiloleucel, Axi-Cel, Kite Pharma, Inc), another CAR T-cell therapy was approved by USA FDA for the treatment of adult patients suffering from aggressive non-Hodgkin lymphoma with the history of at least

two failed systemic therapies. DLBCL is the dominant subtype of blood malignancies with a wide range of clinical and genetic heterogeneity (Belson et al., 2007). The majority of DLBCL new cases respond to a standard care therapy consisting of rituximab and chemotherapy; however, approximately 10–15% of them experience a refractory state (Roberts et al., 2017). According the FDA, Yescarta can be administrated in the case of DLBCL, primary mediastinal large B-cell lymphoma, high-grade B-cell lymphoma and DLBCL resulting from follicular lymphoma (Updated 02/20/2018).

Yescarta is a CD19-directed ex-vivo modulated autologous T cells transfected with gamma-retroviral vector. It expresses a CAR consisting of an extracellular murine anti-CD19 single-chain variable fragment fused to a cytoplasmic domain comprising of CD28 and CD3-zeta co-stimulatory domains. The autologous T-cells harvested from patients by leukapheresis would be shipped to legal centers. The transferred T-cells are then enriched in a closed system in the Yescarta manufacturing center. The activated T-cells (by anti-CD3 and IL-2 treatment) would be transduced with a retroviral vehicle expressing the anti-CD19 CAR gene. Finally, following less than 10 days for manufacturing processes, the Yescarta CAR T-cell is ready for infusion back into the patient (Roberts et al., 2017).

The genetically manipulated autologous CAR T cells can target and eliminate CD19-positive cells when infused back into the patient. 2×10^6 CAR-positive viable T cells per kg body weight is recommended dose which is administrated following a lymphodepleting chemotherapy. It is not designate for the cure of patients with primary CNS lymphoma (Updated 02/20/2018).

In a phase II trial, 101 patients received Yescarta and the result indicated an 82% objective response rate and a 54% complete response rate with a 52% overall rate of survival within 18 months. The common observed side effects were neutropenia (78%) and anemia (43%) (Locke et al., 2017; Neelapu et al., 2017). The Zuma-1 results showed approximately six-fold higher complete remission rates when compared to the achievements of the SCHOLAR-1, the benchmark multi-cohort retrospective study on refractory DLBCL patients outcomes (Neelapu et al., 2016).

Invossa (TissueGene-C)

Invossa (TissueGene-C) has completed phase III trials in the USA and attained marketing approval in Korea by KolonTissueGene as a first-in-class cell mediated gene therapy strategy for the treatment of symptomatic and persistent knee osteoarthritis (OA). It contains 3:1 mixture ratio of non-transformed and retrovirally transduced allogenic chondrocytes that upregulate transforming growth factor $\beta 1$ (TGF $\beta 1$) (Lee, 2018).

Since the beginning of 2015, application of the Invossa progressed rapidly through its clinical trials and commercialization in the USA. To assess the safety and efficacy of Invossa, several phase II and III clinical trials and also post marketing surveillance studies were completed or recruited in Korea and USA. Studies showed that Invossa

TABLE 1 | History and featured data of twenty approved gene and cell based gene therapy products.

Product name	Developer	Approval authority	Therapeutic indication and target tissue	Structure and mechanism of action	Dose	Limitations of use	Price 2018/2019	Clinical trials ID	Key publication
Fomivirsen (Vitravene)	Isis Pharmaceuticals with Novartis Ophthalmics	USA FDA (1998) EMA (1999)	Treatment of CMV retinitis -eye	ASO -inhibition of immediate-early IE-2 protein	330 mg intravitreal injection on day 1 and day 15 of treatment for induction, then 330 mg monthly 1.0×10^{12} vector particles per vial	Not recorded	withdrawn	NCT00002187 NCT00002156 NCT00002356 NCT00002355	Freedman (2001), Group (2002b), Group (2002c), Goldberg et al. (2005)
Gendicine	Shenzhen SIBionoGeneTech	China FDA (2003)	Head and neck squamous cell carcinoma (HNSCC)	Adenovirus contains wild-type p53.P53 production in cancer cells and inhibition of cell proliferation and induce apoptosis.	1.0×10^{12} vector particles per vial	Women at pregnant or during breast-feeding period. Patients with generalized infection and fever of inflammation	\$387 per injection	ChiCTR-TRC-08000094 ChiCTR-TRC-08000092 ChiCTR-TRC-08000091	Li et al. (2015) Zhang et al. (2018)
Pegaptanib (Macugen)	Eyetech Pharmaceuticals and Pfizer, Inc	USA FDA (2004)	Treatment of neovascular age-related macular degeneration - eye	Polynucleotide aptamer - inhibition VEGF165 isoform	0.3 mg once every 6 weeks by intravitreal injection	Need for repeated and frequent intravitreal injection	\$765 per dose	NCT00858208 NCT01245387 NCT00788177 NCT00549055 NCT00460408 NCT00787319 NCT00312351 NCT00327470 NCT01712035 NCT02579564	Gragoudas et al. (2004), D'Amico (2005), Battaglia Parodi et al. (2018)
H101 (Oncorine)	Shanghai Sunway Biotech	China FDA (2005)	Patients with last stage refractory nasopharyngeal cancer(Head and neck cancer), NSCLC lung cancer, Liver cancer, malignant pleural and peritoneal effusion and pancreatic cancer	E1B-55 kDa gene-deleted replication-selective adenovirus	Intratumoral injected at the dose of 1-3 vials per day, combined with 5-FU and cisplatin	Self-relieving local injection, mild to medium grade fever and influenza-like symptoms	–	NCT02579564	Nemunaitis et al. (2000) Xia et al. (2004)
Rexin-G	Epeius Biotechnologies	Philippine FDA (2007) USA FDA (2010)	metastatic solid tumours(pancreatic cancer, Breast cancer, Osteosarcoma, Sarcoma)	Retroviral vector contains cyclin G1. Causes cell death by blocking the cell cycle event in G1 phase and promoting apoptosis in cancer cells	5×10^9 (cfu)/ml	Pregnant or nursing women Fertile patients	\$5000 per bag 500,000 \$ per treatment	NCT00505271 NCT00572130 NCT00505713 NCT00504998 NCT00121745	(Chawla et al., 2010) (Gordon and Hall, 2010)
Neovasculgen (PI-VEGF165)	Human Stem Cells Institute	Russian Ministry of Healthcare (2012)	Atherosclerotic Peripheral Arterial Disease (PAD), including Critical Limb Ischemia (CLI) intramuscular transfer (calf muscles)	Plasmid DNA encoding the 165-amino-acid isoform of human vascular endothelial growth factor (pCMV - VEGF165)	1.2 mg per injection (totally 2.4 mg) In 2 ml injection water Each injection at 4-5 sites at calf muscles	Not recorded	\$6600 for treatment course	NCT03068585 NCT02538705 NCT02369809	(Deev et al., 2015) Deev et al. (2015) Deev et al. (2017) Mäkinen et al. (2002)
Alipogenetiparvovec (Glybera)	Amsterdam Molecular Therapeutics	European Commission (EC) (2012)	Familial LPLD, Leg muscle	AAV1- contains the human lipoprotein lipase (LPL) gene variant LPL ^{S447X}	1×10^{12} gc/kg	higher risk of bleeding and muscle disease in patients with immunodeficiency	over \$1.2 m per patient	NCT02904772 NCT03293810	Bryant et al. (2013) Ferreira et al. (2014)

(Continued)

TABLE 1 | Continued

Product name	Developer	Approval authority	Therapeutic indication and target tissue	Structure and mechanism of action	Dose	Limitations of use	Price 2018/2019	Clinical trials ID	Key publication
Mipomersen, (Kynamro)	ISIS Pharmaceuticals	USA FDA (2013)	Treatment of homozygous familial hypercholesterolemia - Liver, hepatocytes	ASO - Inhibition of the synthesis of apoB100	200 mg once weekly subcutaneous injection	It is not established in patients with hypercholesterolemia and without HoFH	\$6910 For 1-ml vial	NCT00607373 NCT00280995 NCT00477594 NCT00694109	Santos et al. (2015)
Imlygic	Amgen	USA FDA (2016)	Multiple solid tumors (Melanoma, Pancreatic Cancer)	HSV-1 oncolytic virus with deletions in the γ 34.5 and α 47 regions, which GM-CSF gene inserted into the deleted γ 34.5 loci. Tumor Lysis and induce antitumor immune responses	10^6 - 10^8 PFU/ml	Pregnancy, Immunocompromised Patients	\$65000 per treatment	NCT02014441 NCT01740297 NCT02366195 NCT00402025	Hu et al. (2006) Senzler et al. (2009)
Eteplirsen (Exondys 51)	Sarepta Therapeutics	USA FDA (2016)	Treatment of Duchenne muscular dystrophy -Striated muscles	PhosphomorpholidateMorpholino Oligomer (PMO) -Exon skipping	30 mg/kg once weekly by IV infusion	Not recorded	\$1678 For 2 ml vial	NCT02420379 NCT01540409 NCT01396239 NCT00844597 NCT02286947 NCT03218995 NCT02255552 NCT00159250	Kinali et al. (2009) Cirak et al. (2011), Mendell et al. (2013), Geary et al. (2015)
Spinraza (nusinersen)	Biogen Ionis Pharmaceuticals, Inc.	USA FDA EMA (2016)	spinal muscular atrophy (SMA)/motor neurons and CNS	A modified 2'-O-2-methoxyethyl phosphorothioate ASO	12 mg intrathecally	Safe Lumbar puncture complications Intestinal problems Respiratory complications	\$125000 per injection	NCT01494701 NCT01780246 NCT02052791 NCT01703988 NCT02292537 NCT02193074 NCT02594124	Finkel et al. (2016) Mercuri et al. (2017) Finkel et al. (2017) Bertini et al. (2017)
Defibrotide (defibrotide sodium)	Jazz Pharmaceuticals plc	USA FDA EMA (2016)	SOS/VOD with multi-organ dysfunction	Single-stranded oligodeoxyribonucleotides	4 doses per day (every 6 h) of 6.25 mg/kg 2-h Infusion	Hemorrhage Hypersensitivity Reactions Fertility, Pregnancy	\$825 per 200-milligram or 2.5-milliliter vial	NCT00003966 NCT00628498 NCT00358501 NCT00272948	Corbacioglu et al. (2004, 2012) Richardson et al. (1998, 2010, 2016)
Luxturna (VoretigeneNeparvovec-rzyl)	Novartis Inc.	FDA (2017) EMA (2018)	Inherited Retinal Dystrophies (IRD), retinal pigment epithelial (RPE) cells	AAV2 is laboring a normal copy of the RPE65 gene	0.3 ml/eye	Conjunctival hyperemia, cataract, increased intraocular pressure and retinal tear, holes, and inflammation	\$850000 or \$425000 per eye	NCT00999609 NCT03602820	Miraldi et al. (2018) Russell et al. (2017)
Patisiran (Onpattro)	Alnylam Pharmaceuticals Inc.	FDA EMA (2018)	Familial amyloid polyneuropathy (FAP), Liver, Peripheral nerves Heart, Kidney, Gastrointestinal tract	dsRNA encapsulated in aliposome	0.3 mg/kg every 3 weeks (24 weeks)	Infusion-related reactions (IRRs)	\$345000 per 2 mg/ml	NCT01559077 NCT01617967 NCT01961921 NCT01960348	Adams et al. (2018) Adams et al. (2017) Suhr et al. (2015)
Zolgensma (onasemnogene abeparvovec-xioi)	AveXis/ Novartis	USA FDA (2019)	Pediatric individuals less than 2 years of age diagnosed with SMA having bi-allelic mutations in SMN1 gene	Non-replicating rAAV9 containing cDNA of the human SMN gene under the control of CMV enhancer/chicken- β -actin-hybrid promoter	A sterile solution of 2.0×10^{13} vector genomes (vg)/ml [1.1×10^{14} vg/kg (body weight)] by intravenous infusion	The use of the drug in patients with advanced SMA Not recorded	Range (\$2.125-5.0M/ treatment), the average lifetime cost/patient is \$4.2-6.6M	NCT03421977 NCT03505099 NCT02122952 NCT0330627 NCT03461289	AL-Zaidy et al. (2019), Rao et al. (2018), Dabbous et al. (2019), Waldrop and Kolb, (2019), Malone et al. (2019)

(Continued)

TABLE 1 | Continued

Product name	Developer	Approval authority	Therapeutic indication and target tissue	Structure and mechanism of action	Dose	Limitations of use	Price 2018/2019	Clinical trials ID	Key publication
Strimvelis (GSK2696273)	GlaxoSmithKline (GSK)	EMA (2016)	ADA-SCID	Autologous HSC expressing ADA - retroviral vector transduced	2-20 million CD34+ cells per kg should be administered only once	Limitation of use in HCV infected patients (> 15 IU/ml nucleic acid test)	594,000 euros, or \$648000 3 may 2017	NCT00598481	Cicalese et al. (2016) Aiuti et al. (2017) Cicalese et al. (2018)
Zalmoxis	MolMed SPAA	EMA (2016)	Hematopoietic Stem Cell Transplantation Graft vs Host Disease	Allogenic T cell expressing HSV-TK suicide gene- retroviral vector transduced	1 ± 0.2 x 10 ⁷ cells/kg infusion 21-49 days from transplantation, in the absence of GVHD. repeated every month (max 4 times) to reach the ≥100 T cell count per µl.	Not used in children or T cell ≥100/µl in circulation	€149000 Dec 2017 €163900 EUR in Germany: ex-factory price	TK007 TK008	Denny et al. (2003) Ciceri et al. (2009) Vago et al. (2012)
Kymriah (Tisagenlecleucel CTL019)	Novartis Pharmaceuticals Corporation	USA FDA (2017)	Patients up to 25 years of age with B-cell precursor ALL that is refractory or in second or later relapse	Autologous CART cell targeting CD19- Lentiviral vector transduced	0.2-5 x 10 ⁶ CAR-positive viable T cells per kg for <50kg patients or 0.1-2.5 x 10 ⁸ /kg for >50kg cases	lymphodepleting chemotherapy pretreatment is needed Not indicated for treatment of primary CNS lymphoma	\$475000 Feb 2018	NCT02435849 NCT02445248	Schuster et al. (2017) Maude et al. (2018)
Yescarta (Axicabtageneclisoleucel)	Kite Pharma, Incorporated	USA FDA (2017)	Non-Hodgkin lymphoma. B-cell lymphoma, high grade B-cell lymphoma, and DLBCL arising from follicular lymphoma	Autologous CAR T cell targeting CD19 -Retroviral vector transduced	2 x 10 ⁶ CAR-positive viable T cells per kg	lymphodepleting chemotherapy pretreatment is needed Not indicated for treatment of primary CNS lymphoma	\$373000 April 2018	NCT02348216	Roberts et al. (2017) Locke et al. (2017) Neelapu et al. (2017)
INVOSS (TissueGene-C or Tonogenchoncel-L)	TissueGene (now called KolonTissueGene)	South Korea's Ministry of Food and Drug Safety (MFDS) (2017)	Kellgren-Lawrence Grade 3 knee osteoarthritis (OA) intra-articular space of the knee via X-ray guidance	Retrovirally transduced allogenic chondrocytes that upregulates TGF β1	1.8x10 ⁷ cells per 3 ml (mixture of transduced and non-transformed chondrocytes)	Peripheral edema arthralgia joint swelling Injection-site pain	Not recorded	NCT01221441 NCT03412864	Cho et al. (2015) Cherian et al. (2015) Cho et al. (2017) Lee et al. (2018)

treatment can improve knee OA significantly (Cherian et al., 2015). In another phase II clinical trial on patients with a confirmed diagnosis of knee OA, Invossa treatment improved pain, sport activities and quality of daily life (Cho et al., 2015). Additionally, in the randomized double blind, multi-center, placebo-controlled phase III trial on 156 patients with confirmed knee OA, Invossa treatment successfully improved the quality of patients' lives (Cho et al., 2017). In the most recent phase III clinical trial on 163 OA patients, Invossa significantly improved the function and pain relief with structural refinement and mild to severe adverse effects. The most frequent side effects in the Invossa are reported as peripheral edema, arthralgia, joint swelling, and injection-site pain (Lee, 2018).

Finally, KolonTissueGene achieved medical product approval of Invossa in July 2017. A post-marketing surveillance of Invossa injection in patients with OA is now involving to evaluate the safety and effectiveness of this new drug. What's more, several studies have investigated the underlying molecular mechanisms in the effectiveness of Invossa on OA. The majority of studies suggest that Invossa provides an anti-inflammatory surroundings in the arthritic knee joints by means of macrophage polarization which is crucial for modifying the disease (Choi et al., 2017; Lee et al., 2018).

Timeline and the Future Trend of Gene Therapy Market

In 1998, Vitravene was the first clinic antisense gene therapy product for the treatment of CMV retinitis in HIV-infected patients that was approved by USA FDA (de Smet et al., 1999). For targeting HNSCC, a common form of cancer in China, Gendicine was approved in 2003. It is a recombinant human adenovirus expressing Tp53 gene, which entered the Chinese market in 2004 (Pearson et al., 2004; Peng, 2005). Also in 2004, Macugen was approved by USA FDA as the first therapeutic RNA treatment of AMD (Gragoudas et al., 2004). Oncorine is the second approved gene therapy drug in China which is used for cancer treatment (Liang, 2018). The next gene therapy drug targeting pancreas cancer is Rixin-G which was approved by USA FDA in 2007 (Gordon and Hall, 2010). Neovasculgen gene therapy drug is only developed and approved for Russian market in 2012 (Deev et al., 2015). Europe approved their first gene therapy drug, Glybera, in 2012 to treat LPLD. Glybera is nowadays not anymore listed on its developer company product pipeline (Hampson et al., 2017). Kynamro is another antisense drug inhibiting apoB in HoFH patients which was approved by USA FDA in 2013 (Raal et al., 2010; McGowan et al., 2012; Stein et al., 2012; Thomas et al., 2013). Imlygic was approved by the USA FDA for targeting melanoma in 2015 (Reach, 2015; Printz, 2016; Chakradhar, 2017). As the tenth clinical gene therapy product, Exondys51 was approved by USA FDA in 2016 for treatment of DMD patients (U.S Food and Drug Administration, 2016; Andre et al., 2017). Spinraza as the first gene therapy to treat SMA patients was approved by USA FDA in 2016 and by EMA in 2017 (Garber, 2016). Defibrotide was developed to treat patients with hepatic SOS/VOD prior

to HSCT and was approved by USA FDA in March 2016 (Richardson et al., 2016). In 2016, Strimvelis, was approved in Europe a gene therapy for treating ADA-SCID patients. 2017 was promising year for stem cell based gene therapy (Hershfield, 2009), (Whitmore and Gaspar, 2016). The United States finally approved *ex vivo* CAR-T therapies. Kymriah is considered the first CAR-T therapy targeting ALL which was approved by USA FDA in August of 2017. As second CAR-T therapy product, Yescarta, treating adult patients with diffuse large B-cell lymphoma, was approved by USA FDA in October of 2017 (Roberts et al., 2017; Terwilliger and Abdul-Hay, 2017). Zalmoxis and Invossa are stem cells based gene therapy products than entered the clinic in 2016 and 2017 respectively (Lorentino et al., 2017) (Lee, 2018). Spark Therapeutics has been granted an FDA approval for their Luxturna gene therapy drug in 2018. This pharma company utilizes an AAV system to deliver RPE65 gene into the patients eye suffering from retinal dystrophy caused by RPE mutations (Ramlogan-Steel et al., 2018). As the first RNAi drug, Patisiran was approved by FDA in 2018 (Adams et al., 2018). In May 2019, Zolgensma was approved by USA FDA as the twentieth approved gene therapy products until now (**Figure 2**). Finally, at the beginning of 2019, Zaynteglo, also known as Lentiglobin got conditional approval by EMA for beta thalassemia. Furthermore, BMN 270 a drug for hemophilia A, and GT-AADC a product for AADC deficiency are two highlighted gene therapy products that may be approved until 2020.

Table 2 summarized more gene therapy clinical trials that are currently ongoing for advanced stages of trials and/or approval.

In recent years a number of gene therapies based on gene editing tools, especially CRISPR system, advanced to human clinical trial stage. Gene editing technologies including CRISPR, Zinc Finger Nuclease (ZFN) and TALEN allow scientists to undertake precise genomic modifications at desired human genome positions yielding tremendous beneficial results in modern medicine and the field of genetics. Ongoing human gene therapy trials based on gene editing systems are listed in **Table 3**. The upcoming trials mark the maturation of the gene editing tools into a clinical-grade technology.

Recent advances in understanding molecular mechanism of human diseases and treatment are boosting the global gene therapy market. This market is categorized into cancers, neurological diseases, rare genetic diseases, cardiovascular disorders, and infectious diseases. Cancers and monogenic diseases had the highest market share in recent years respectively. Viral vectors (mainly retrovirus, lentivirus, adeno-associated virus) in comparison to non-viral vectors are the preferred gene therapy vehicles in the clinic. High-efficiency of gene transduction, specific gene delivery and targeting, safety and reduced administration dose are the main benefits of viral vectors. North America and Europe are dominant players and drive advancements in gene therapy market of cancer and rare genetic diseases. Nowadays, the key trends of the gene therapy market are high prevalence of human cancers and genetic disorders, clarifying gene therapy guidelines

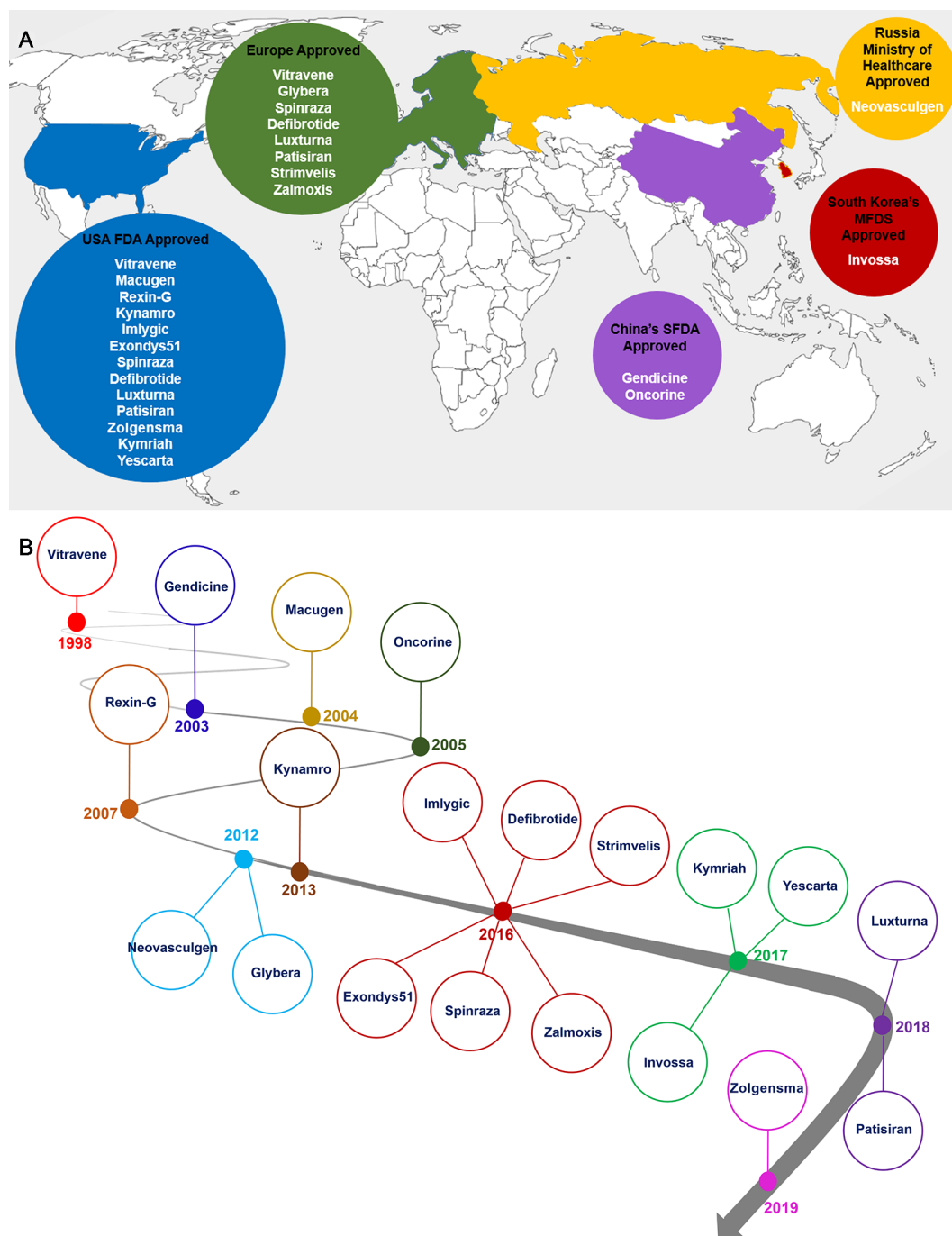


FIGURE 2 | Distribution countries of approval gene therapy drugs **(A)** and timeline of gene therapy development **(B)**. As it is shown the North America and Europe are main parts of the gene therapy development and market **(A)**. Gene therapy market began with approval of Vitravene drug in 1998, and it was continued with approval of Zolgensma drug in 24 May, 2019. 2016 and 2017 years were promising points in gene therapy market since near 10 gene therapy products such as Imlygic, Defibrotide, Spinraza, Zalmoxis, Exondys51, Strimvelis, Invossa, Yescarta and Kymriah were approved by relevant authority **(B)**.

and rising financial support of gene therapy and cell-based gene therapy in clinical trials. However, safety and efficacy problems, prolonged laboratory procedures for conducting clinical studies, unknown product interactions with host, and high cost of gene therapy drugs are major barriers in the way of gene therapy market.

CONCLUSION

Despite considerable efforts in gene therapy segment, only a few of the twenty approved gene and cell-based gene therapy products were translated into the clinic (May 2019). In the previous year, numerous promising results attested progress in clinical

TABLE 2 | Highlighted ongoing gene therapy products.

Product name	Developer/ sponsor	Structure and mechanism of action	Therapeutic indication and target tissue	Clinical trials ID
Zynteglo (Lentiglobin BB305)	Bluebird Bio	<i>Ex vivo</i> transplantation of hematopoietic stem cells manipulated by a lentiviral vehicle transferring HBB	Beta thalassemia (conditional approved by EMA)	NCT01745120 NCT02151526 NCT03207009 NCT02140554 NCT02906202 NCT03207009
BMN 270 (Valoctocogene roxaparvovec)	BioMarin Pharmaceutical	An AAV vector transferring factor VIII in hemophilia A individuals with Residual FVIII Levels ≤ 1 IU/dl	Hemophilia A	NCT03392974 NCT03370913 NCT02576795 NCT03520712
AAV2-hAADC	PTC Therapeutics	AAV serotype 2 expressing human AADC (hAADC)	Aromatic L-amino Acid Decarboxylase (AADC9 deficiency AADC	NCT02926066 NCT02852213 NCT02399761 NCT02926066 NCT01395641
Lenti-D	Bluebird Bio	<i>Ex vivo</i> Transplantation of CD34+ stem cells manipulated by lentiviral vector to contain a functional ABCD1 gene	Cerebral adrenoleukodystrophy (CALD), ABCD1 gene	NCT03852498 NCT01896102 NCT02698579
NSR-REP1 (AAV2-hCHM)	Spark Therapeutics Nightstar Therapeutics	An AAV2 vector transferring REP1 gene into the eye. The drug is delivered surgically by injection into the sub-retinal space	Retinal Gene Therapy for choroideremia	NCT02553135 NCT02077361 NCT03507686 NCT03496012 NCT01461213 NCT03584165
ADVM-043 (AAVrh.10hA1AT)	Adverum	An AAV expressing Alpha-1 Antitrypsin (A1AT) gene to patients with A1AT deficiency	A1AT Deficiency	NCT02168686 NCT03804021
LYS-SAF302	Lysogene, Regenxbio, Sarepta Therapeutics	AAVrh10 virus expressing N-sulfoglucosamine sulfohydrolase (SGSH)	Sanfilippo type A syndrome	NCT03612869 NCT02746341
AMT-061	uniQure	An AAV5 viral vector expressing the Padua variant of Factor IX (FIX-Padua)	Hemophilia B	NCT03569891 NCT03489291
Fidanacogene elaparvovec (SPK-9001)	Pfizer	An AAV with engineered capsid that expresses human coagulation factor IX gene	Hemophilia B	NCT03587116 NCT03307980
OTL-103 (GSK2696275)	Orchard Therapeutics	<i>Ex vivo</i> autologous transplantation of hematopoietic stem cells manipulated by a lentiviral vehicle expressing WAS gene	Wiskott Aldrich syndrome	NCT01515462 NCT03837483
OTL-200 (GSK2696274)	Orchard Therapeutics	<i>Ex vivo</i> autologous transplantation of CD34+ stem cells manipulated by lentiviral vector expressing human arylsulfatase A (ARSA) gene	Metachromatic Leukodystrophy (MLD)	NCT03392987
Instiladrin (rAd-IFN/Syn3)	FKD Therapies, Ferring Pharmaceuticals	Nonreplicating recombinant adenovirus type 5 expressing interferon alpha-2b (IFN α 2b) gene, fused with the excipient Syn3 (rAd-IFN/Syn3)	Bacillus Calmette-Guérin (BCG)- unresponsive nonmuscle invasive bladder cancer (NMIBC)	NCT02773849
OTL-101	Orchard Therapeutics	Lentiviral transduced CD34+ cells to express ADA gene	Adenosine deaminase severe combined immune deficiency (ADA-SCID)	NCT03765632
RT-100 (Ad5.hAC6)	Renova Therapeutics	AAV5 vehicle expressing human adenylyl cyclase type 6 which it is directly injected throughout cardiac catheterization into the arteries feeding the heart	Heart failure and reduced ejection fraction	NCT03360448 NCT00787059
ADXS11-001 (Axalimogene filolisbac)	Advaxis	Targeted immunotherapy /attenuated <i>Listeria monocytogenes</i> manipulated for secretion of antigen/ adjuvant fusion proteins	Metastatic cervical cancer	NCT02853604 NCT02002182 NCT02164461

(Continued)

TABLE 2 | Continued

Product name	Developer/ sponsor	Structure and mechanism of action	Therapeutic indication and target tissue	Clinical trials ID
OXB-301 (MVA-5T4)	Oxford BioMedica	An attenuated engineered vaccinia virus Ankara that delivers the 5T4 oncofetal antigen gene to stimulate the immune cells against cancer cells	Ovarian cancer, colorectal cancer	NCT01556841 NCT01569919
Pexa-Vec (JX-594)	SillaJen, Transgene	Wyeth strain vaccinia virus engineered to directly lyse tumor cells and stimulate anti-tumor immunity	Hepatocellular carcinoma	NCT02562755 NCT03071094
SPK-8011	Spark Therapeutics	Manipulated AAV vehicle (to specifically transfer the human factor VIII (hFVIII) gene into liver	Hemophilia A	NCT03003533 NCT03432520 NCT03876301
Toca 511 and Toca FC (vocimagene amiretrorepvec)	Tocagen, ApolloBio	Toca 511 is a retroviral vector expressing cytosine deaminase. Toca FC is prodrug 5-fluorocytosine which is converted into 5-fluorouracil by CD	Glioblastoma Multiforme And Anaplastic Astrocytoma	NCT02414165 NCT02576665 NCT02598011
VB-111 (ofranergene obadenovec)	Vascular Biogenics	Targeted anticancer gene-based biologic administered as an IV infusion once every 2 months. VB-111 was developed through VBL's Vascular Targeting System (VTS™). VBL says the mechanism combines blockade of tumor vasculature with an anti-tumor immune response	Solid tumors, including recurrent platinum-resistant ovarian cancer	NCT03398655
VGX-3100	Inovio Pharmaceuticals	A synthetic DNA vehicle targeting the E6 and E7 proteins of HPV types 16 and 18, also is considered as a DNA vaccine	Cervical high-grade squamous intraepithelial lesion infected by HPV subtypes of 16 and 18	NCT03185013 NCT03721978 NCT03603808 NCT03499795 NCT03180684.
VM202	ViroMed	DNA vehicle expressing two isoforms of hepatocyte growth factor, HGF728 and HGF723	Painful diabetic peripheral neuropathy, Chronic nonhealing ischemic foot ulcer in diabetes, Critical limb ischemia, Amyotrophic lateral sclerosis, Acute myocardial infarction	NCT02563522 NCT02427464 NCT03363165 NCT03404024
QTA020V	Astellas	A rAAV transfects retinal ganglion cell to express BDNF and one of its receptors (TrkB) resulting reduction in cell death	Glaucoma	—
EB-101	Abeona Therapeutics	This drug was designed to transfer a healthy copy of COL7A1 gene using the patients' own skin cells	RDEB (skin disease)	NCT01263379
AT132	Audentes Therapeutics	An AAV8 vector containing a functional copy of the MTM1 gene	X-Linked Myotubular Myopathy (ASPIRO)	NCT03199469
NSR-RPGR	Biogen/Nightstar	An AAV8 encoding retinitis pigmentosa GTPase Regulator (RPGR)	X-linked retinitis pigmentosa (XLRP)	NCT03116113
Generx (Ad5FGF-4)	Angionetics, Huapont Life Sciences	A human serotype 5 adenovirus that express fibroblast growth factor-4 (FGF-4) gene	Angina, Stable	NCT02928094 NCT01550614 NCT00346437
GS010	GenSight Biologics	AAV2 vector that expresses the human wild-type ND4 protein	Leber Hereditary Optic Neuropathy (LHON) caused by mutation of the ND4 gene	NCT02652780 NCT02652767 NCT03293524 NCT03406104

gene therapies for monogenic diseases, inherited blindness, certain inherited neurodegenerative diseases, metabolic genetic disorders and a number of bone marrow and lymph nodes cancers (Dunbar et al., 2018).

Growth and development of viral delivery systems emerge as effective tools for gene manipulation and gene therapy approaches such as CRISPR/Cas has revolutionized the realm of gene therapy (Dicarlo et al., 2017; Ehrke-Schulz et al., 2017).

Although Glybera was withdrawn from the European market in early 2017, approval of six gene therapy and cell-based gene therapy products, known as Invossa, Luxturna, Kymriah, Yescarta, Patisiran, and recently approved Zolgensma (May 2019), promise a new era in gene therapy for untreatable genetic disorders.

As such, the global gene therapy market has grown commensurately in recent years and is expected to grow at a high rate through 2030 according to Grand View Research. A

TABLE 3 | Recruited clinical trials based on genome editing technologies (e.g. CRISPR/Cas, ZFN and TALEN).

Product name	Developer/ sponsor	Structure and mechanism of action	Therapeutic indication and target tissue	Clinical trials ID
AGN-151587 (EDIT-101)	Allergan	An AAV5 vector employed in EDIT-101 contains two gRNAs and Cas9 to correct IVS26 mutation in CEP290 gene	Leber Congenital Amaurosis 10 (LCA10)	NCT03872479
Cyclophosphamide (PD-1 Knockout T Cells)	Sichuan University	<i>Ex vivo</i> gene manipulation of T cells with CRISPR/Cas system to target non-small cell lung cancer	Metastatic NSCLC	NCT02793856
PD-1 Knockout T Cells	Yang Yang	<i>Ex vivo</i> gene manipulation of Peripheral blood lymphocytes with CRISPR/Cas system to knockout PD1 gene. Then EBV-CTL will be produced (PD-1 Knockout EBV-CTL)	Advanced-stage EBV-associated malignancies	NCT03044743
PD-1 Knockout T Cells	Hangzhou Cancer Hospital	<i>Ex vivo</i> gene manipulation of T cells with CRISPR/Cas system to target esophageal cancer	Esophageal Cancer	NCT03081715
CTX001	Vertex Pharmaceuticals Incorporated	Autologous CD34+ Hematopoietic Stem and Progenitor Cells (hHSPCs) manipulated with CRISPR/Cas9 system at the enhancer position of the BCL11A gene leads to increase in fetal hemoglobin	Hematologic diseases (Hemoglobinopathies)	NCT03655678 NCT03745287
SB-FIX	Sangamo Therapeutics	An AAV2/6 virus with ZFN inserting Factor 9 gene under the control of albumin promoter for liver expression	Hemophilia B	NCT02695160
ST-400	Sangamo Therapeutics	Autologous Hematopoietic Stem Cell Transplant/ST-400 is composed of the patient's own blood stem cells which are genetically modified with ZFN technology to disrupt a precise and specific sequence of the enhancer of the BCL11A gene	Transfusion Dependent Beta-thalassemia	NCT03432364
CCR5 gene modification	Affiliated Hospital to Academy of Military Medical Sciences	CD34+ hematopoietic stem/progenitor cells from donor are manipulated with CRISPR/Cas9 aiming CCR5 gene deletion from cell surface	HIV-1-infection	NCT03164135
SB-318	Sangamo	An AAV2/6 viral vehicle combined with ZFN to incorporated IDUA under the control of albumin promoter for liver expression	MPS I (Hurler syndrome)	NCT02702115
SB-913	Sangamo	An AAV2/6 viral vehicle linked with ZFN tool to incorporated IDS gene under the control of albumin promoter in liver cells	MPS II (Hunter's syndrome)	NCT03041324
UCART123	Cellectis S.A.	Allogeneic engineered T-cells with TALEN system expressing anti-CD123 chimeric antigen receptor	AML	NCT03190278
UCART019	Chinese PLA General Hospital	CRISPR/Cas9 mediated CAR-T Cells Targeting CD19 in individuals with relapsed or refractory CD19+ leukemia and lymphoma	leukemia and lymphoma	NCT03166878
Anti-Mesothelin CAR-T cells	Chinese PLA General Hospital	Knocking out of PD-1 and TCR genes using CRISPR/Cas in CAR T Cells	Solid Tumor, Adult	NCT03545815
NY-ESO-1 redirected autologous T cells with CRISPR edited endogenous TCR and PD-1	University of Pennsylvania	Autologous T cells manipulated with a lentiviral vector for expressing NY-ESO-1 and transfected with CRISPR guide RNA to for knocking out of endogenous TCR α , TCR β and PD-1 (NYCE T Cells)	Multiple Myeloma	NCT03399448
Universal Dual Specificity CD19 and CD20 or CD22 CAR-T Cells	Chinese PLA General Hospital	Using CRISPR/Cas9 to target CD19 and CD20 or CD22 in CAR-T Cells for treatment of Relapsed or Refractory Leukemia and Lymphoma	Leukemia Lymphoma	NCT03398967

recent report published by Roots Analysis on 'Gene Therapy Market (2nd Edition), 2018-2030', stated that nearly 300 product candidates are currently under various stages of development for a diverse range of applications.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contribution to the work and approved it for publication.

REFERENCES

- Updated 02/20/2018. U.S. Food & Drug Administration, YESCARTA (axicabtagene ciloleucel), Product information [Online]. Available: <https://www.fda.gov/BiologicsBloodVaccines/CellularGeneTherapyProducts/ApprovedProducts/ucm581222.htm> [Accessed].
- Updated 5/09/2016. European Medicines Agency, Zalmoxis (Allogeneic T cells genetically modified with a retroviral vector encoding for a truncated form of the human low affinity nerve growth factor receptor (DLNGFR) and the herpes simplex 1 virus thymidine kinase (HSV-TK Mut2)), Product Information [Online]. Available: <https://www.ema.europa.eu/en/medicines/human/EPAR/zalmoxis> [Accessed].
- updated 08/06/2016. European Medicines Agency, Strimvelis (autologous CD34+ enriched cell fraction that contains CD34+ cells transduced with retroviral vector that encodes for the human ADA cDNA sequence), Product information [Online]. Available: <https://www.ema.europa.eu/en/medicines/human/EPAR/stirimvelis> [Accessed].
- Updated: 05/07/2018 U.S. Food & Drug Administration, KYMRIAH (tisagenlecleucel), product information [Online]. Available: <https://www.fda.gov/BiologicsBloodVaccines/CellularGeneTherapyProducts/ApprovedProducts/ucm573706.htm> [Accessed].
- Abken, H. (2015). Adoptive therapy with CAR redirected T cells: the challenges in targeting solid tumors. *Immunotherapy* 7, 535–544. doi: 10.2217/imt.15.15
- Adams, D., Cauquil, C., Labeyrie, C., Beaudonnet, G., Algalarrondo, V., and Théaudin, M. (2016). 11TTR kinetic stabilizers and TTR gene silencing: a new era in therapy for familial amyloidotic polyneuropathies. *Expert Opin. Pharmacother.* 17, 791–802. doi: 10.1517/14656566.2016.1145664
- Adams, D., Gonzalez-Duarte, A., O’rordan, W. D., Yang, C.-C., Ueda, M., Kristen, A. V., et al. (2018). 1Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *N. Engl. J. Med.* 379, 11–21. doi: 10.1056/NEJMoa1716153
- Adams, D., Suhr, O. B., Dyck, P. J., Litchy, W. J., Leahy, R. G., Chen, J., et al. (2017). 12Ttrial design and rationale for APOLLO, a Phase 3, placebo-controlled study of patisiran in patients with hereditary ATTR amyloidosis with polyneuropathy. *BMC Neurol.* 17, 181–181. doi: 10.1186/s12883-017-0948-5
- Aiuti, A., Roncarolo, M. G., and Naldini, L. (2017). Gene therapy for ADA-SCID, the first marketing approval of an *ex vivo* gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products. *EMBO Mol. Med.* 9, 737–740. doi: 10.15252/emmm.201707573
- Al-Zaidy, S., Pickard, A. S., Kotha, K., Alfano, L. N., Lowes, L., Paul, G., et al. (2019). Health outcomes in spinal muscular atrophy type 1 following AVXS-101 gene replacement therapy. *Pediatr. Pulmonol.* 54, 179–185. doi: 10.1002/ppul.24203
- Anderson, K. P., Fox, M. C., Brown-Driver, V., Martin, M. J., and Azad, R. F. (1996). Inhibition of human cytomegalovirus immediate-early gene expression by an antisense oligonucleotide complementary to immediate-early RNA. *Antimicrob. Agents Chemother.* 40, 2004–2011. doi: 10.1128/AAC.40.9.2004
- Andre, M. S., Johnson, M., Bansal, P. N., Wellen, J., Robertson, A., Opsahl, A., et al. (2017). A mouse anti-myostatin antibody increases muscle mass and improves muscle strength and contractility in the mdx mouse model of Duchenne muscular dystrophy and its humanized equivalent, domagrozumab (PF-06252616), increases muscle volume in cynomolgus monkeys. *Skeletal Muscle* 7, 25. doi: 10.1186/s13395-017-0141-y
- Andrew E Libby, H. W. (2013). 33An update on gene therapy for the treatment of lipoprotein lipase deficiency. *DovePress* 4, 47–54. doi: 10.2147/ODRR.S39979
- Andtbacka, R. H., Kaufman, H. L., Collichio, F., Amatruda, T., Senzer, N., Chesney, J., et al. (2015). Talimogene laherparepvec improves durable response rate in patients with advanced melanoma. *J. Clin. Oncol.* 33, 2780–2788. doi: 10.1200/JCO.2014.58.3377
- Azad, R. F., Driver, V., Tanaka, K., Crooke, R., and Anderson, K. (1993). Antiviral activity of a phosphorothioate oligonucleotide complementary to RNA of the human cytomegalovirus major immediate-early region. *Antimicrob. Agents Chemother.* 37, 1945–1954. doi: 10.1128/AAC.37.9.1945
- Battaglia Parodi, M., Di Bartolo, E., Brue, C., Cappello, E., Furino, C., Giuffrida, S., et al. (2018). Pegaptanib: choroidal neovascularization in patients with age-related macular degeneration and previous arterial thromboembolic events. *Eur. J. Ophthalmol.* 28, 58–62. doi: 10.5301/ejo.5001060
- Baumgartner, I., Pieczek, A., Manor, O., Blair, R., Kearney, M., Walsh, K., et al. (1998). Constitutive expression of phVEGF165 after intramuscular gene transfer promotes collateral vessel development in patients with critical limb ischemia. *Circulation* 97, 1114–1123. doi: 10.1161/01.CIR.97.12.1114
- Bearman, S. I. (1995). The syndrome of hepatic veno-occlusive disease after marrow transplantation. *Blood* 85, 3005–3020.
- Belson, M., Kingsley, B., and Holmes, A. (2007). Risk Factors for Acute Leukemia in Children: A Review. *Environ. Health Perspect.* 115, 138–145. doi: 10.1289/ehp.9023
- Bennett, J., Wellman, J., Marshall, K. A., Mccague, S., Ashtari, M., Distefano-Pappas, J., et al. (2016). 29Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. *Lancet (London, England)* 388, 661–672. doi: 10.1016/S0140-6736(16)30371-3
- Bertini, E., Hwu, W.-L., Reyna, S., Farwell, W., Gheuens, S., Sun, P., et al. (2017). Efficacy and safety of nusinersen in infants with presymptomatic spinal muscular atrophy (SMA): Interim results from the NURTURE study. *Eur. J. Paediatr. Neurol.* 21, e14. doi: 10.1016/j.ejpn.2017.04.1218
- Bianchi, G., Barone, D., Lanzarotti, E., Tettamanti, R., Porta, R., Moltrasio, D., et al. (1993). Defibrotide, a single-stranded polydeoxyribonucleotide acting as an adenosine receptor agonist. *Eur. J. Pharmacol.* 238, 327–334. doi: 10.1016/0014-2999(93)90864-E
- Bommareddy, P. K., Patel, A., Hossain, S., and Kaufman, H. L. (2017). Talimogene laherparepvec (T-VEC) and other oncolytic viruses for the treatment of melanoma. *Am. J. Clin. Dermatol.* 18, 1–15. doi: 10.1007/s40257-016-0238-9
- Bondar, O., Shevchenko, V., Martynova, A., Salakhieva, D., Savina, I., Shtyrlin, Y., et al. (2015). Intracellular delivery of VEGF165 encoding gene therapeutic using trifunctional copolymers of ethylene oxide and propylene oxide. *Eur. Polymer J.* 68, 680–686. doi: 10.1016/j.eurpolymj.2015.03.042
- Bonini, C., Peccatori, J., Stanghellini, M. T., Vago, L., Bondanza, A., Cieri, N., et al. (2015). Haploidentical HSCT: a 15-year experience at San Raffaele. *Bone Marrow Transplant.* 50 Suppl 2, S67–S71. doi: 10.1038/bmt.2015.99
- Booth, C., and Gaspar, H. B. (2009). Pegademase bovine (PEG-ADA) for the treatment of infants and children with severe combined immunodeficiency (SCID). *Biol. Targets Ther.* 3, 349–358. doi: 10.2147/BTT.S3103
- Bressler, S. B. (2009). Introduction: understanding the role of angiogenesis and antiangiogenic agents in age-related macular degeneration. *Ophthalmology* 116, S1–S7. doi: 10.1016/j.ophtha.2009.06.045
- Brichta, L., Hofmann, Y., Hahnen, E., Siebzehnrubl, F., Raschke, H., Blumcke, I., et al. (2003). Valproic acid increases the SMN2 protein level: a well-known drug as a potential therapy for spinal muscular atrophy. *Hum. Mol. Genet.* 12, 2481–2489. doi: 10.1093/hmg/ddg256
- Brown, D. M., Michels, M., Kaiser, P. K., Heier, J. S., Sy, J. P., and Ianchulev, T. (2009). Ranibizumab versus verteporfin photodynamic therapy for neovascular age-related macular degeneration: two-year results of the ANCHOR study. *Ophthalmology* 116, 57–65. doi: 10.1016/j.ophtha.2008.10.018
- Bryant, L. M., Christopher, D. M., Giles, A. R., Hinderer, C., Rodriguez, J. L., Smith, J. B., et al. (2013). 23Lessons learned from the clinical development and market authorization of Glybera. *Hum. Gene Ther. Clin. Dev.* 24, 55–64. doi: 10.1089/humc.2013.087
- Burki, U., Keane, J., Blain, A., O’donovan, L., Gait, M. J., Laval, S. H., et al. (2015). Development and application of an ultrasensitive hybridization-based ELISA method for the determination of peptide-conjugated phosphorodiamidate morpholino oligonucleotides. *Nucleic Acid Ther.* 25, 275–284. doi: 10.1089/nat.2014.0528
- Butera, S. (2018). CAR-T: trailblazing the path from clinical development to the clinic. *Gene Ther.* 25, 163–164. doi: 10.1038/s41434-018-0013-z
- Butler, J. S., Chan, A., Costelha, S., Fishman, S., Willoughby, J. L. S., Borland, T. D., et al. (2016). 8Preclinical evaluation of RNAi as a treatment for transthyretin-mediated amyloidosis. *Amyloid* 23, 109–118. doi: 10.3109/13506129.2016.1160882
- Castro, D., and Iannaccone, S. T. (2014). Spinal muscular atrophy: therapeutic strategies. *Curr. Treat. Options Neurol.* 16, 316. doi: 10.1007/s11940-014-0316-3
- Chakradhar, S. (2017). Viral vanguard: designing cancer-killing viruses to chase metastatic tumors. *Nat. Med.* 23, 652–655. doi: 10.1038/nm0617-652
- Chalandon, Y., Roosnek, E., Mermillod, B., Newton, A., Ozsahin, H., Wacker, P., et al. (2004). Prevention of veno-occlusive disease with defibrotide after

- allogeneic stem cell transplantation. *Biol. Blood Marrow Transplant.* 10, 347–354. doi: 10.1016/j.bbmt.2004.01.002
- Chang, J.-G., Hsieh-Li, H.-M., Jong, Y.-J., Wang, N. M., Tsai, C.-H., and Li, H. (2001). Treatment of spinal muscular atrophy by sodium butyrate. *Proc. Natl Acad. Sci.* 98, 9808–9813. doi: 10.1073/pnas.171105098
- Chawla, S. P., Chua, V. S., Fernandez, L., Quon, D., Blackwelder, W. C., Gordon, E. M., et al. (2010). Advanced phase I/II studies of targeted gene delivery *in vivo*: intravenous REXIN-G for gemcitabine-resistant metastatic pancreatic cancer. *Mol. Ther.* 18, 435–441. doi: 10.1038/mt.2009.228
- Chen, C., Pan, J., and Xu, L. (2003). Recombinant adenovirus p53 agent injection combined with radiotherapy in treatment of nasopharyngeal carcinoma: a phase II clinical trial. *Zhonghua Yi Xue Za Zhi* 83, 2033–2035.
- Chen, G.-X., Zhang, S., He, X.-H., Liu, S.-Y., Ma, C., and Zou, X.-P. (2014). Clinical utility of recombinant adenoviral human p53 gene therapy: current perspectives. *OncoTargets Ther.* 7, 1901. doi: 10.2147/OTT.S50483
- Cherian, J., Parvizi, J., Bramlet, D., Lee, K., Romness, D., and Mont, M. (2015). Preliminary results of a phase II randomized study to determine the efficacy and safety of genetically engineered allogeneic human chondrocytes expressing TGF- β 1 in patients with grade 3 chronic degenerative joint disease of the knee. *Osteoarthr. Cartil.* 23, 2109–2118. doi: 10.1016/j.joca.2015.06.019
- Chiriboga, C. A. (2017). Nusinersen for the treatment of spinal muscular atrophy. *Expert Rev. Neurother.* 17, 955–962. doi: 10.1080/14737175.2017.1364159
- Chmielewski, M., and Abken, H. (2015). TRUCKs: the fourth generation of CARs. *Expert Opin. Biol. Ther.* 15, 1145–1154. doi: 10.1517/14712598.2015.1046430
- Chmielewski, M., Hombach, A. A., and Abken, H. (2014). Of CARs and TRUCKs: chimeric antigen receptor (CAR) T cells engineered with an inducible cytokine to modulate the tumor stroma. *Immunol. Rev.* 257, 83–90. doi: 10.1111/immr.12125
- Cho, J., Kim, T., Park, Y., and Lee, B. (2015). TissueGene-C (INVOSSA™) in patients with osteoarthritis: a phase II trials. *Osteoarthr. Cartil.* 23, A170. doi: 10.1016/j.joca.2015.02.936
- Cho, J., Kim, T., Shin, J., Kang, S., and Lee, B. (2017). A phase III clinical results of INVOSSA™ (TissueGene C): a clues for the potential disease modifying OA drug. *Cytotherapy* 19, S148. doi: 10.1016/j.jcyt.2017.02.221
- Choi, K., Lee, H., Kim, D., Kim, M., Lim, C.-L., Lee, Y.-J., et al. (2017). Invossa™ (TISSUEGENE-C) induces an anti-inflammatory environment in the arthritic knee joints via macrophage polarization. *Osteoarthr. Cartil.* 25, S157. doi: 10.1016/j.joca.2017.02.267
- Chopra, R., Eaton, J., Grassi, A., Potter, M., Shaw, B., Salat, C., et al. (2000). Defibrotide for the treatment of hepatic veno-occlusive disease: results of the European compassionate-use study. *Br. J. Haematol.* 111, 1122–1129. doi: 10.1046/j.1365-2141.2000.02475.x
- Chung, D. C., Bertelsen, M., Lorenz, B., Pennesi, M. E., Leroy, B. P., Hamel, C. P., et al. (2018). The natural history of inherited retinal dystrophy due to biallelic mutations in the RPE65 Gene. *Am. J. Ophthalmol.* 199, 58–70. doi: 10.1016/j.ajo.2018.09.024
- Cicalese, M. P., Ferrua, F., Castagnaro, L., Pajno, R., Barzaghi, F., Giannelli, S., et al. (2016). Update on the safety and efficacy of retroviral gene therapy for immunodeficiency due to adenosine deaminase deficiency. *Blood* 128, 45–54. doi: 10.1182/blood-2016-01-688226
- Cicalese, M. P., Ferrua, F., Castagnaro, L., Rolfé, K., De Boever, E., Reinhardt, R. R., et al. (2018). Gene therapy for adenosine deaminase deficiency: a comprehensive evaluation of short- and medium-term safety. *Mol. Ther.* 26, 917–931. doi: 10.1016/j.ymthe.2017.12.022
- Ciceri, F., Bonini, C., Stanghellini, M. T., Bondanza, A., Traversari, C., Salomoni, M., et al. (2009). Infusion of suicide-gene-engineered donor lymphocytes after family haploidentical haemopoietic stem-cell transplantation for leukaemia (the TK007 trial): a non-randomised phase I-II study. *Lancet Oncol.* 10, 489–500. doi: 10.1016/S1470-2045(09)70074-9
- Cirak, S., Arechavala-Gomez, V., Guglieri, M., Feng, L., Torelli, S., Anthony, K., et al. (2011). Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study. *Lancet* 378, 595–605. doi: 10.1016/S0140-6736(11)60756-3
- Colella, P., Ronzitti, G., and Mingozzi, F. (2018). Emerging issues in AAV-mediated *in vivo* gene therapy. *Mol. Ther. Methods Clin. Dev.* 8, 87–104. doi: 10.1016/j.omtm.2017.11.007
- Communes Internationales Des Substances, D. (2016). International Nonproprietary Names for Pharmaceutical Substances (INN). *WHO Drug Information* 30.
- Corbacioglu, S., Cesaro, S., Faraci, M., Valteau-Couanet, D., Gruhn, B., Rovelli, A., et al. (2012). Defibrotide for prophylaxis of hepatic veno-occlusive disease in paediatric haemopoietic stem-cell transplantation: an open-label, phase 3, randomised controlled trial. *Lancet* 379, 1301–1309. doi: 10.1016/S0140-6736(11)61938-7
- Corbacioglu, S., Greil, J., Peters, C., Wulffraat, N., Laws, H., Dilloo, D., et al. (2004). Defibrotide in the treatment of children with veno-occlusive disease (VOD): a retrospective multicentre study demonstrates therapeutic efficacy upon early intervention. *Bone Marrow Transplant.* 33, 189. doi: 10.1038/sj.bmt.1704329
- Corey, D. R. (2017). Nusinersen, an antisense oligonucleotide drug for spinal muscular atrophy. *Nat. Neurosci.* 20, 497. doi: 10.1038/nn.4508
- Corrigan-Curay, J., O'Reilly, M., Kohn, D. B., Cannon, P. M., Bao, G., Bushman, F. D., et al. (2015). Genome editing technologies: defining a path to clinic: genomic editing: establishing preclinical toxicology standards, Bethesda, Maryland 10 June 2014. *Mol. Ther.* 23, 796–806. doi: 10.1038/mt.2015.54
- Crooke, R. M., Graham, M. J., Lemonidis, K. M., Whipple, C. P., Koo, S., and Perera, R. J. (2005). An apolipoprotein B antisense oligonucleotide lowers LDL cholesterol in hyperlipidemic mice without causing hepatic steatosis. *J. Lipid Res.* 46, 872–884. doi: 10.1194/jlr.M400492-JLR200
- Crooke, S. T., and Geary, R. S. (2013). Clinical pharmacological properties of mipomersen (Kynamro), a second generation antisense inhibitor of apolipoprotein B. *Br. J. Clin. Pharmacol.* 76, 269–276. doi: 10.1111/j.1365-2125.2012.04469.x
- D'Amico, D. J. (2005). VEGF inhibition study in ocular neovascularization (VISION): Second year efficacy data. *Invest. Ophthalmol. Visual Sci.* 46, 2309–2309.
- Dydevall, C., and Sumner, C. J. (2015). Spinal muscular atrophy therapeutics: where do we stand? *Neurotherapeutics* 12, 303–316. doi: 10.1007/s13311-015-0337-y
- Dabbous, O., Sproule, D. M., Feltner, D. E., Ogrinc, F. G., Menier, M., Droege, M., et al. (2019). Event-free survival and motor milestone achievement following AVXS-101 and nusinersen interventions contrasted to natural history for Type I spinal muscular atrophy patients. *Neurology*, 92 (15 Supplement), S25.005.
- de Smet, M. D., Meenken, C., and Van Den Horn, G. J. (1999). Fomivirsen—a phosphorothioate oligonucleotide for the treatment of CMV retinitis. *Ocul. Immunol. Inflammation* 7, 189–198. doi: 10.1076/ocii.7.3.189.4007
- Deayton, J. R., Wilson, P., Sabin, C. A., Davey, C. C., Johnson, M. A., Emery, V. C., et al. (2000). Changes in the natural history of cytomegalovirus retinitis following the introduction of highly active antiretroviral therapy. *Aids* 14, 1163–1170. doi: 10.1097/00002030-200006160-00013
- Decker, J., and Goldstein, J. C. (1982). Risk factors in head and neck cancer. *N. Engl. J. Med.* 306, 1151–1155. doi: 10.1056/NEJM198205133061905
- Deev, R., Plaksa, I., Bozo, I., and Isaev, A. (2017). Results of an international postmarketing surveillance study of pl-VEGF165 safety and efficacy in 210 patients with peripheral arterial disease. *Am. J. Cardiovasc. Drugs* 17, 235–242. doi: 10.1007/s40256-016-0210-3
- Deev, R. V., Bozo, I. Y., Mzhavanadze, N. D., Voronov, D. A., Gavrilenko, A. V., Chervyakov, Y. V., et al. (2015). pCMV-vegf165 intramuscular gene transfer is an effective method of treatment for patients with chronic lower limb ischemia. *J. Cardiovasc. Pharmacol. Ther.* 20, 473–482. doi: 10.1177/1074248415574336
- Denny, W. A. (2003). Prodrugs for gene-directed enzyme-prodrug therapy (Suicide Gene Therapy). *J. Biomed. Biotechnol.* 2003, 48–70. doi: 10.1155/S110724303209098
- Dias, M. F., Joo, K., Kemp, J. A., Fialho, S. L., Da Silva Cunha, A., Woo, S. J., et al. (2018). 22Molecular genetics and emerging therapies for retinitis pigmentosa: Basic research and clinical perspectives. *Prog. Retin. Eye Res.* 63, 107–131. doi: 10.1016/j.preteyeres.2017.10.004
- Dicarlo, J. E., Deeconda, A., and Tsang, S. H. (2017). Viral vectors, engineered cells and the CRISPR revolution. *Adv. Exp. Med. Biol.* 1016, 3–27. doi: 10.1007/978-3-319-63904-8_1
- Drolet, D. W., Nelson, J., Tucker, C. E., Zack, P. M., Nixon, K., Bolin, R., et al. (2000). Pharmacokinetics and safety of an anti-vascular endothelial growth factor aptamer (NX1838) following injection into the vitreous humor of rhesus monkeys. *Pharm. Res.* 17, 1503–1510. doi: 10.1023/A:1007657109012

- Dunbar, C. E., High, K. A., Joung, J. K., Kohn, D. B., Ozawa, K., and Sadelain, M. (2018). Gene therapy comes of age. *Science* 359, eaan4672. doi: 10.1126/science.aan4672
- Ehrke-Schulz, E., Schiwon, M., Leitner, T., Dávid, S., Bergmann, T., Liu, J., et al. (2017). CRISPR/Cas9 delivery with one single adenoviral vector devoid of all viral genes. *Sci. Rep.* 7, 17113. doi: 10.1038/s41598-017-17180-w
- Ervasti, J. M. (2007). Dystrophin, its interactions with other proteins, and implications for muscular dystrophy. *BBA-Mol. Basis Dis* 1772, 108–117. doi: 10.1016/j.bbadis.2006.05.010
- Escolano-Lozano, F., Barreiros, A. P., Birklein, F., and Geber, C. (2017). 4Transthyretin familial amyloid polyneuropathy (TTR-FAP): Parameters for early diagnosis. *Brain Behav* 8, e00889–e00889. doi: 10.1002/brb3.889
- Faustinella, F., Chang, A., Van Biervliet, J. P., Rosseneu, M., Vinaimont, N., Smith, L. C., et al. (1991). Catalytic triad residue mutation (Asp156—Gly) causing familial lipoprotein lipase deficiency. Co-inheritance with a nonsense mutation (Ser447—Ter) in a Turkish family. *J. Biol. Chem.* 266, 14418–14424.
- Ferreira, V., Petry, H., and Salmon, F. (2014). 30Immune Responses to AAV-Vectors, the Glybera Example from Bench to Bedside. *Front. Immunol.* 5, 82–82. doi: 10.3389/fimmu.2014.00082
- Ferrua, E., and Aiuti, A. (2017). Twenty-five years of gene therapy for ADA-SCID: from bubble babies to an approved drug. *Human Gene Ther.* 28, 972–981. doi: 10.1089/hum.2017.175
- Fine, S. L., Berger, J. W., Maguire, M. G., and Ho, A. C. (2000). Age-related macular degeneration. *N. Engl. J. Med.* 342, 483–492. doi: 10.1056/NEJM200002173420707
- Finkel, R. S., Chiriboga, C. A., Vajsaar, J., Day, J. W., Montes, J., De Vivo, D. C., et al. (2016). Treatment of infantile-onset spinal muscular atrophy with nusinersen: a phase 2, open-label, dose-escalation study. *Lancet* 388, 3017–3026. doi: 10.1016/S0140-6736(16)31408-8
- Finkel, R. S., Kuntz, N., Mercuri, E., Muntoni, F., Chiriboga, C., and Darras, B. Primary efficacy and safety results from the phase 3 ENDEAR study of nusinersen in infants diagnosed with spinal muscular atrophy (SMA). 43rd annual congress of the British Paediatric Neurology Association, 2017. doi: 10.1016/j.ejpn.2017.04.1219
- Francischetti, I. M., Oliveira, C. J., Ostera, G. R., Yager, S. B., Debierre-Grockiego, F., Carregaro, V., et al. (2012). Defibrotide interferes with several steps of the coagulation-inflammation cycle and exhibits therapeutic potential to treat severe malaria. *Arterioscler. Thromb. Vasc. Biol.* 32, 786–798. doi: 10.1161/ATVBAHA.111.240291
- Freedman, W. R. (2001). Retinal toxic effects associated with intravitreal fomivirs. *Arch. Ophthalmol.* 119, 458–458.
- Friedman, D. S., Ocolmain, B. J., Munoz, B., Tomany, S. C., Mccarty, C., De Jong, P., et al. (2004). Prevalence of age-related macular degeneration in the United States. *Arch. Ophthalmol.* 122, 564–572. doi: 10.1001/archophth.122.4.564
- Friedmann, T. (2007). A decade of accomplishments: gene therapy and the ASGT. *Mol. Ther.* 15, 1576–1578. doi: 10.1038/sj.mt.6300284
- Fukuhara, H., Ino, Y., and Todo, T. (2016). Oncolytic virus therapy: a new era of cancer treatment at dawn. *Cancer Sci.* 107, 1373–1379. doi: 10.1111/cas.13027
- Gallant, J. E., Moore, R. D., Richman, D. D., Keruly, J., Chaisson, R. E., and Group, Z. E. S. (1992). Incidence and natural history of cytomegalovirus disease in patients with advanced human immunodeficiency virus disease treated with zidovudine. *J. Infect. Dis.* 166, 1223–1227. doi: 10.1093/infdis/166.6.1223
- Garber, K. (2016). Big win possible for Ionis/Biogen antisense drug in muscular atrophy. *Nat. Biotechnol.* 34, 1002–1003. doi: 10.1038/nbt1016-1002
- Gaudet, D., Méthot, J., Déry, S., Brisson, D., Essiembre, C., Tremblay, G., et al. (2013). 31Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial. *Gene Ther.* 20, 361–369. doi: 10.1038/gt.2012.43
- Gaudet, D., Méthot, J., and Kastelein, J. (2012). 28Gene therapy for lipoprotein lipase deficiency. *Curr Opin. Lipidol.* 23, 310–320. doi: 10.1097/MOL.0b013e3283555a7e
- Gaudet, D., Stroes, E. S., Méthot, J., Brisson, D., Tremblay, K., Bernelot Moens, S. J., et al. (2016). 35Long-term retrospective analysis of gene therapy with alipogene tiparvovec and its effect on lipoprotein lipase deficiency-induced pancreatitis. *Human Gene Ther.* 27, 916–925. doi: 10.1089/hum.2015.158
- Geary, R. S., Norris, D., Yu, R., and Bennett, C. F. (2015). Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. *Adv. Drug Delivery Rev.* 87, 46–51. doi: 10.1016/j.addr.2015.01.008
- Goldberg, A. C., Hopkins, P. N., Toth, P. P., Ballantyne, C. M., Rader, D. J., Robinson, J. G., et al. (2011). Familial hypercholesterolemia: screening, diagnosis and management of pediatric and adult patients: clinical guidance from the National Lipid Association Expert Panel on Familial Hypercholesterolemia. *J. Clin. Lipidol.* 5, S1–S8. doi: 10.1016/j.jacl.2011.04.003
- Goldberg, D. E., Smithen, L. M., Angelilli, A., and Freeman, W. R. (2005). HIV-associated retinopathy in the HAART era. *Retina* 25, 633–649. doi: 10.1097/00006982-200507000-00015
- Gordon, E. M., Cornelio, G. H., Lorenzo, C. C., Levy, J. P., Reed, R. A., Liu, L., et al. (2004). First clinical experience using a ‘pathotropic’injectable retroviral vector (Rexin-G) as intervention for stage IV pancreatic cancer. *Int. J. Oncol.* 24, 177–185. doi: 10.3892/ijo.24.1.177
- Gordon, E. M., and Hall, F. L. (2010). Rexin-G, a targeted genetic medicine for cancer. *Expert Opin. Biol. Ther.* 10, 819–832. doi: 10.1517/14712598.2010.481666
- Gordon, E. M., Lopez, F. F., Cornelio, G. H., Lorenzo, C. C., Levy, J. P., Reed, R. A., et al. (2006). Pathotropic nanoparticles for cancer gene therapy Rexin-G™ IV: Three-year clinical experience. *Int. J. Oncol.* 29, 1053–1064. doi: 10.3892/ijo.29.5.1053
- Gragoudas, E. S., Adamis, A. P., Cunningham, E. T., Jr., Feinsod, M., and Guyer, D. R. (2004). Pegaptanib for neovascular age-related macular degeneration. *N. Engl. J. Med.* 351, 2805–2816. doi: 10.1056/NEJMoa042760
- Group, V. S. (2002a). A randomized controlled clinical trial of intravitreal fomivirs for treatment of newly diagnosed peripheral cytomegalovirus retinitis in patients with aids1. *Am. J. Ophthalmol.* 133, 467–474. doi: 10.1016/S0002-9394(02)01327-2
- Group, V. S. (2002b). Randomized dose-comparison studies of intravitreal fomivirs for treatment of cytomegalovirus retinitis that has reactivated or is persistently active despite other therapies in patients with AIDS1. *Am. J. Ophthalmol.* 133, 475–483. doi: 10.1016/S0002-9394(02)01326-0
- Group, V. S. (2002c). Safety of intravitreal fomivirs for treatment of cytomegalovirus retinitis in patients with AIDS1. *Am. J. Ophthalmol.* 133, 484–498. doi: 10.1016/S0002-9394(02)01332-6
- Hampson, G., Towse, A., Pearson, S. D., Dreitlein, W. B., and Henshall, C. (2017). 24Gene therapy: evidence, value and affordability in the US health care system. *J. Comp. Eff. Res.* 7, 15–28. doi: 10.2217/ceer-2017-0068
- Han, D., Huang, Z., Zhang, W., Yu, Z., Wang, Q., Ni, X., et al. (2003). Effectiveness of recombinant adenovirus p53 injection on laryngeal cancer: phase I clinical trial and follow up. *Zhonghua Yi Xue Za Zhi* 83, 2029–2032.
- Hartmann, J., Schüßler-Lenz, M., Bondanza, A., and Buchholz, C. J. (2017). Clinical development of CAR T cells—challenges and opportunities in translating innovative treatment concepts. *EMBO Mol. Med.* 9, 1183–1197. doi: 10.15252/emmm.201607485
- Hassan, A., Booth, C., Brightwell, A., Allwood, Z., Veys, P., Rao, K., et al. (2012). Outcome of hematopoietic stem cell transplantation for adenosine deaminase-deficient severe combined immunodeficiency. *Blood* 120, 3615–24; quiz 3626. doi: 10.1182/blood-2011-12-396879
- Hershfield, M. S. (2009). “Adenosine deaminase deficiency,” in *Encyclopedia of Molecular Mechanisms of Disease*. Ed. F. Lang (Berlin, Heidelberg: Springer)
- Hovingh, G. K., Davidson, M. H., Kastelein, J. J., and O’connor, A. M. (2013). Diagnosis and treatment of familial hypercholesterolaemia. *Eur. Heart J.* 34, 962–971. doi: 10.1093/eurheartj/ehd015
- Hoy, S. M. (2017). Nusinersen: first global approval. *Drugs* 77, 473–479. doi: 10.1007/s40265-017-0711-7
- Hu, J. C., Coffin, R. S., Davis, C. J., Graham, N. J., Groves, N., Guest, P. J., et al. (2006). A phase I study of OncoVEXGM-CSF, a second-generation oncolytic herpes simplex virus expressing granulocyte macrophage colony-stimulating factor. *Clin. Cancer Res* 12, 6737–6747. doi: 10.1158/1078-0432.CCR-06-0759
- Johnson, D. B., Puzanov, I., and Kelley, M. C. (2015). Talimogene laherparepvec (T-VEC) for the treatment of advanced melanoma. *Immunotherapy* 7, 611–619. doi: 10.2217/imt.15.35
- Kato, Z., Okuda, M., Okumura, Y., Arai, T., Teramoto, T., Nishimura, M., et al. (2009). Oral administration of the thyrotropin-releasing hormone (TRH) analogue, taltireline hydrate, in spinal muscular atrophy. *J. Child Neurol.* 24, 1010–1012. doi: 10.1177/0883073809333535

- Kaufman, H. L., Kim, D. W., Deraffele, G., Mitcham, J., Coffin, R. S., and Kim-Schulze, S. (2010). Local and distant immunity induced by intralesional vaccination with an oncolytic herpes virus encoding GM-CSF in patients with stage IIIC and IV melanoma. *Ann Surg Oncol* 17, 718–730. doi: 10.1245/s10434-009-0809-6
- Kell, J. (2016). Considerations and challenges for patients with refractory and relapsed acute myeloid leukaemia. *Leuk Res* 47, 149–160. doi: 10.1016/j.leukres.2016.05.025
- Kempen, J. H., Martin, B. K., Wu, A. W., Barron, B., Thorne, J. E., Jabs, D. A., et al. (2003). The effect of cytomegalovirus retinitis on the quality of life of patients with AIDS in the era of highly active antiretroviral therapy. *Ophthalmology* 110, 987–995. doi: 10.1016/S0161-6420(03)00089-7
- Kerschen, P., and Planté-Bordeneuve, V. (2016). 10Current and Future Treatment Approaches in Transthyretin Familial Amyloid Polyneuropathy. *Curr. Treat. Options Neurol.* 18, 53. doi: 10.1007/s11940-016-0436-z
- Kinali, M., Arechavala-Gomez, V., Feng, L., Cirak, S., Hunt, D., Adkin, C., et al. (2009). Local restoration of dystrophin expression with the morpholino oligomer AVI-4658 in Duchenne muscular dystrophy: a single-blind, placebo-controlled, dose-escalation, proof-of-concept study. *Lancet Neurol* 8, 918–928. doi: 10.1016/S1474-4422(09)70211-X
- Kohlhapp, F. J., and Kaufman, H. L. (2016). Molecular pathways: mechanism of action for talimogene laherparepvec, a new oncolytic virus immunotherapy. *Clin. Cancer Res.* 22, 1048–1054. doi: 10.1158/1078-0432.CCR-15-2667
- Kole, R., and Krieg, A. M. (2015). Exon skipping therapy for Duchenne muscular dystrophy. *Adv. Drug Delivery Rev.* 87, 104–107. doi: 10.1016/j.addr.2015.05.008
- Kumar, S. R., Markusic, D. M., Biswas, M., High, K. A., and Herzog, R. W. (2016). Clinical development of gene therapy: results and lessons from recent successes. *Mol. Ther. Methods Clin. Dev.* 3, 16034. doi: 10.1038/mtm.2016.34
- Lee, B. (2018). INVOSSA, a first-in-class of cell and gene therapy for osteoarthritis treatment: the phase III trial. *Osteoarthr. Cartil.* 26, S43–S44. doi: 10.1016/j.joca.2018.02.103
- Lee, C. S., Bishop, E. S., Zhang, R., Yu, X., Farina, E. M., Yan, S., et al. (2017). Adenovirus-mediated gene delivery: potential applications for gene and cell-based therapies in the new era of personalized medicine. *Genes Dis.* 4, 43–63. doi: 10.1016/j.gendis.2017.04.001
- Lee, H., Choi, K., Kim, H., Kim, D., Lee, Y., Lee, B., et al. (2018). INVOSSA-K induces an anti-inflammatory environment in a rat mia model via macrophage polarization. *Osteoarthr. Cartil.* 26, S125. doi: 10.1016/j.joca.2018.02.272
- Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Violette, L., et al. (1995). Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80, 155–165. doi: 10.1016/0092-8674(95)90460-3
- Levin, A. A., Rosie, Z. Y., Geary, R. S. (2007). “Basic principles of the pharmacokinetics of antisense oligonucleotide drugs.” In *Antisense drug technology*. 2nd ed. Ed. S. T. Crooke (CRC Press).
- Li, Y., Li, B., Li, C.-J., and Li, L.-J. (2015). Key points of basic theories and clinical practice in RAD-p53 (Gendicine™) gene therapy for solid malignant tumors. *Expert Opin. Biol. Ther.* 15, 437–454. doi: 10.1517/14712598.2015.990882
- Liang, M. (2018). Oncorine, the world first oncolytic virus medicine and its update in China. *Curr. Cancer Drug Targets* 18, 171–176. doi: 10.2174/1568009618666171129221503
- Lim, K. R. Q., Maruyama, R., and Yokota, T. (2017). Eteplirsens in the treatment of Duchenne muscular dystrophy. *Drug Design Dev. Ther.* 11, 533. doi: 10.2147/DDDT.S97635
- Locke, F. L., Neelapu, S. S., Bartlett, N. L., Lekakis, L. J., Miklos, D. B., Jacobson, C. A., et al. (2017). Clinical and biologic covariates of outcomes in ZUMA-1: a pivotal trial of axicabtagene ciloleucel (axi-cel; KTE-C19) in patients with refractory aggressive non-Hodgkin lymphoma (r-NHL). *J. Clin. Oncol.* 35, 7512–7512. doi: 10.1200/JCO.2017.35.15_suppl.7512
- Lorentino, F., Labopin, M., Fleischhauer, K., Cicieri, F., Mueller, C. R., Ruggeri, A., et al. (2017). The impact of HLA matching on outcomes of unmanipulated haploidentical HSCT is modulated by GVHD prophylaxis. *Blood Adv.* 1, 669–680. doi: 10.1182/bloodadvances.2017006429
- Lowes, L., Al-Zaidy, S., Shell, R., Arnold, W. D., Rodino-Klapac, L., Prior, T. W., et al. (2017). AVXS-101 Phase 1 gene therapy clinical trial in SMA Type 1: correlation between CHOP-INTEND and motor milestone achievements (S13.002). *Neurology* 88, S13.002.
- Lu, W., Zheng, S., Li, X.-F., Huang, J.-J., Zheng, X., and Li, Z. (2004). Intratumor injection of H101, a recombinant adenovirus, in combination with chemotherapy in patients with advanced cancers: a pilot phase II clinical trial. *World J. Gastroenterol.* 10, 3634. doi: 10.3748/wjg.v10.i24.3634
- Maharshi, V., and Hasan, S. (2017). Nusinersen: the first option beyond supportive care for spinal muscular atrophy. *Clin Drug Investig.* 37, 807–817. doi: 10.1007/s40261-017-0557-5
- Mahley, R. (2001). “Biochemistry and physiology of lipid and lipoprotein metabolism,” in *Principles and Practice of Endocrinology and Metabolism*, 3rd ed. Ed. K. L. Becker (Lippincott Williams and Wilkins).
- Mäkinen, K., Manninen, H., Hedman, M., Matsi, P., Mussalo, H., Alhava, E., et al. (2002). Increased vascularity detected by digital subtraction angiography after VEGF gene transfer to human lower limb artery: a randomized, placebo-controlled, double-blinded phase II study. *Mol. Ther.* 6, 127–133. doi: 10.1006/mthe.2002.0638
- Malone, D. C., Dean, R., Arjunji, R., Jensen, I., Cyr, P., Miller, B., et al. (2019). Cost-effectiveness analysis of using onasemnogene abeparvovec (AVXS-101) in spinal muscular atrophy type 1 patients. *J. Mark. Access Health Policy* 7, 1601484. doi: 10.1080/20016689.2019.1601484
- Mattis, V. B., Fosso, M. Y., Chang, C.-W., and Lorson, C. L. (2009). Subcutaneous administration of TC007 reduces disease severity in an animal model of SMA. *BMC Neurosci.* 10, 142. doi: 10.1186/1471-2202-10-142
- Maude, S. L., Laetsch, T. W., Buechner, J., Rives, S., Boyer, M., Bittencourt, H., et al. (2018). Tisagenlecleucel in Children and Young Adults with B-Cell Lymphoblastic Leukemia. *N. Engl. J. Med.* 378, 439–448. doi: 10.1056/NEJMoa1709866
- Mcgowan, M. P., Tardif, J.-C., Ceska, R., Burgess, L. J., Soran, H., Gouni-Berthold, I., et al. (2012). Randomized, placebo-controlled trial of mipomersen in patients with severe hypercholesterolemia receiving maximally tolerated lipid-lowering therapy. *PLoS one* 7, e49006. doi: 10.1371/journal.pone.0049006
- Mendell, J. R., Goemans, N., Lowes, L. P., Alfano, L. N., Berry, K., Shao, J., et al. (2016). Longitudinal effect of eteplirsens versus historical control on ambulation in Duchenne muscular dystrophy. *Ann. Neurol.* 79, 257–271. doi: 10.1002/ana.24555
- Mendell, J. R., Al-Zaidy, S., Shell, R., Arnold, W. D., Rodino-Klapac, L., Prior, T. W., et al. (2017). AVXS-101 Phase 1 gene therapy clinical trial in SMA Type 1: Event free survival and achievement of developmental milestones. *Eur. J. Paediatr. Neurol.* 21, e13–e14. doi: 10.1016/j.ejpn.2017.04.1216
- Mendell, J. R., and Lloyd-Puryear, M. (2013). Report of MDA muscle disease symposium on newborn screening for Duchenne muscular dystrophy. *Muscle Nerve* 48, 21–26. doi: 10.1002/mus.23810
- Mendell, J. R., Rodino-Klapac, L. R., Sahenk, Z., Roush, K., Bird, L., Lowes, L. P., et al. (2013). Eteplirsens for the treatment of Duchenne muscular dystrophy. *Ann. Neurol.* 74, 637–647. doi: 10.1002/ana.23982
- Mercuri, E., and Bertini, E. (2012). Stem cells in severe infantile spinal muscular atrophy. *Neuromuscul. Disord.* 22, 1105. doi: 10.1016/j.nmd.2012.11.001
- Mercuri, E., Finkel, R., Kirschner, J., Chiriboga, C., Kuntz, N., Sun, P., et al. (2017). Efficacy and safety of nusinersen in children with later-onset spinal muscular atrophy (SMA): end of study results from the phase 3 CHERISH study. *Neuromuscul. Disord.* 27, S210. doi: 10.1016/j.nmd.2017.06.418
- Mercuri, E., Finkel, R., Montes, J., Mazzone, E. S., Sormani, M. P., Main, M., et al. (2016). Patterns of disease progression in type 2 and 3 SMA: implications for clinical trials. *Neuromuscul. Disord.* 26, 126–131. doi: 10.1016/j.nmd.2015.10.006
- Miraldi Utz, V., Coussa, R. G., Antaki, F., and Traboulsi, E. I. (2018). Gene therapy for RPE65-related retinal disease. *Ophthalmic Genet.* 39, 671–677. doi: 10.1080/13816810.2018.1533027
- Moat, S. J., Bradley, D. M., Salmon, R., Clarke, A., and Hartley, L. (2013). Newborn bloodspot screening for Duchenne muscular dystrophy: 21 years experience in Wales (UK). *Eur. J. Human Genet.* 21, 1049. doi: 10.1038/ejhg.2012.301
- Mohty, M., Labopin, M., Velardi, A., Van Lint, M. T., Bunjes, D., Bruno, B., et al. (2016). Allogeneic genetically modified T Cells (HSV-TK) as adjunctive treatment in haploidentical hematopoietic stem-cell transplantation (haplo-HSCT) of adult patients with high-risk hematological malignancies: a pair-matched analysis from the acute Leukemia working party of EBMT. *Blood* 128, 672.
- Mohty, M., Malard, F., Abecassis, M., Aerts, E., Alaskar, A., Aljurf, M., et al. (2015). Sinusoidal obstruction syndrome/veno-occlusive disease: current situation

- and perspectives—a position statement from the European Society for Blood and Marrow Transplantation (EBMT). *Bone Marrow Transplant*. 50, 781. doi: 10.1038/bmt.2015.52
- Moulton, J. (2016). Guide for morpholino users: toward therapeutics. *J. Drug Discov. Dev. Deliv.* 3, 1023.
- Mulamba, G. B., Hu, A., Azad, R. F., Anderson, K. P., and Coen, D. M. (1998). Human cytomegalovirus mutant with sequence-dependent resistance to the phosphorothioate oligonucleotide fomivirsen (ISIS 2922). *Antimicrob. Agents Chemother.* 42, 971–973. doi: 10.1128/AAC.42.4.971
- Mullard, A. (2017). 2016 EMA drug approval recommendations. (Nature Publishing Group). doi: 10.1038/nrd.2017.17
- Neelapu, S. S., Locke, F. L., Bartlett, N. L., Lekakis, L., Miklos, D., Jacobson, C. A., et al. (2016). Kte-C19 (anti-CD19 CAR T Cells) induces complete remissions in patients with refractory diffuse large B-Cell Lymphoma (DLBCL): results from the pivotal Phase 2 Zuma-1. *Blood* 128, LBA-6-LBA-6.
- Neelapu, S. S., Locke, F. L., Bartlett, N. L., Lekakis, L. J., Miklos, D. B., Jacobson, C. A., et al. (2017). Axicabtagene ciloleucel CAR T-Cell therapy in refractory large B-Cell lymphoma. *N. Engl. J. Med.* 377, 2531–2544. doi: 10.1056/NEJMoa1707447
- Nemunaitis, J., Ganly, I., Khuri, F., Arseneau, J., Kuhn, J., Mccarty, T., et al. (2000). Selective replication and oncolysis in p53 mutant tumors with ONYX-015, an E1B-55kD gene-deleted adenovirus, in patients with advanced head and neck cancer: a phase II trial. *Cancer Res.* 60, 6359–6366.
- Nikoo, H. R., Ardebili, A., Ravanshad, M., Rezaei, F., Teimoori, A., Khanizadeh, S., et al. (2017). E6-Specific detection and typing of human Papillomaviruses in oral cavity specimens from Iranian patients. *Iran. Biomed. J.* 21, 411.
- Pearson, S., Jia, H., and Kandachi, K. (2004). China approves first gene therapy. *Nat. Biotechnol.* 22, 3–4. doi: 10.1038/nbt0104-3
- Peng, Z. (2005). Current status of gendicine in China: recombinant human Ad-p53 agent for treatment of cancers. *Human Gene Ther.* 16, 1016–1027. doi: 10.1089/hum.2005.16.1016
- Plante-Bordeneuve, V. (2018). 3Transthyretin familial amyloid polyneuropathy: an update. *J. Neurol.* 265, 976–983. doi: 10.1007/s00415-017-8708-4
- Prakash, V. (2017). Spinraza—a rare disease success story. *Gene Ther.* 24, 497. doi: 10.1038/gt.2017.59
- Printz, C. (2016). FDA approves new treatment of metastatic melanoma. *Cancer* 122, 1149–1149. doi: 10.1002/cncr.29998
- Puzanov, I., Milhem, M. M., Andtbacka, R. H. I., Minor, D. R., Hamid, O., Li, A., et al. (2014). Primary analysis of a phase 1b multicenter trial to evaluate safety and efficacy of talimogene laherparepvec (T-VEC) and ipilimumab (ipi) in previously untreated, unresected stage IIIB-IV melanoma. *J. Clin. Oncol.* 32, 9029. doi: 10.1200/jco.2014.32.15_suppl.9029
- Raal, F. J., and Santos, R. D. (2012). Homozygous familial hypercholesterolemia: current perspectives on diagnosis and treatment. *Atherosclerosis* 223, 262–268. doi: 10.1016/j.atherosclerosis.2012.02.019
- Raal, F. J., Santos, R. D., Blom, D. J., Marais, A. D., Charng, M.-J., Cromwell, W. C., et al. (2010). Mipomersen, an apolipoprotein B synthesis inhibitor, for lowering of LDL cholesterol concentrations in patients with homozygous familial hypercholesterolemia: a randomised, double-blind, placebo-controlled trial. *Lancet* 375, 998–1006. doi: 10.1016/S0140-6736(10)60284-X
- Ramlogan-Steel, C. A., Murali, A., Andrzejewski, S., Dhungel, B., Steel, J. C., and Layton, C. J. (2018). Gene therapy and the adeno associated virus in the treatment of genetic and acquired ophthalmic diseases in humans: trials, future directions and safety considerations. *Clin. Exp. Ophthalmol.* 47, 521–536. doi: 10.1111/ceo.13416
- Rao, V. K., Kapp, D., Schroth, M. (2018). Gene therapy for spinal muscular atrophy: an emerging treatment option for a devastating disease. *J. Manag. Care Spec. Pharm.* 24(12-a Suppl), S3–S16. doi: 10.18553/jmcp.2018.24.12-a.s3
- Reach, T. (2015). FDA Approves First Oncolytic Virus Therapy: Imlygic for Melanoma.
- Reid, T. R., Freeman, S., Post, L., McCormick, F., and Sze, D. Y. (2005). Effects of Onyx-015 among metastatic colorectal cancer patients that have failed prior treatment with 5-FU/leucovorin. *Cancer Gene Ther.* 12, 673. doi: 10.1038/sj.cgt.7700819
- Richardson, P. G., Corbacioglu, S., Ho, V. T.-V., Kernan, N. A., Lehmann, L., Maguire, C., et al. (2013). Drug safety evaluation of defibrotide. *Expert Opin. Drug Saf.* 12, 123–136. doi: 10.1517/14740338.2012.749855
- Richardson, P. G., Elias, A. D., Krishnan, A., Wheeler, C., Nath, R., Hoppensteadt, D., et al. (1998). Treatment of severe veno-occlusive disease with defibrotide: compassionate use results in response without significant toxicity in a high-risk population. *Blood* 92, 737–744.
- Richardson, P. G., Riches, M. L., Kernan, N. A., Brochstein, J. A., Mineishi, S., Termuhlen, A. M., et al. (2016). Phase 3 trial of defibrotide for the treatment of severe veno-occlusive disease and multi-organ failure. *Blood* 127, 1656–1665. doi: 10.1182/blood-2015-10-676924
- Richardson, P. G., Soiffer, R. J., Antin, J. H., Uno, H., Jin, Z., Kurtzberg, J., et al. (2010). Defibrotide for the treatment of severe hepatic veno-occlusive disease and multiorgan failure after stem cell transplantation: a multicenter, randomized, dose-finding trial. *Biol. Blood Marrow Transplant.* 16, 1005–1017. doi: 10.1016/j.bbmt.2010.02.009
- Ricotta, D. N., and Frishman, W. (2012). Mipomersen: a safe and effective antisense therapy adjunct to statins in patients with hypercholesterolemia. *Cardiol. Rev.* 20, 90–95. doi: 10.1097/CRD.0b013e31823424be
- Robert Hettle, M. C., Hinde, S., Hodgson, R., Jones-Diette, J., Woolcott, N., and Palmer, S. (2017). 34The assessment and appraisal of regenerative medicines and cell therapy products: an exploration of methods for review, economic evaluation and appraisal. Southampton (UK): NIHR Journals Library.
- Roberts, Z. J., Better, M., Bot, A., Roberts, M. R., and Ribas, A. (2017). Axicabtagene ciloleucel, a first-in-class CAR T cell therapy for aggressive NHL. *Leuk. Lymphoma* 59, 1785–1796. doi: 10.1080/10428194.2017.1387905
- Robinson, C. J., and Stringer, S. E. (2001). The splice variants of vascular endothelial growth factor (VEGF) and their receptors. *J. Cell Sci.* 114, 853–865.
- Rodriguez, A., Brown, C., and Badie, B. (2017). Chimeric antigen receptor T-cell therapy for glioblastoma. *Transl Res* 187, 93–102. doi: 10.1016/j.trsl.2017.07.003
- Rosie, Z. Y., Lemonidis, K. M., Graham, M. J., Matson, J. E., Crooke, R. M., Tribble, D. L., et al. (2009). Cross-species comparison of *in vivo* PK/PD relationships for second-generation antisense oligonucleotides targeting apolipoprotein B-100. *Biochem. Pharmacol.* 77, 910–919. doi: 10.1016/j.bcp.2008.11.005
- Ruckman, J., Green, L. S., Beeson, J., Waugh, S., Gillette, W. L., Henninger, D. D., et al. (1998). 2'-Fluoropyrimidine RNA-based aptamers to the 165-amino acid form of vascular endothelial growth factor (VEGF165) inhibition of receptor binding and VEGF-induced vascular permeability through interactions requiring the exon 7-encoded domain. *J. Biol. Chem.* 273, 20556–20567. doi: 10.1074/jbc.273.32.20556
- Russell, S., Bennett, J., Maguire, A. M., and High, K. A. (2018). 17Voretigene neparvovec-rzyl for the treatment of biallelic RPE65 mutation-associated retinal dystrophy. *Expert Opin. Orphan Drugs* 6, 457–464. doi: 10.1080/21678707.2018.1508340
- Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z.-F., Tillman, A., et al. (2017). 19Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet (London, England)* 390, 849–860. doi: 10.1016/S0140-6736(17)31868-8
- Santos, R. D., Raal, F. J., Catapano, A. L., Witztum, J. L., Steinhagen-Thiessen, E., and Tsimikas, S. (2015). Mipomersen, an antisense oligonucleotide to apolipoprotein B-100, reduces lipoprotein (a) in various populations with hypercholesterolemia: results of 4 phase III trials. *Arteriosclerosis, thrombosis, and vascular biology, ATVBAHA.* 114.304549. doi: 10.1161/ATVBAHA.114.304549
- Sarwar, S., Clearfield, E., Soliman, M. K., Sadiq, M. A., Baldwin, A. J., Hanout, M., et al. (2016). Aflibercept for neovascular age-related macular degeneration. *Cochrane Database Syst. Rev.* 2, CD011346. doi: 10.1002/14651858.CD011346.pub2
- Schuster, S. J., Bishop, M. R., Tam, C. S., Waller, E. K., Borchmann, P., McGuirk, J. P., et al. (2017). Primary analysis of juliet: a global, pivotal, phase 2 Trial of CTL019 in adult patients with relapsed or refractory diffuse large B-cell lymphoma. *Blood* 130, 577–577. doi: 10.1016/j.jclml.2017.07.213
- Senzer, N. N., Kaufman, H. L., Amatruda, T., Nemunaitis, M., Reid, T., Daniels, G., et al. (2009). Phase II clinical trial of a granulocyte-macrophage colony-stimulating factor-encoding, second-generation oncolytic herpesvirus in patients with unresectable metastatic melanoma. *J. Clin. Oncol.* 27, 5763–5771. doi: 10.1200/JCO.2009.24.3675
- Shababi, M., Lorson, C. L., and Rudnik-Schöneborn, S. S. (2014). Spinal muscular atrophy: a motor neuron disorder or a multi-organ disease? *J. Anatomy* 224, 15–28. doi: 10.1111/joa.12083
- Singh, N. N., Shishimorova, M., Cao, L. C., Gangwani, L., and Singh, R. N. (2009). A short antisense oligonucleotide masking a unique intronic motif prevents

- skipping of a critical exon in spinal muscular atrophy. *RNA Biol.* 6, 341–350. doi: 10.4161/rna.6.3.8723
- Smith, A. J., Oertle, J., Warren, D., and Prato, D. (2016). Chimeric antigen receptor (CAR) T cell therapy for malignant cancers: summary and perspective. *J. Cell. Immunother.* 2, 59–68. doi: 10.1016/j.jocit.2016.08.001
- Solomon, S. D., Lindsley, K., Vedula, S. S., Krzystolik, M. G., and Hawkins, B. S. (2014). Anti-vascular endothelial growth factor for neovascular age-related macular degeneration. *Cochrane Database Syst. Rev.* 8, CD005139. doi: 10.1002/14651858.CD005139.pub3
- Stein, C. A. (2016). Eteplirsen approved for Duchenne muscular dystrophy: the FDA faces a difficult choice. *Mol. Ther.* 24, 1884–1885. doi: 10.1038/mt.2016.188
- Stein, E. A., Dufour, R., Gagne, C., Gaudet, D., East, C., Donovan, J. M., et al. (2012). Apolipoprotein B synthesis inhibition with mipomersen in heterozygous familial hypercholesterolemia: results of a randomized, double-blind, placebo controlled trial to assess efficacy and safety as add-on therapy in patients with coronary artery disease. *Circulation* 126, 2283–2292. doi: 10.1161/CIRCULATIONAHA.112.104125
- Strouse, C., Richardson, P., Prentice, G., Korman, S., Hume, R., Nejadnik, B., et al. (2016). Defibrotide for treatment of severe veno-occlusive disease in pediatrics and adults: an exploratory analysis using data from the Center for International Blood and Marrow Transplant Research. *Biol. Blood Marrow Transplant.* 22, 1306–1312. doi: 10.1016/j.bbmt.2016.04.011
- Studies Of Ocular Complications Of Aids Research Group, I. C. W. T. A. C. T. G. (1992). Mortality in patients with the acquired immunodeficiency syndrome treated with either foscarnet or ganciclovir for cytomegalovirus retinitis. *N. Engl. J. Med.* 326, 213–220. doi: 10.1056/NEJM199201233260401
- Suhr, O. B., Coelho, T., Buades, J., Pouget, J., Conceicao, I., Berk, J., et al. (2015). Efficacy and safety of patisiran for familial amyloidotic polyneuropathy: a phase II multi-dose study. *Orphanet J. Rare Dis.* 10, 109. doi: 10.1186/s13023-015-0326-6
- Summerton, J., and Weller, D. (1997). Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev.* 7, 187–195. doi: 10.1089/oli.1.1997.7.187
- Sun, Y., Grimmer, M., Schwarzer, V., Schoenen, F., Fischer, U., and Wirth, B. (2005). Molecular and functional analysis of intragenic SMN1 mutations in patients with spinal muscular atrophy. *Hum. Mutat.* 25, 64–71. doi: 10.1002/humu.20111
- Terwilliger, T., and Abdul-Hay, M. (2017). Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J.* 7, e577. doi: 10.1038/bcj.2017.53
- Thomas, G. S., Cromwell, W. C., Ali, S., Chin, W., Flaim, J. D., and Davidson, M. (2013). Mipomersen, an apolipoprotein B synthesis inhibitor, reduces atherogenic lipoproteins in patients with severe hypercholesterolemia at high cardiovascular risk: a randomized, double-blind, placebo-controlled trial. *J. Am. Col. Cardiol.* 62, 2178–2184. doi: 10.1016/j.jacc.2013.07.081
- Tiercy, J.-M. (2016). How to select the best available related or unrelated donor of hematopoietic stem cells? *Haematologica* 101, 680–687. doi: 10.3324/haematol.2015.141119
- U.S. Food and Drug Administration. (2013) FDA approves new orphan drug Kynamro to treat inherited cholesterol disorder [Online]. Available: <https://wayback.archive-it.org/7993/20170112023906/http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm337195.htm>
- U.S. Food and Drug Administration. (2016). FDA grants accelerated approval to first drug for Duchenne muscular dystrophy [Online]. Available: <https://www.fda.gov/news-events/press-announcements/fda-grants-accelerated-approval-first-drug-duchenne-muscular-dystrophy>
- U.S. Food and Drug Administration. (2017). FDA briefing document: Oncologic Drugs Advisory Committee meeting; BLA 125646; Tisagenlecleucel, Available: <https://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/OncologicDrugsAdvisoryCommittee/UCM566166.pdf>
- Utsav Patel, M. B., De Léséleuc, L., and Visintini, S. 2018. 18Voretigene Neparvovec: An Emerging Gene Therapy for the Treatment of Inherited Blindness.
- Uwaydat, S. H., and Li, H. K. (2002). Pigmentary retinopathy associated with intravitreal fomivirsen. *Arch. Ophthalmol.* 120, 854–857.
- Vago, L., Oliveira, G., Bondanza, A., Noviello, M., Soldati, C., Ghio, D., et al. (2012). T-cell suicide gene therapy prompts thymic renewal in adults after hematopoietic stem cell transplantation. *Blood* 120, 1820–1830. doi: 10.1182/blood-2012-01-405670
- Varani, S., Spezzacatena, P., Manfredi, R., Chiodo, F., Mastroianni, A., Ballarini, P., et al. (2000). The incidence of cytomegalovirus (CMV) antigenemia and CMV disease is reduced by highly active antiretroviral therapy. *Eur. J. Epidemiol.* 16, 433–437. doi: 10.1023/A:1007619323939
- Veenstra, D. L., Guzauskas, G. F., Villa, K. F., and Boudreau, D. M. (2017). The budget impact and cost-effectiveness of defibrotide for treatment of veno-occlusive disease with multi-organ dysfunction in patients post-hematopoietic stem cell transplant. *J. Med. Econ.* 20, 453–463. doi: 10.1080/13696998.2016.1275652
- Vinore, S. (2003). Technology evaluation: pegaptanib, Eyetech/Pfizer. *Curr. Opin. Mol. Ther.* 5, 673–679.
- Vormittag, P., Gunn, R., Ghorashian, S., and Veraitch, F. S. (2018). A guide to manufacturing CAR T cell therapies. *Curr Opin. Biotechnol.* 53, 164–181. doi: 10.1016/j.copbio.2018.01.025
- Waldrop, M.A., and Kolb, S.J. (2019). Current treatment options in neurology—SMA therapeutics. *Curr. Treat. Options Neurol.* 21, 25. doi: 10.1007/s11940-019-0568-z
- Whitmore, K. V., and Gaspar, H. B. (2016). Adenosine Deaminase Deficiency – More Than Just an Immunodeficiency. *Front. Immunol.* 7, 314. doi: 10.3389/fimmu.2016.00314
- Wilson, J. M. (2005). Gendicine: The first commercial gene therapy product; Chinese translation of editorial. *Human Gene Ther.* 16, 1014–1015. doi: 10.1089/hum.2005.16.1014
- Wong, T., Chakravarthy, U., Klein, R., Mitchell, P., Zlateva, G., Buggage, R., et al. (2008). The natural history and prognosis of neovascular age-related macular degeneration: a systematic review of the literature and meta-analysis. *Ophthalmology* 115, 116–126. doi: 10.1016/j.ophtha.2007.03.008
- Wong, W. L., Su, X., Li, X., Cheung, C. M. G., Klein, R., Cheng, C.-Y., et al. (2014). Global prevalence of age-related macular degeneration and disease burden projection for 2020 and 2040: a systematic review and meta-analysis. *Lancet Global Health* 2, e106–e116. doi: 10.1016/S2214-109X(13)70145-1
- Wood, H. (2018). 2FDA approves patisiran to treat hereditary transthyretin amyloidosis. *Nat. Rev. Neurol.* 14, 570–570. doi: 10.1038/s41582-018-0065-0
- Xia, Z.-J., Chang, J.-H., Zhang, L., Jiang, W.-Q., Guan, Z.-Z., Liu, J.-W., et al. (2004). Phase III randomized clinical trial of intratumoral injection of E1B gene-deleted adenovirus (H101) combined with cisplatin-based chemotherapy in treating squamous cell cancer of head and neck or esophagus. *Ai zheng*, 23, 1666–1670.
- Zanetta, C., Nizzardo, M., Simone, C., Monguzzi, E., Bresolin, N., Comi, G. P., et al. (2014). Molecular therapeutic strategies for spinal muscular atrophies: current and future clinical trials. *Clin. Ther.* 36, 128–140. doi: 10.1016/j.clinthera.2013.11.006
- Zhang, S., Xiao, S., Liu, C., Sun, Y., Su, X., Li, D., et al. (2003). Treatment of head and neck squamous cell carcinoma by recombinant adenovirus-p53 combined with radiotherapy: a phase II clinical trial of 42 cases. *Zhonghua yi xue za zhi* 83, 2023–2028.
- Zhang, S., Xiao, S., Liu, C., Sun, Y., Su, X., Li, D., et al. (2005). Recombinant adenovirus-p53 gene therapy combined with radiotherapy for head and neck squamous-cell carcinoma. *Zhonghua zhong liu za zhi [Chinese journal of oncology]* 27, 426–428.
- Zhang, W.-W., Li, L., Li, D., Liu, J., Li, X., Li, W., et al. (2018). The first approved gene therapy product for cancer ad-p53 (gendicine): 12 years in the clinic. *Human Gene Ther.* 29, 160–179. doi: 10.1089/hum.2017.218

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shahryari, Saghaeian Jazi, Mohammadi, Razavi Nikoo, Nazari, Hosseini, Burtscher, Mowla and Lickert. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genotype–Phenotype Association Analysis Reveals New Pathogenic Factors for Osteogenesis Imperfecta Disease

Jingru Shi^{1†}, Meng Ren^{1†}, Jinmeng Jia¹, Muxue Tang¹, Yongli Guo^{2,3,4*}, Xin Ni^{2,3,4*} and Tielu Shi^{1,2*}

¹ Center for Bioinformatics and Computational Biology, and the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai, China, ² Big Data and Engineering Research Center, Beijing Key Laboratory for Pediatric Diseases of Otolaryngology, Head and Neck Surgery, MOE Key Laboratory of Major Diseases in Children, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China, ³ Biobank for Clinical Data and Samples in Pediatrics, Beijing Children's Hospital, National Center for Children's Health, Beijing Pediatric Research Institute, Capital Medical University, Beijing, China, ⁴ Department of Otolaryngology, Head and Neck Surgery, Beijing Children's Hospital, National Center for Children's Health, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:

Peter Vee Sin Lee,
The University of Melbourne,
Australia

Reviewed by:

Amit P. Bhavsar,
University of Alberta,
Canada
Qing Lyu,
University of Rochester,
United States

*Correspondence:

Yongli Guo
guoyongli@bch.com.cn
Xin Ni
nixin@bch.com.cn
Tielu Shi
tlshi@bio.ecnu.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 16 November 2018

Accepted: 17 September 2019

Published: 15 October 2019

Citation:

Shi J, Ren M, Jia J, Tang M, Guo Y,
Ni X and Shi T (2019) Genotype–
Phenotype Association Analysis
Reveals New Pathogenic Factors for
Osteogenesis Imperfecta Disease.
Front. Pharmacol. 10:1200.
doi: 10.3389/fphar.2019.01200

Osteogenesis imperfecta (OI), mainly caused by structural abnormalities of type I collagen, is a hereditary rare disease characterized by increased bone fragility and reduced bone mass. Clinical manifestations of OI mostly include multiple repeated bone fractures, thin skin, blue sclera, hearing loss, cardiovascular and pulmonary system abnormalities, triangular face, dentinogenesis imperfecta (DI), and walking with assistance. Currently, 20 causative genes with 18 subtypes have been identified for OI, of them, variations in *COL1A1* and *COL1A2* have been demonstrated to be major causative factors to OI. However, the complexity of the bone formation process indicates that there are potential new pathogenic genes associated with OI. To comprehensively explore the underlying mechanism of OI, we conducted association analysis between genotypes and phenotypes of OI diseases and found that mutations in *COL1A1* and *COL1A2* contributed to a large proportion of the disease phenotypes. We categorized the clinical phenotypes and the genotypes based on the variation types for those 155 OI patients collected from literature, and association study revealed that three phenotypes (bone deformity, DI, walking with assistance) were enriched in two variation types (the Gly-substitution missense and groups of frameshift, nonsense, and splicing variations). We also identified four novel variations (c.G3290A (p.G1097D), c.G3289C (p.G1097R), c.G3289A (p.G1097S), c.G3281A (p.G1094D)) in gene *COL1A1* and two novel variations (c.G2332T (p.G778C), c.G2341T (p.G781C)) in gene *COL1A2*, which could potentially contribute to the disease. In addition, we identified several new potential pathogenic genes (*ADAMTS2*, *COL5A2*, *COL8A1*) based on the integration of protein–protein interaction and pathway enrichment analysis. Our study provides new insights into the association between genotypes and phenotypes of OI and novel information for dissecting the underlying mechanism of the disease.

Keywords: osteogenesis imperfecta, genotype, phenotype, novel candidate pathogenic genes, novel candidate pathogenic variations

INTRODUCTION

Osteogenesis imperfecta (OI) is a phenotypically and genetically heterogeneous group of bone disorders characterized by bone fragility and skeletal deformity, owing to the abnormality of type I collagen formed by two $\alpha 1(I)$ chains (encoded by *COL1A1* gene) and one $\alpha 2(I)$ chain (encoded by *COL1A2* gene). Individuals with OI have low bone mass, which results in deformity of long bones, vertebral anomalies and fractures, shortening of extremities, and skull defect (Marini et al., 2007). The observed extra-skeletal phenotypes include dentinogenesis imperfecta (DI), thin skin, blue sclera, scoliosis, cardiovascular and pulmonary system abnormalities, triangular face, and hearing impairment (Foster et al., 2014; Marini et al., 2017). Previous studies categorize OI into four subtypes (types I–IV) based on clinical findings, inheritance patterns, and radiographic features: OI type I is the mildest form, OI type II is the perinatal lethal form, while OI type III is the most severe form, and OI type IV is characterized by the mild to moderate form (Sillence et al., 1979; Rauch et al., 2010; Lin et al., 2015; Mrosk et al., 2018). With an in-depth understanding of OI disease, more subtypes have been defined and added into OI's original classification system, making the number of subtypes updated to 18 (Forlino and Marini, 2016; Marini et al., 2017; Lu et al., 2019).

Current evidences demonstrate that *COL1A1* and *COL1A2* are the main factors in the cause of OI, as approximately 85% to 90% of cases are disturbed by them, and all of the four subtypes are involved in *COL1A1* and *COL1A2* genes (<http://www.le.ac.uk/ge/collagen/>). There are two general categories of mutational defects occurred in *COL1A1/COL1A2*. The first is missense mutation, mainly involving glycine replacement within the Gly-Xaa-Yaa repeat (the Gly-substitution missense), which results in the synthesis of collagen with abnormal structure (Lin et al., 2015). The second is a group of variations that include frameshift, nonsense, and splicing mutations, which mainly lead to the reduced amount of normal type I collagen. Previous studies have shown that the second variation group is often associated with milder phenotypes, while the Gly-substitution missense usually lead to more severe phenotypes (Rauch et al., 2010; Zhang et al., 2012). Considering the phenotypic specificity of the Gly-substitution missense, we would like to investigate more potentially pathogenic Gly-substitution mutations for OI mechanism exploration.

In addition to the confirmed OI-related collagen genes (*COL1A1* and *COL1A2*), in the past decade, a series of studies have found that a set of new non-collagen gene defects affect normal post-translational processing, molecular folding of type I collagen, fibril formation, osteoblast differentiation, and mineralization, leading to rare autosomal recessive, dominant, and X-linked forms of OI (Bregou Bourgeois et al., 2016; Lindert et al., 2016; Marom et al., 2016; Marini et al., 2017). With the rapid development of next-generation sequencing technology, almost 18 pathogenic non-collagen genes have gradually been identified (Forlino and Marini, 2016; Marini et al., 2017; Mrosk et al., 2018), including *BMP1*, *CRTAP*, *P3H1*, *PP1B*, *TMEM38B*, *SERPINH1*, *FKBP10*, *PLOD2*, *IFITM5*, *SERPINF1*, *WNT1*, *CREB3L1*, *SP7*, *SPARC*, *MBTPS2*, *P4HB*, *PLS3*, and

SEC24D. Based on the complexity of bone formation and clinical observation, we believe that new potential disease-related genes remain to be identified.

Genotype and phenotype associations can provide new insights into understanding the disease mechanism (Geng et al., 2017; Li et al., 2017). The phenotypic severity depends not only on the affected gene, but also on the position of the mutation in the gene. To identify new missense mutations associated with OI, in the present study, we firstly collected genotypic and phenotypic information on 155 patients from literature and evaluated the genotype–phenotype associations. Next, we identified a set of disease-associated variations in *COL1A1* and *COL1A2* by integrative analysis with several software designed to predict functional effect of human missense mutations. In addition, considering the fact that each biological function is accomplished by the interactions of multiple proteins, we performed network-based analysis and pathway enrichment analysis to identify novel candidate risk genes potentially contributing to the development of OI. Considering limited availability of the patient size and the complex pathogenesis for OI, our comprehensive analysis could promote better understanding of OI in the clinical diagnose, genetic counseling, and prenatal diagnosis.

MATERIALS AND METHODS

Data Resources

The 20 confirmed OI pathogenic genes and their related reported variant information were extracted from the Osteogenesis Imperfecta Variant Database (OIVD) (e.g., “DNA change,” “mutation effect,” “protein,” “reference,” etc.) (Fokkema et al., 2011). We manually collected 155 OI patient information from published literature, including genes, mutation type, phenotypes, age, and gender.

InWeb_InBioMap is a scored human protein–protein interaction data resource. It is generated by combining interactions from eight protein interaction databases and providing confidence score for each interacting pair and relevance score for every protein. The confidence score represents a lower bound on the probability for the interaction being a true positive, and the relevance score represents a relationship between one protein and others in a specific network. We selected those interactive pairs with both confidence score of 1 and a relevance score of 1 for subsequent study (Li et al., 2017).

We retrieved the expression patterns for those predicted genes and pathogenic genes in different tissues from GTEx database. GTEx is a data resource platform for exploring correlations between genetic variants and gene expression in multiple human tissues (Consortium, 2013).

We standardized the phenotypic description for those 155 patients based on eRAM system (Jia et al., 2018; Ni and Shi (2017)). The collected information of those cases has been stored into eRAM and PedAM (Jia et al., 2018). eRAM is a comprehensively standardized data resource for rare disease by integrating massive text mining results and data from multiple databases. Currently, there are standardized 15,942 rare diseases with corresponding phenotypes recorded in eRAM.

WikiPathways is an open and integrative database, providing biology pathway information (Slenter et al., 2018). WebGestalt supports multiple functional enrichment analysis based on selected different organism, enrichment method and functional databases (Wang et al., 2017). We conducted pathway enrichment analysis based on WikiPathway functional database and only focused on OI-related pathways among those significant ones ($P < 0.05$).

The Pathogenic Analysis for Variants in OI-Related Genes

To avoid analysis bias resulted from the insufficient reported cases, we only performed the prediction of new pathogenic OI-related variants on *COL1A1* and *COL1A2* genes. All variations information on *COL1A1* and *COL1A2* genes was downloaded from ANNOVAR database (Wang et al., 2010).

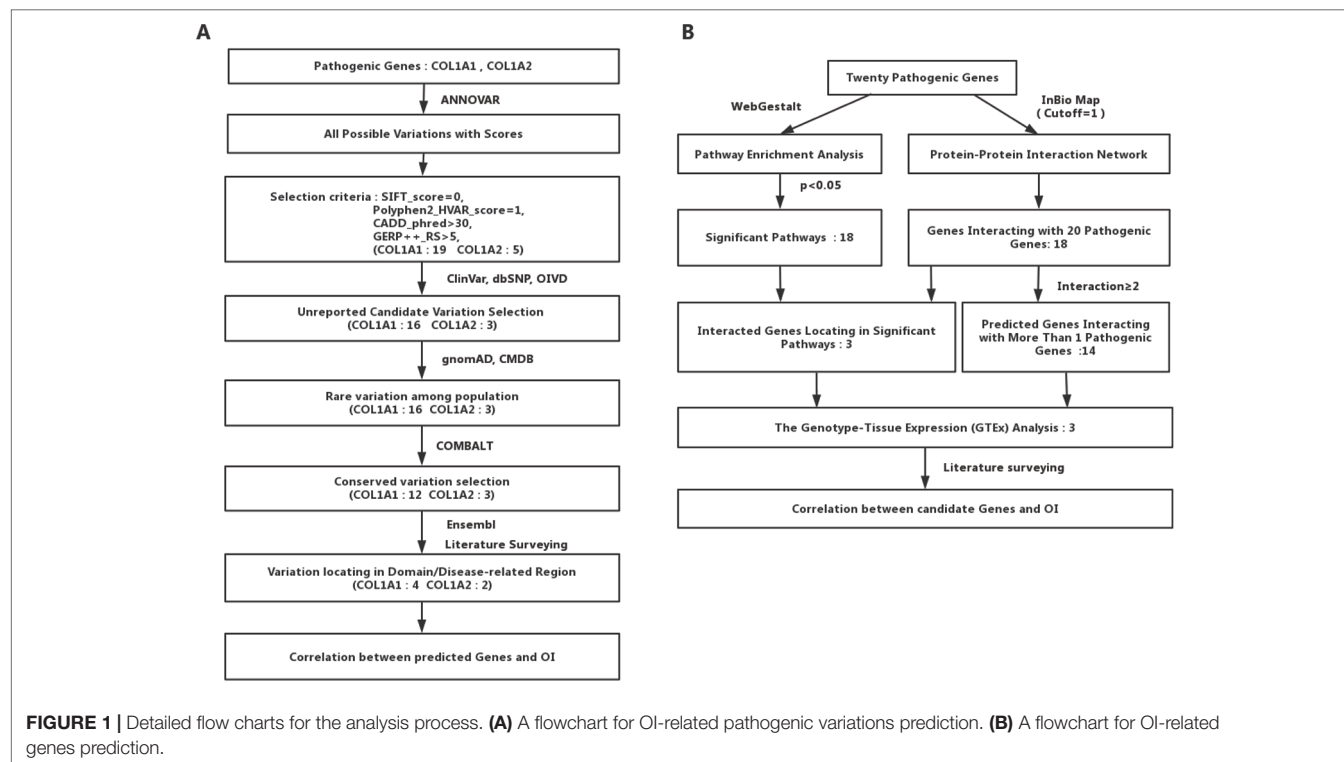
ANNOVAR provides functional annotation of single-nucleotide variants (SNVs), insertions, and deletions with more than 10 different tools, including SIFT, CADD, PolyPhen-2, and GERP++. The pathogenic ability of each variation is predicted by those tools with defined cutoff values. SIFT is a tool that uses sequence homology to predict whether a substitution affects protein function and whether amino acid substitutions at specific positions of the protein have phenotypic effects (Ng and Henikoff, 2001). CADD integrates allelic diversity, pathogenicity, functional annotations, and severity of disease, which has the ability to rank known pathogenic mutations by disease severity in individual genome (Kircher et al., 2014). PolyPhen-2 predicts the functional significance of allelic replacement with multiple

parameters. We used HumVar-trained results in this tool (Adzhubei et al., 2010). GERP++ recognizes high-resolution regions with nucleotide substitution defects and measures these defects as “rejected substitutions” (Cooper et al., 2005). We used the default cutoff values for each of those four softwares (SIFT_score = 0, Polyphen2_HVAR_score = 1, CADD_phred > 30, GERP++_RS > 5) to define the possible pathogenic ability of each variation on *COL1A1*/*COL1A2* gene.

We collected the allele frequency for corresponding variants from gnomAD (Lek et al., 2016) and Chinese Millionome Database (CMDB) (Liu et al., 2018). To further analyze the pathogenicity of those predicted mutations based on the regional conservation, we used COMBALT (Constraint-based Multiple Alignment Tool) to conduct multiple-sequence alignment and identify the conservation for those mutations (Papadopoulos and Agarwala, 2007). At last, we searched the domain information of every predicted mutation by literature surveying and Ensembl genome browser 98 (Cunningham et al., 2019). The whole process was displayed in **Figure 1A**.

Novel Candidate Risk Genes Identification

To identify potential new risk factor for OI, we first mapped those 20 experimentally confirmed pathogenic genes into InWeb_InBioMap data resource to form a disease gene centralized protein-protein interaction network. Those genes with the largest confidence score (= 1) and the highest relevance score (= 1) to the OI pathogenic genes were selected to be interacting genes for OI. We then constructed a sub-network of pathogenic proteins and their interacting protein partners and focused on



those interacting proteins (the predicted gene set A) that directly interact with more than one OI pathogenic proteins. In addition, we used WebGestalt to conduct pathway enrichment analysis based on Wikipathway database and selected those significant pathways ($P < 0.05$). Subsequently, we mapped those interacting gene partners of OI pathogenic genes into those significant pathways and selected those gene partners (the predicted gene set B) involved in these pathways. We then checked the expression patterns of those predicted OI-related pathogenic genes (the union of predicted gene set A and predicted gene set B). Last, we conducted literature survey to further verify those novel candidate risk genes. The whole process was displayed in **Figure 1B**.

Statistical Analysis

Two-sided Fisher's exact test was used to test the difference ($P < 0.05$) between the Gly-substitution missense and other types of variation (frameshift, nonsense, splicing) among OI subtypes, gender, pathogenic genes, and the presence of each clinical features in OI patients. One-sided Fisher's exact test was used to test co-occurrence relationship ($P < 0.05$) between two different phenotypes. All the calculated progress was conducted by R software 3.4.3.

RESULT

We extracted all experimentally confirmed OI-related pathogenic genes from the OIVD, and those pathogenic genes are *COL1A1*, *COL1A2*, *BMP1*, *CRTAP*, *P3H1*, *PPIB*, *TMEM38B*, *SERPINH1*, *FKBP10*, *PLOD2*, *IFITM5*, *SERPINF1*, *WNT1*, *CREB3L1*, *SP7*, *SPARC*, *MBTPS2*, *P4HB*, *PLS3*, and *SEC24D*. As less reported cases available for other 18 pathogenic non-collagen genes, we just collected cases with the variations reported in collagen gene *COL1A1* and *COL1A2* from literature.

Currently, there are four different subtypes (OI type I to IV) mainly caused by *COL1A1* and *COL1A2*, which correlate with different defects on type I collagen caused by different variation types (Forlino and Marini, 2016). We classified all variations into two variation categories (Rauch et al., 2010; Lin et al., 2015; Lindahl et al., 2015). The first variation type is missense in *COL1A1/COL1A2*, which mainly includes the Gly-substitution missense in the triple helix, leading to qualitative effects on type I collagen. The second variation type includes frameshift, nonsense, and splicing variations in *COL1A1/COL1A2* (Lin et al., 2015), causing a quantitative influence on type I collagen. The type of every variation in all patients was determined based on the results from literature and OIVD. Variations without recorded variation type were excluded. Among all the 155 patients, there were 59 Gly-substitution missenses, 70 variations of frameshift, nonsense, and splicing, and 26 non-Gly-substitution missenses or without variation type. We just analyzed 129 patients with Gly-substitution missense, frameshift, nonsense, or splicing variation.

Genotype–Phenotype Association Analysis

To explore the associations between genotypes of pathogenic genes and OI clinical phenotypes, we first standardized the phenotypic description based on our eRAM system for those

collected 155 patients (**Supplementary Table 1**), such as bone deformity, dense metaphyseal bands, blue sclera, hearing loss, and so on, and obtained 12 phenotypic categories. Among 129 analyzed patients, there were 49 OI type I patients, 1 OI type II patient, 10 OI type III patients, 29 OI type IV patients, and 40 patients without defined OI type. The OI type II is generally perinatal lethal type, and the data for this type were insufficient and excluded for subsequent analysis.

Based on the two types of variation, we performed statistical significance test between the two variation types and 12 phenotypes recorded in the collected 129 cases. As a result, these two variation groups were significantly different among OI type I, OI type III, and OI type IV ($p = 1.526e-07$), most OI type I were caused by the second variation type (frameshift, nonsense, and splicing). The second variation type seemed to occur mainly in *COL1A1* ($p = 1.833e-13$), and no variation type difference was observed between male and female patients ($p = 0.7218$).

Among those 12 phenotypes, blue sclera was one of the most common phenotypes in OI disease. The number of patients with blue sclera was up to 128, accounting for 99.2% of 129 patients. However, there were no significant differences between two variation types in blue sclera ($p = 0.4243$) (**Table 1A**). In addition, no significant difference was observed between patients with the Gly-substitution missense and patients with other variations (frameshift, nonsense and splicing) for hypermobile joints ($p = 1$), dense metaphyseal bands ($p = 0.4231$), vertebral fracture ($p = 1$), osteopenia ($p = 1$), hearing loss ($p = 0.6024$), triangular face ($p = 0.389$), and popcorn calcif ($p = 1$) (**Table 1A**).

Results revealed that patients with the Gly-substitution missense tended to develop bone deformity ($p = 0.02946$) and DI ($p = 0.03189$). In addition, we also found that patients with DI were prone to have bone deformity ($p = 0.0005576$) or vertebral anomalies ($p = 0.01832$) (**Table 1B**). Scoliosis, which composes part of vertebral anomalies, is one of the mostly common phenotypes in OI. Previous study also indicated that children with DI had a large probability in having scoliosis, pathological kyphosis, and basilar impression (Engelbert et al., 1998). The association between the first mutation type and bone deformity/DI provides new insight into the research on spinal-related abnormalities disease, like spinal complications. Although DI had a strong relationship with vertebral anomalies, no significant difference was observed between two variation types in vertebral anomalies (vertebral anomalies, $p = 0.06021$). In addition, patients with the Gly-substitution missense had a poorer walking ability ($p = 0.0001345$).

Novel Candidate Pathogenic Variations Identification

We obtained 55,032 possible mutations of *COL1A1* and 110,016 possible mutations of *COL1A2* in ANNOVAR. After filtering out mutations based on the cutoff value (SIFT_score = 0, Polyphen2_HVAR_score = 1, CADD_phred > 30, GERP++_RS > 5), 19 *COL1A1* mutations and 5 *COL1A2* mutations were kept as pathogenic variations. Strikingly, 5 of the 24 mutations

TABLE 1 | Statistical analysis result on genotype–phenotype correlation and phenotype–phenotype correlation.

(A) Relationship between clinical characteristics and different variation types (Gly-substitution mutation vs. Frameshift, nonsense, and splicing mutation) in COL1A1 and COL1A2 of 129 patients with OI.

Phenotype	Gly-substitution missense (n = 59)	Other variation (Frameshift, nonsense, and splicing) (n = 70)	Total case (Gly/other)	P value
OI Type(I/III/IV)	23.7%/23.7%/52.6%	80%/2%/18%	38/50	1.53E-07
Gender(F/M)	54.2%/45.8%	58.6%/41.4%	59/70	0.7218
COL1A1/COL1A2 mutation	40.7%/59.3%	97.1%/2.9%	58/70	1.83E-13
Bone deformity	80%	58.50%	45/53	0.02946
Hypermobility joints	83.3%	100%	12/3	1
Dense metaphyseal bands	81.8%	50.0%	11/2	0.4231
Vertebral anomalies	51.1%	31.3%	45/48	0.06021
Vertebral fracture	40.0%	42.1%	20/19	1
Osteopenia	100%	100%	1/2	1
Dentinogenesis imperfecta	56.9%	35.2%	51/54	0.03189
Blue sclera	70.7%	77.1%	58/72	0.4243
Hearing loss	14.0%	18.5%	50/54	0.6024
Walking with assistance	31.0%	0%	29/42	0.0001345
Triangular face	25.8%	16.7%	31/42	0.3889
Popcorn calcif	25.0%	0%	12/2	1

(B) Relationship between dentinogenesis imperfecta and bone deformity/vertebral anomalies. Phenotype with significant difference ($p < 0.05$) are represented in bold font.

Phenotype	Dentinogenesis imperfecta (Y)	Dentinogenesis imperfecta (N)	Total case (DI(Y)/DI(N))	P value
Bone deformity	85.10%	52.90%	47/51	0.0005576
Vertebral anomalies	53.50%	30%	43/50	0.01832

have been reported to be pathogenic to OI in ClinVar, dbSNP or OIVD Database (**Figures 2A, B**). The remaining 16 candidate mutations in *COL1A1* and three candidate mutations in *COL1A2* currently have no supporting evidence.

Next, we checked the frequency of those candidate variations in gnomAD and CMDB databases, and no frequency has been reported in any ethnic groups. The result indicated that all the variations are an extremely rare site with high probability of lethality for the disease. The multiple sequence alignment of *COL1A1/COL1A2* homologous genes among seven species (*Homo sapiens*, *Pan troglodytes*, *Papio anubis*, *Macaca mulatta*, *Mus musculus*, *Xenopus laevis*, *Caenorhabditis elegans*) with COMBALT revealed that 12 mutations in *COL1A1* and 3 mutations in *COL1A2* located in highly conserved regions among those species. Finally, we mapped the 15 candidate variations into the domains of $\alpha 1(I)$ chain/ $\alpha 2(I)$ chain (encoded by *COL1A1/COL1A2*) in Ensembl, and found only four mutations (c.G3290A (p.G1097D), c.G3289C (p.G1097R), c.G3289A (p.G1097S), c.G3281A (p.G1094D)) in *COL1A1* located in collagen triple helix repeat domain, which influence type I collagen formation. In contrast, two mutations (c.G2332T (p.G778C), c.G2341T (p.G781C)) in *COL1A2* were not in any domain region, but located in two reported lethal spaces in *COL1A2* protein sequence. We supposed that these two candidate variations also associate with OI disease.

Novel Candidate Risk Genes Identification

To identify new risk genes related to OI, we carried out network-based integrative analysis, the assumption of our approach is that

if a protein directly interacts with more than one OI causative proteins or is enriched in the same pathway with known causal genes, theoretically, the protein would be a candidate that also contributes to OI disease. Therefore, we first mapped all 20 pathogenic genes into the InWeb_InBioMap database to construct a PPI network, then we selected those proteins with confidence score cutoff ($= 1$) and relevance score cutoff ($= 1$). The resulting network contains 20 pathogenic proteins with 18 directly interactive partners (*ADAMTS2*, *ADAMTS3*, *ADAMTS14*, *COL4A6*, *COL5A2*, *COL8A1*, *COL19A1*, *COL20A1*, *COL21A1*, *COL22A1*, *COL24A1*, *COL27A1*, *COL28A1*, *TLL1*, *CNIH1*, *CNIH3*, *SEC16B*, *WNT8B*) (≥ 1 interaction) (**Figure 3A**). We extracted 14 interactive proteins (*ADAMTS2*, *ADAMTS3*, *ADAMTS14*, *COL4A6*, *COL5A2*, *COL8A1*, *COL19A1*, *COL20A1*, *COL21A1*, *COL22A1*, *COL24A1*, *COL27A1*, *COL28A1*, *TLL1*) that interact with more than one pathogenic proteins from PPI network as the predicted gene set A (≥ 2 interactions). Next, we performed pathway enrichment analysis on 20 disease-causing genes and obtained 18 significant pathways ($p < 0.05$). Among the 18 interacting gene partners, 3 of them (*COL4A6*, *COL5A2*, *WNT8B*) fell into those 18 significant pathways (the predicted gene set B) (**Table 2**). Based on the integrated result of the protein–protein network (14 interactive proteins from the predicted gene set A) and pathway enrichment analysis (3 interactive proteins from the predicted gene set B), we finally identified 15 genes (*ADAMTS2*, *ADAMTS3*, *ADAMTS14*, *COL4A6*, *COL5A2*, *COL8A1*, *COL19A1*, *COL20A1*, *COL21A1*, *COL22A1*, *COL24A1*, *COL27A1*, *COL28A1*, *TLL1*, *WNT8B*) as the potential causal genes to OI disease. To further verify those predicted genes for OI disease, we checked the expression

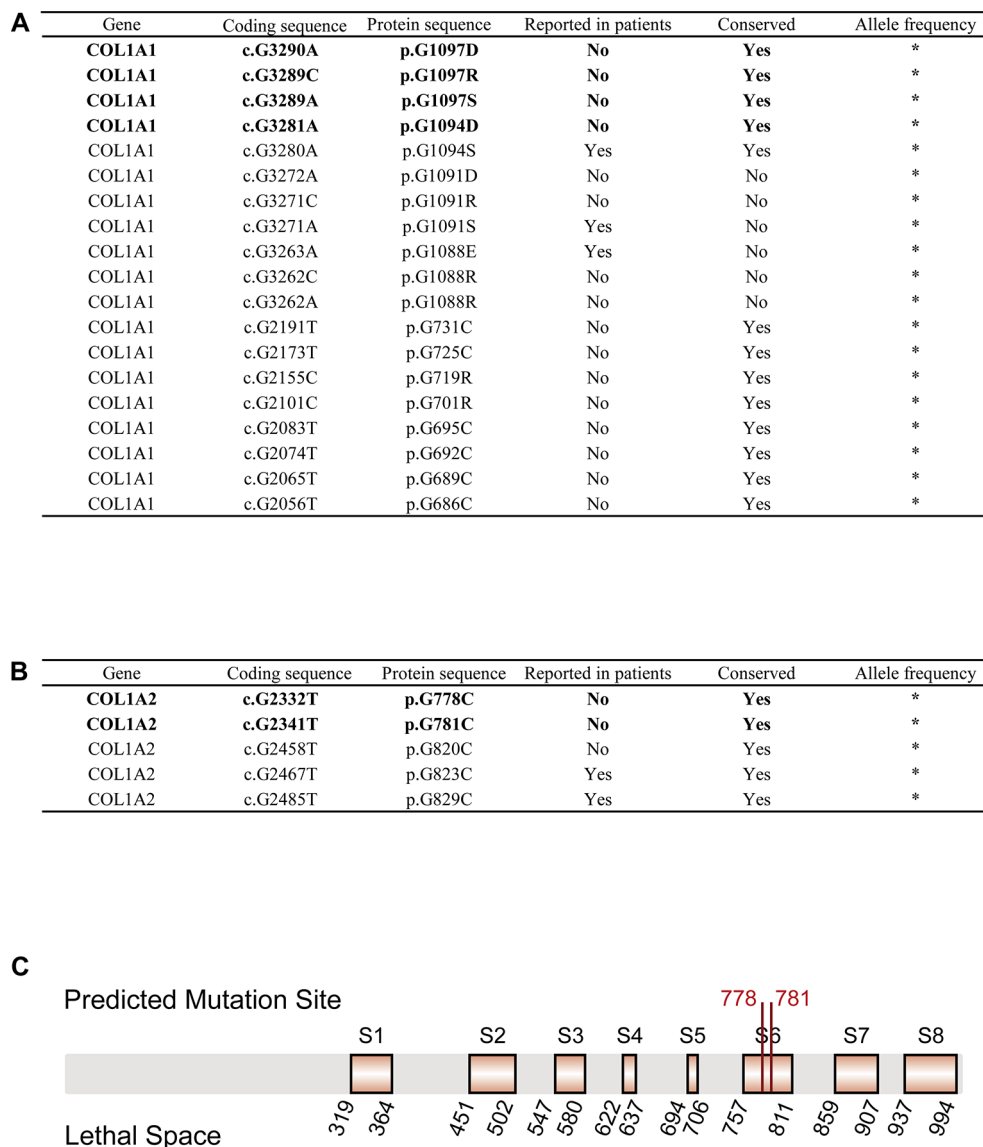


FIGURE 2 | OI-related candidate mutations in *COL1A1* and *COL1A2*. **(A)** OI-related candidate mutations in *COL1A1* ("*": currently unknown). **(B)** OI-related candidate mutations in *COL1A2* ("*": currently unknown). **(C)** Localization of variants c.G2332T (p.G778C) and c.G2341T (p.G781C) in *COL1A2* protein sequence. Predicted pathogenic variations are represented in bold font (Eight orange rectangles with black border denote eight lethal spaces).

pattern of those 15 genes in different human tissues based on GTEx data. As related to collagen hereditary disease, most of the OI causative genes are expressed in artery aorta and transformed fibroblasts (**Supplementary Figure**). Artery aorta is rich in connective tissue, and fibroblasts secrete a variety of substrates, collagen and fibers. After filtering out those predicted genes that do not show similar tissue expression pattern with that of most pathogenic genes, only three candidate genes were left (*ADAMTS2*, *COL5A2*, and *COL8A1*). Those three candidate genes presented similarly tissue expression pattern with their seven interactive OI pathogenic genes (*COL1A1*, *COL1A2*, *PPIB*, *SERPINH1*, *P3H1*, *BMP1*, *CRTAP*), especially showed high expression in transformed fibroblasts (**Figure 3B**).

DISCUSSION

OI is a rare disease with bone disorders characterized by bone fragility and skeletal deformity. Clinical observation indicates that currently identified pathogenic genes and variations cannot fully decipher the phenotypic and genetic heterogeneity of the disease. To better explore the underlying mechanism of OI, we performed genotype and phenotype association analysis based on manually collected 155 patients' data. According to the variation type, we classified mutations into two groups, one group refers to the Gly-substitution missense, and another includes frameshift, nonsense, and splicing mutation. Most of the Gly-substitution missenses result in structural abnormalities of type I

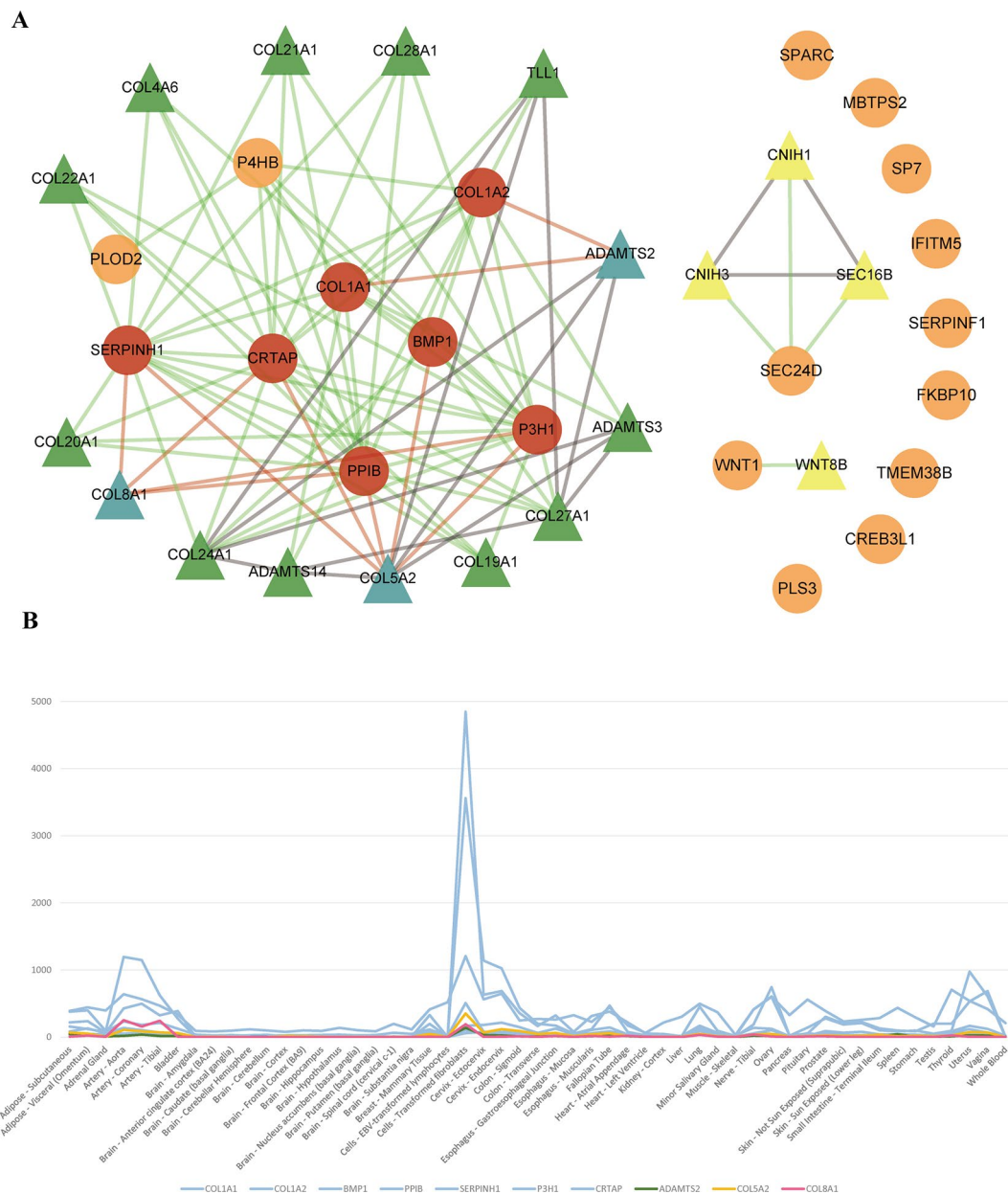


FIGURE 3 | (A) Network of OI pathogenic genes and candidate genes. 93 interactions between 20 OI pathogenic genes and 18 interacted genes. The circles and triangles represent 20 OI pathogenic genes and 18 interacted genes (≥ 1 interaction), respectively. Yellow triangles denote four predicted genes ($= 1$ interaction). Green and blue triangles totally denote 14 predicted genes (the predicted gene set A) (≥ 2 interactions), and blue triangles denote three candidate genes (gene expression is consistent with OI pathogenic genes), respectively. Red circles denote seven OI pathogenic genes (interact with three candidate genes). Grey lines denote 15 interactions between predicted genes. Green lines and red lines totally denote 78 interactions between OI pathogenic genes and predicted genes, and red lines denote 11 interactions between three candidate genes and seven interacted OI pathogenic genes, respectively. **(B)** The tissue expression distribution of predicted genes and their interacted pathogenic genes. The green, yellow, red, and blue lines denote candidate gene *ADAMTS2*, *COL5A2*, *COL8A1* and interacted pathogenic genes (*COL1A1*, *COL1A2*, *BMP1*, *PPIB*, *SERPINH1*, *P3H1*, *CRTAP*). All these genes are expressed highly in “Cells_Transformed_fibroblasts” and have a similar expression distribution among all human tissues.

collagen, whereas the second mutation group caused by the early termination of codons lead to insufficient collagen synthesis. Our results showed that the two mutation types played different roles in bone deformity, DI, and walking with assistance, which was consistent with the results of other studies (Lin et al., 2015).

In addition, we found that patients with DI were more likely to have bone deformity ($p = 0.0005576$) and vertebral anomalies ($p = 0.01832$). Similarly, Lin et al. also reported that patients with DI were more susceptible to bone deformities and scoliosis (Lin et al., 2015). Most of the patients in our collected

TABLE 2 | Significant pathways with related pathogenic genes and predicted genes.

Significant Pathways	20 Pathogenic Genes and 18 Interacted Genes	P value
Epithelial to mesenchymal transition in colorectal cancer	COL4A6, WNT8B, SPARC, WNT1	0.00083567
Focal Adhesion-PI3K-Akt-mTOR-signaling pathway	COL4A6, COL5A2, COL1A1, COL1A2, CREB3L1	0.0010862
miR-509-3p alteration of YAP1/ECM axis	COL1A1, SPARC	0.0011611
Focal Adhesion	COL4A6, COL5A2, COL1A1, COL1A2	0.0018166
Inflammatory Response Pathway	COL1A1, COL1A2	0.0035058
miRNA targets in ECM and membrane receptors	COL5A2, COL1A2	0.0059006
PI3K-Akt Signaling Pathway	COL4A6, COL1A1, COL1A2, CREB3L1	0.012283
Sterol Regulatory Element-Binding Proteins (SREBP) signaling	MBTPS2, SEC24D	0.01637
Hypothetical Craniofacial Development Pathway	WNT1	0.021586
ncRNAs involved in Wnt signaling in hepatocellular carcinoma	SERPINF1, WNT1	0.023282
LncRNA involvement in canonical Wnt signaling and colorectal cancer	SERPINF1, WNT1	0.031757
Senescence and Autophagy in Cancer	COL1A1, SPARC	0.034042
EDA Signaling in Hair Follicle Development	BMP1	0.037488
Osteoblast Signaling	COL1A1	0.037488
Wnt Signaling Pathway	SERPINF1, WNT1	0.040029
Canonical and Non-Canonical TGF- β signaling	BMP1	0.045347
Regulation of Wnt/ β -catenin Signaling by Small Molecule Compounds	WNT1	0.047953
Adipogenesis	BMP1, WNT1	0.04837

Eighteen significant pathways ($P < 0.05$) were identified with 20 OI causative genes and their 18 interacting gene partners from protein–protein interaction network. Candidate genes (the predicted gene set B) are represented in bold font.

data with both two phenotypes contain the Gly-substitution missense (DI and bone deformity: 62.5%; DI and vertebral anomalies: 78.3%). Based on the observation, we supposed that the collagen structural abnormalities caused by the Gly-substitution missense could result in the abnormal development of the whole body bone morphology, which could also be one of the pathogenic factors for other bone disorders. Especially, the strong co-occurrence of DI and vertebral anomalies implied that *COL1A1* and *COL1A2* genes could also contribute to other spinal and vertebral diseases.

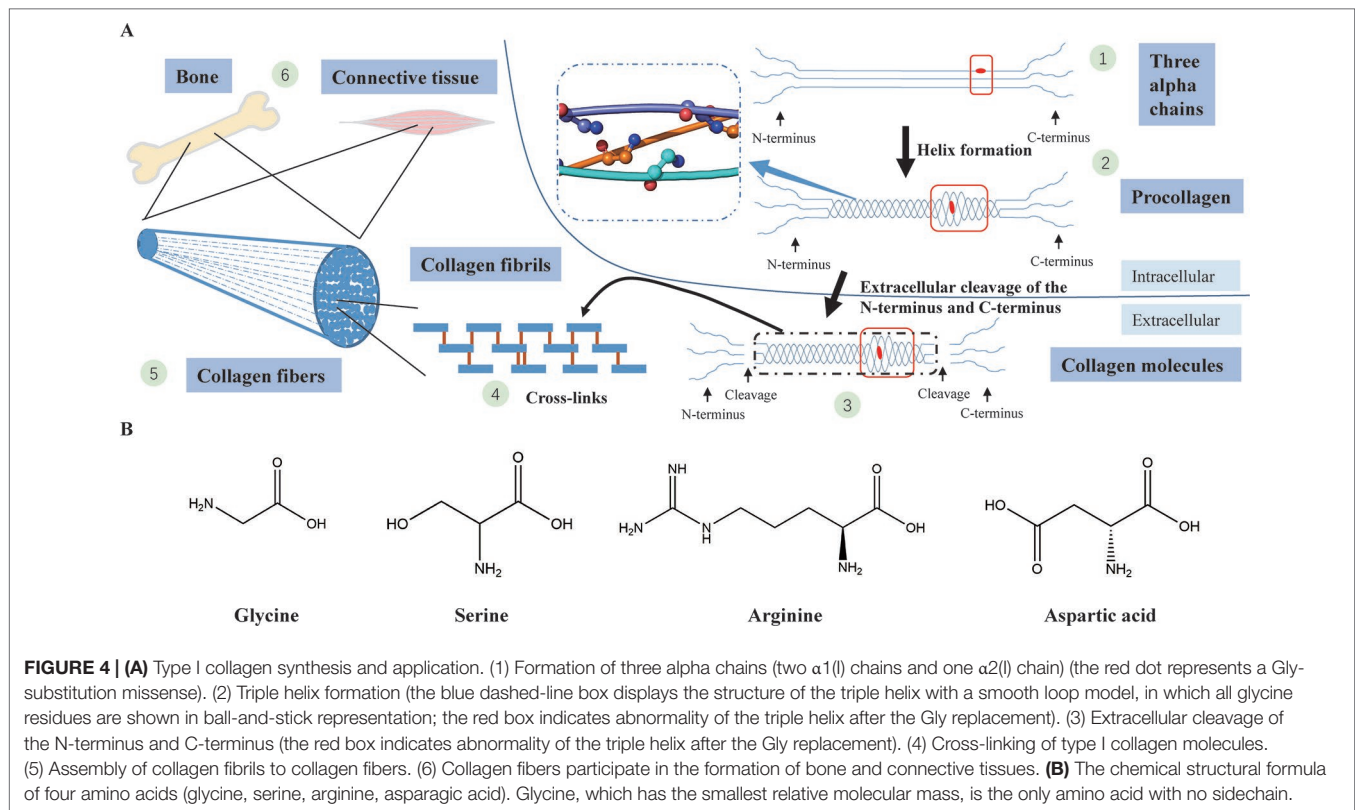
For the 24 pathogenic mutations from ANNOVAR annotation, we searched different variation annotation databases, including ClinVar, dbSNP, and OIVD, to explore whether they have any disease-related report. As a result, 19 of the mutations have no report linked to OI patients. However, multiple alignment with homologous genes from different species with COMBALT showed that only six variations (*COL1A1*: c.G3290A (p.G1097D), c.G3289C (p.G1097R), c.G3289A (p.G1097S), and c.G3281A

(p.G1094D); *COL1A2*: c.G2332T (p.G778C) and c.G2341T (p.G781C)) locate in the conserved regions, which indicates that these four positions (amino acid position: 1094 and 1097 (on $\alpha 1(I)$ chain); 778 and 781 (on $\alpha 2(I)$ chain)) have a significant impact on protein structure or function. Among the 6 candidate variations, four of them (in *COL1A1*) locate in the triple helical region and the other two variations (in *COL1A2*) locate in the collagen region.

A collagen triple helix is formed by three chains (two $\alpha 1(I)$ chains and one $\alpha 2(I)$ chain) supercoiling around the common axis and glycine, framing almost 338 Gly-Xaa-Yaa repeats in the region, is the only residue small enough to be accommodated in the limited interior of the helical space (Ramachandran and Kartha, 1955; Rich and Crick, 1961; Brodsky and Persikov, 2005). In the collagen triple helix, the Gly-substitution missense will produce structural deformation of the triple helix, leading to destabilization of the helical structure, affecting the synthesis of collagen (**Figure 4**) (Brodsky and Persikov, 2005; Qiu et al., 2018).

To validate the pathogenicity of the candidate variations in *COL1A1*, we checked the specificity of their locations (positions of the four candidate mutations: 1094 and 1097). Evidence from the protein families database (Pfam) (El-Gebali et al., 2019) demonstrate that the locations of all four variations belong to the collagen triple helix region (PF01391: Collagen triple helix repeat (1079–1137)). Structurally, different abnormalities in the collagen helix are associated with the identity of the residue replacing Gly (Bryan et al., 2011; Qiu et al., 2018), which also influence the severity of OI patients (residues replacing Gly of four candidate mutations: Asp, Arg, and Ser). Through the statistical analysis on the location of Glysubstitution mutations in a large number of OI patients, Beck et al. found that all Gly→Asp in the $\alpha 1(I)$ chain led to OI type II (perinatal lethal form) (Beck et al., 2000). In addition, the study of the impact of various Gly replacements discovered that the three replaced form (Gly→Arg, Gly→Ser, and Gly→Cys) had a stronger association with OI lethality than the other replaced forms (Beck et al., 2000). In all, these conclusions indicate that the four candidate mutations of *COL1A1* we identified are highly likely to cause lethal OI phenotypes.

For the two candidate variations in *COL1A2*, we found that the two mutations locate in a special region which is enriched with lethal mutations. Previous study reported that OI-related lethal mutations normally accumulate in eight regularly spaced clusters along the chain $\alpha 2(I)$ (Marini et al., 2007). Truly, both c.G2332T (p.G778C) and c.G2341T (p.G781C) locate in the lethal space 6 (S6) (**Figure 2C**) which belongs to the binding region of proteoglycans (keratan sulfate and heparan sulfate proteoglycan) and type I collagen fibril, the deformation of this region affects the interaction between type I collagen and other components in extracellular matrix (San Antonio et al., 1994; Schaefer, 2014). Abnormalities in this region can lead to abnormal binding of collagen to the matrix, which will affect some of the corresponding biological functions, such as the signal interruption or the distortion of the three-dimensional framework of tissue and organ molecules (Bella and Hulmes, 2017). These biological function disorders then affect bone



development and differentiation, leading to bone abnormalities (Schaefer, 2014). Therefore, we conclude that these two variations possibly generate a severe effect on people carrying the mutations, and are most likely related to OI. In summary, those evidences all support the pathogenesis of these six candidate mutations to OI.

Among the three identified potential new risk factors to OI, *ADAMTS2* and *COL5A2* seem to be more associated with OI than *COL8A1*. In fact, *ADAMTS2* is highly expressed in the skin, bones, tendons, and aorta and has a strong correlation with type I collagen (Bekhouche and Colige, 2015). When procollagen is transferred to the extracellular space, the amino-terminal-propeptide (N-terminal propeptide) and carboxy-terminal propeptide (C-terminal propeptide) at both ends of procollagen would be proteolytically removed to form collagen fibrils (Nijhuis et al., 2019). Next, a large number of collagen fibrils are aggregated into collagen fibers, and finally assembled into the structural framework of cells and tissues and bone (Figure 4A) (Kadler et al., 1996; Van Dijk and Sillence, 2014; Bekhouche and Colige, 2015; Marini et al., 2017). *ADAMTS2* encodes procollagen I N-proteinase, which is used to excise the N-terminal propeptides in procollagen. Abnormal expression of the *ADAMTS2* gene will lead to accumulation of pN-procollagen (collagen molecules with an N-terminal propeptide sequence that has not been cleaved), which will result in the polymerization of the abnormal collagen fibers (Nusgens et al., 1992; Colige et al., 1997; De Coster et al., 2007). According to the phenotypic level,

it has been reported that mutations in *ADAMTS2* caused type I collagen disorders, resulting in DI (one feature of OI) (De Coster et al., 2007). Evidences from Mouse Genome Informatics (MGI) database (Eppig, 2017; Law and Shaw, 2018) also demonstrate that *ADAMTS2* mutated mice show thin skin, triangular face, abnormal cutaneous collagen fibril morphology, abnormal hair follicle morphology, and other phenotypes. Similarly, we noticed that two phenotypes “thin skin” and “triangular face” also exist in the human phenotypes of OI, which further supports the conclusion that *ADAMTS2* gene is most likely to be associated with OI (Marini et al., 2017).

COL5A2, which encodes type V collagen, plays an important role in tissue-specific matrix assembly and the regulation of fibrillogenesis (Wenstrup et al., 2004). From a molecular point of view, most collagen fibril consist of a large number of type I collagen and a very small number (~2%) of type V collagen. Although type V collagen only occupies small portion of collagens in all tissues, it is crucial for collagen fibril nucleation, regarded as collagen fibrillogenesis (Connizzo et al., 2015; Makuszewska et al., 2019). Mouse model results indicate that the deficiency or abnormality of type V collagen leads to a lack of fibril formation in mouse embryonic tissue and even death in the stage of early embryos (Wenstrup et al., 2004; Connizzo et al., 2015), which is similar to the phenotypes in human OI type II (perinatal lethal type). Evidences in MGI shows that mice with *COL5A2* mutation present abnormal cardiovascular system physiology, abnormal skeleton development, abnormal cutaneous collagen

fibril morphology, abnormal cornea morphology, embryonic lethality during organogenesis, neonatal lethality, respiratory distress, thin dermal layer, and other phenotypes (Chen et al., 2017), most of these phenotypes are very similar to the human phenotypes of OI, therefore, we suggest that the *COL5A2* gene is tightly correlated with OI, especially OI type II.

Type VIII collagen, co-encoded by *COL8A1* and *COL8A2*, is a major component of the Descemet's membrane of corneal endothelial cells and is also expressed in other tissues, such as the cornea, sclera, blood vessels, heart, kidneys, and lungs (Hopfer et al., 2005). One clinical feature of OI is the blue sclera, OI also presents thin corneal thickness, smaller corneal diameter, retinal detachment, corneal opacities, myopia, smaller globe length, primary open-angle glaucoma, and other eye abnormalities (Wallace et al., 2014; Hald et al., 2018; Lagrou et al., 2018). Dimasi et al. performed an eye measurement on 28 OI type I patients and found that their mean central corneal thickness (CCT) was lower than that of people in the normal population (Dimasi et al., 2010), indicating that there is a correlation between OI type I and low CCT. Previous studies revealed that mutations in type VIII collagen could lead to lower CCT and thinner Descemet's membrane in the Caucasian and Asian populations (Desronvil et al., 2010). In addition, based on the evidence from MGI, *COL8A1* mutated mice show decreased cornea thickness, abnormal Descemet membrane and other eye abnormalities, we suppose that *COL8A1* is associated with low CCT. In conclusion, *COL8A1* might be associated with eye abnormalities and could also be related to OI.

Taken together, we have explored the association between genotypes and phenotypes in OI with the collected cases from literature. We also have systematically analyzed the impact of each predicted variants in pathogenic genes and identified the potential risk genes for OI, which provide new insights into the underlying mechanism of OI disease. However, our method also has certain limitations, one of them is that many cases do not have their clinical phenotypes fully recorded, which resulted in the insignificant association between certain genotypes and phenotypes. Meanwhile, the phenotypes described in different patients could also be inconsistent. In addition, ANNOVAR

is only focused on coding region, which make the non-coding region unavailable for pathogenic analysis. Nevertheless, our research provides an alternative way to study a new mechanism for rare diseases (Jia and Shi, 2017). Our findings should shed light on the better understanding of OI disease and its effective disease diagnosis.

AUTHOR CONTRIBUTIONS

In this study, TS, YG, and XN designed the study. JS, MR, JJ and MT conducted the data collection and data analysis. JS, MR, JJ, TS, YG, and XN interpreted data in context of BSCL biology. JS, MR, and JJ drafted the manuscript. TS, XN, and YG revised and finalized the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the China Human Proteome Project (Grant No.2014DFB30010, 2014DFB30030), National Natural Science Foundation of China (31671377, 81472369, and 81502144), Clinical Application Research Funds of Capital Beijing (Z171100001017051), Beihang University & Capital Medical University Advanced Innovation Center for Big Data-Based Precision Medicine Plan (BHME- 201804) and Shanghai 111 Project (B14019).

ACKNOWLEDGMENTS

We thank the supercomputer center of East China Normal University for their support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2019.01200/full#supplementary-material>

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7 (4), 248–249. doi: 10.1038/nmeth0410-248
- Beck, K., Chan, V. C., Shenoy, N., Kirkpatrick, A., Ramshaw, J. A., and Brodsky, B. (2000). Destabilization of osteogenesis imperfecta collagen-like model peptides correlates with the identity of the residue replacing glycine. *Proc. Natl. Acad. Sci. U.S.A.* 97 (8), 4273–4278. doi: 10.1073/pnas.070050097
- Bekhouche, M., and Colige, A. (2015). The procollagen N-proteinases ADAMTS2, 3 and 14 in pathophysiology. *Matrix Biol.* 44–46, 46–53. doi: 10.1016/j.matbio.2015.04.001
- Bella, J., and Hulmes, D. J. (2017). Fibrillar Collagens. *Subcell Biochem.* 82, 457–490. doi: 10.1007/978-3-319-49674-0_14
- Bregou Bourgeois, A., Aubry-Rozier, B., Bonafe, L., Laurent-Applegate, L., Pioletti, D. P., and Zambelli, P. Y. (2016). Osteogenesis imperfecta: from diagnosis and multidisciplinary treatment to future perspectives. *Swiss Med. Wkly.* 146, w14322. doi: 10.4414/SMW.2016.14322
- Brodsky, B., and Persikov, A. V. (2005). Molecular structure of the collagen triple helix. *Adv. Protein Chem.* 70, 301–339. doi: 10.1016/S0065-3233(05)70009-7
- Bryan, M. A., Cheng, H., and Brodsky, B. (2011). Sequence environment of mutation affects stability and folding in collagen model peptides of osteogenesis imperfecta. *Biopolymers* 96 (1), 4–13. doi: 10.1002/bip.21432
- Chen, G., Chen, J., Yang, J., Chen, L., Qu, X., Shi, C., et al. (2017). Significant variations in alternative splicing patterns and expression profiles between human-mouse orthologs in early embryos. *Sci. China Life Sci.* 60 (2), 178–188. doi: 10.1007/s11427-015-0348-5
- Colige, A., Li, S. W., Sieron, A. L., Nusgens, B. V., Prockop, D. J., and Lapiere, C. M. (1997). cDNA cloning and expression of bovine procollagen I N-proteinase: a new member of the superfamily of zinc-metalloproteinases with binding sites for cells and other matrix components. *Proc. Natl. Acad. Sci. U.S.A.* 94 (6), 2374–2379. doi: 10.1073/pnas.94.6.2374
- Connizzo, B. K., Freedman, B. R., Fried, J. H., Sun, M., Birk, D. E., and Soslowsky, L. J. (2015). Regulatory role of collagen V in establishing mechanical properties of tendons and ligaments is tissue dependent. *J. Orthop. Res.* 33 (6), 882–888. doi: 10.1002/jor.22893

- Consortium, G. T. (2013). The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* 45 (6), 580–585. doi: 10.1038/ng.2653
- Cooper, G. M., Stone, E. A., Asimenos, G., Program, N. C. S., Green, E. D., Batzoglu, S., et al. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* 15 (7), 901–913. doi: 10.1101/gr.3577405
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M., et al. (2019). Ensembl 2019. *Nucleic Acids Res.* 47(D1), D745–D751. doi: 10.1093/nar/gky1113
- De Coster, P. J., Cornelissen, M., De Paepe, A., Martens, L. C., and Vral, A. (2007). Abnormal dentin structure in two novel gene mutations [COL1A1, Arg134Cys] and [ADAMTS2, Trp795-to-ter] causing rare type I collagen disorders. *Arch. Oral. Biol.* 52 (2), 101–109. doi: 10.1016/j.archoralbio.2006.08.007
- Desronvil, T., Logan-Wyatt, D., Abdrabou, W., Triana, M., Jones, R., Taheri, S., et al. (2010). Distribution of COL8A2 and COL8A1 gene variants in Caucasian primary open angle glaucoma patients with thin central corneal thickness. *Mol. Vis.* 16, 2185–2191.
- Dimasi, D. P., Chen, J. Y., Hewitt, A. W., Klebe, S., Davey, R., Stirling, J., et al. (2010). Novel quantitative trait loci for central corneal thickness identified by candidate gene analysis of osteogenesis imperfecta genes. *Hum. Genet.* 127 (1), 33–44. doi: 10.1007/s00439-009-0729-3
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47 (D1), D427–D432. doi: 10.1093/nar/gky995
- Engelbert, R. H., Gerver, W. J., Breslau-Siderius, L. J., van der Graaf, Y., Pruijs, H. E., van Doorne, J. M., et al. (1998). Spinal complications in osteogenesis imperfecta: 47 patients 1–16 years of age. *Acta Orthop. Scand.* 69 (3), 283–286. doi: 10.3109/17453679809000931
- Eppig, J. T. (2017). Mouse Genome Informatics (MGI) resource: genetic, genomic, and biological knowledgebase for the laboratory mouse. *ILAR J.* 58 (1), 17–41. doi: 10.1093/ilar/ilx013
- Fokkema, I. F., Taschner, P. E., Schaafsma, G. C., Celli, J., Laros, J. F., and den Dunnen, J. T. (2011). LOVD v2.0: the next generation in gene variant databases. *Hum. Mutat.* 32 (5), 557–563. doi: 10.1002/humu.21438
- Forlino, A., and Marini, J. C. (2016). Osteogenesis imperfecta. *Lancet* 387 (10028), 1657–1671. doi: 10.1016/S0140-6736(15)00728-X
- Foster, B. L., Ramnitz, M. S., Gafni, R. I., Burke, A. B., Boyce, A. M., Lee, J. S., et al. (2014). Rare bone diseases and their dental, oral, and craniofacial manifestations. *J. Dent. Res.* 93 (7 Suppl), 7S–19S. doi: 10.1177/0022034514529150
- Geng, J., Wang, H., Liu, Y., Tai, J., Jin, Y., Zhang, J., et al. (2017). Correlation between BRAF (V600E) mutation and clinicopathological features in pediatric papillary thyroid carcinoma. *Sci. China Life Sci.* 60 (7), 729–738. doi: 10.1007/s11427-017-9083-8
- Hald, J. D., Folkestad, L., Swan, C. Z., Wanscher, J., Schmidt, M., Gjørup, H., et al. (2018). Osteogenesis imperfecta and the teeth, eyes, and ears—a study of non-skeletal phenotypes in adults. *Osteoporos. Int.* 29 (12), 2781–2789. doi: 10.1007/s00198-018-4663-x
- Hopfer, U., Fukai, N., Hopfer, H., Wolf, G., Joyce, N., Li, E., et al. (2005). Targeted disruption of Col8a1 and Col8a2 genes in mice leads to anterior segment abnormalities in the eye. *FASEB J.* 19 (10), 1232–1244. doi: 10.1096/fj.04-3019com
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60(7), 686–691. doi: 10.1007/s11427-017-9099-3
- Jia, J., An, Z., Ming, Y., Guo, Y., Li, W., Li, X., et al. (2018). PedAM: a database for Pediatric disease Annotation and Medicine. *Nucleic Acids Res.* 46 (D1), D977–DD83. doi: 10.1093/nar/gkx1049
- Jia, J. M., An, Z. X., Ming, Y., Guo, Y. L., Li, W., Liang, Y. X., et al. (2018). eRAM: encyclopedia of rare disease annotations for precision medicine. *Nucleic Acids Res.* 46 (D1), D937–DD43. doi: 10.1093/nar/gkx1062
- Kadler, K. E., Holmes, D. F., Trotter, J. A., and Chapman, J. A. (1996). Collagen fibril formation. *Biochem. J.* 316 (Pt 1), 1–11. doi: 10.1042/bj3160001
- Kircher, M., Witten, D. M., Jain, P., O’Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46 (3), 310–315. doi: 10.1038/ng.2892
- Lagrou, L. M., Gilbert, J., Hannibal, M., Caird, M. S., Thomas, I., Moroi, S. E., et al. (2018). Altered corneal biomechanical properties in children with osteogenesis imperfecta. *J. AAPOS* 22 (3), 183–7 e1. doi: 10.1016/j.jaapos.2017.12.015
- Law, M., and Shaw, D. R. (2018). Mouse Genome Informatics (MGI) is the international resource for information on the laboratory mouse. *Methods Mol. Biol.* 1757, 141–161. doi: 10.1007/978-1-4939-7737-6_7
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536 (7616), 285–291. doi: 10.1038/nature19057
- Li, C., Zhang, J., Li, S., Han, T., Kuang, W., Zhou, Y., et al. (2017). Gene mutations and clinical phenotypes in Chinese children with Blau syndrome. *Sci. China Life Sci.* 60 (7), 758–762. doi: 10.1007/s11427-017-9090-6
- Li, T., Wernersson, R., Hansen, R. B., Horn, H., Mercer, J., Slodkowitz, G., et al. (2017). A scored human protein–protein interaction network to catalyze genomic interpretation. *Nat. Methods* 14 (1), 61–64. doi: 10.1038/nmeth.4083
- Lin, H. Y., Chuang, C. K., Su, Y. N., Chen, M. R., Chiu, H. C., Niu, D. M., et al. (2015). Genotype and phenotype analysis of Taiwanese patients with osteogenesis imperfecta. *Orphanet. J. Rare Dis.* 10, 152. doi: 10.1186/s13023-015-0370-2
- Lindahl, K., Astrom, E., Rubin, C. J., Grigelioniene, G., Malmgren, B., Ljunggren, O., et al. (2015). Genetic epidemiology, prevalence, and genotype–phenotype correlations in the Swedish population with osteogenesis imperfecta. *Eur. J. Hum. Genet.* 23 (8), 1042–1050. doi: 10.1038/ejhg.2015.81
- Lindert, U., Cabral, W. A., Ausavarat, S., Tongkobetch, S., Ludin, K., Barnes, A. M., et al. (2016). MBTPS2 mutations cause defective regulated intramembrane proteolysis in X-linked osteogenesis imperfecta. *Nat. Commun.* 7 (1), 11920. doi: 10.1038/ncomms11920
- Liu, S., Huang, S., Chen, F., Zhao, L., Yuan, Y., Francis, S. S., et al. (2018). Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese Population History. *Cell* 175 (2), 347–59 e14. doi: 10.1016/j.cell.2018.08.016
- Lu, Y., Zhang, S., Wang, Y., Ren, X., and Han, J. (2019). Molecular mechanisms and clinical manifestations of rare genetic disorders associated with type I collagen. *Intractable Rare Dis. Res.* 8 (2), 98–107. doi: 10.5582/irdr.2019.01064
- Makuszewska, M., Bonda, T., Cieslinska, M., Bialuk, I., Winnicka, M. M., Skotnicka, B., et al. (2019). Expression of collagens type I and V in healing rat’s tympanic membrane. *Int. J. Pediatr. Otorhinolaryngol.* 118, 79–83. doi: 10.1016/j.ijporl.2018.12.020
- Marini, J. C., Forlino, A., Cabral, W. A., Barnes, A. M., San Antonio, J. D., Milgrom, S., et al. (2007). Consortium for osteogenesis imperfecta mutations in the helical domain of type I collagen: regions rich in lethal mutations align with collagen binding sites for integrins and proteoglycans. *Hum. Mutat.* 28 (3), 209–221. doi: 10.1002/humu.20429
- Marini, J. C., Forlino, A., Bachinger, H. P., Bishop, N. J., Byers, P. H., Paepe, A., et al. (2017). Osteogenesis imperfecta. *Nat. Rev. Dis. Primers* 3, 17052. doi: 10.1038/nrdp.2017.52
- Marom, R., Lee, Y. C., Grafe, I., and Lee, B. (2016). Pharmacological and biological therapeutic strategies for osteogenesis imperfecta. *Am. J. Med. Genet. C Semin. Med. Genet.* 172 (4), 367–383. doi: 10.1002/ajmg.c.31532
- Mrosk, J., Bhavani, G. S., Shah, H., Hecht, J., Kruger, U., Shukla, A., et al. (2018). Diagnostic strategies and genotype–phenotype correlation in a large Indian cohort of osteogenesis imperfecta. *Bone* 110, 368–377. doi: 10.1016/j.bone.2018.02.029
- Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res.* 11 (5), 863–874. doi: 10.1101/gr.176601
- Nijhuis, W. H., Eastwood, D. M., Allgrove, J., Hvid, I., Weinans, H. H., Bank, R. A., et al. (2019). Current concepts in osteogenesis imperfecta: bone structure, biomechanics and medical management. *J. Child. Orthop.* 13 (1), 1–11. doi: 10.1302/1863-2548.13.180190
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci China Life Sci.* 60 (7), 681–685. doi: 10.1007/s11427-017-9100-1
- Nusgens, B. V., Verellen-Dumoulin, C., Hermanns-Le, T., De Paepe, A., Nuytinck, L., Pierard, G. E., et al. (1992). Evidence for a relationship between Ehlers-Danlos type VII C in humans and bovine dermatosparaxis. *Nat. Genet.* 1 (3), 214–217. doi: 10.1038/ng0692-214
- Papadopoulos, J. S., and Agarwala, R. (2007). COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics* 23 (9), 1073–1079. doi: 10.1093/bioinformatics/btm076
- Qiu, Y., Mekkat, A., Yu, H., Yigit, S., Hamaia, S., Farndale, R. W., et al. (2018). Collagen Gly missense mutations: effect of residue identity on collagen

- structure and integrin binding. *J. Struct. Biol.* 203 (3), 255–262. doi: 10.1016/j.jsb.2018.05.003
- Ramachandran, G. N., and Kartha, G. (1955). Structure of collagen. *Nature* 176 (4482), 593–595. doi: 10.1038/176593a0
- Rauch, F., Lalic, L., Roughley, P., and Glorieux, F. H. (2010). Relationship between genotype and skeletal phenotype in children and adolescents with osteogenesis imperfecta. *J. Bone Miner. Res.* 25 (6), 1367–1374. doi: 10.1359/jbmr.091109
- Rich, A., and Crick, F. H. (1961). The molecular structure of collagen. *J. Mol. Biol.* 3, 483–506. doi: 10.1016/S0022-2836(61)80016-8
- San Antonio, J. D., Lander, A. D., Karnovsky, M. J., and Slayter, H. S. (1994). Mapping the heparin-binding sites on type I collagen monomers and fibrils. *J. Cell Biol.* 125 (5), 1179–1188. doi: 10.1083/jcb.125.5.1179
- Schaefer, L. (2014). Proteoglycans, key regulators of cell-matrix dynamics. *Matrix Biol.* 35, 1–2. doi: 10.1016/j.matbio.2014.05.001
- Sillence, D. O., Rimo, D. L., and Danks, D. M. (1979). Clinical variability in osteogenesis imperfecta-variable expressivity or genetic heterogeneity. *Birth Defects Orig. Artic. Ser.* 15 (5B), 113–129.
- Slenter, D. N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., et al. (2018). WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* 46 (D1), D661–D667. doi: 10.1093/nar/gkx1064
- Van Dijk, F. S., and Sillence, D. O. (2014). Osteogenesis imperfecta: clinical diagnosis, nomenclature and severity assessment. *Am. J. Med. Genet. A* 164A (6), 1470–1481. doi: 10.1002/ajmg.a.36545
- Wallace, D. J., Chau, F. Y., Santiago-Turla, C., Hauser, M., Challa, P., Lee, P. P., et al. (2014). Osteogenesis imperfecta and primary open angle glaucoma: genotypic analysis of a new phenotypic association. *Mol. Vis.* 20, 1174–1181.
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38 (16), e164. doi: 10.1093/nar/gkq603
- Wang, J., Vasaikar, S., Shi, Z., Greer, M., and Zhang, B. (2017). WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 45 (W1), W130–W137. doi: 10.1093/nar/gkx356
- Wenstrup, R. J., Florer, J. B., Brunskill, E. W., Bell, S. M., Chervoneva, I., and Birk, D. E. (2004). Type V collagen controls the initiation of collagen fibril assembly. *J. Biol. Chem.* 279 (51), 53331–53337. doi: 10.1074/jbc.M409622200
- Zhang, Z. L., Zhang, H., Ke, Y. H., Yue, H., Xiao, W. J., Yu, J. B., et al. (2012). The identification of novel mutations in COL1A1, COL1A2, and LEPRE1 genes in Chinese patients with osteogenesis imperfecta. *J. Bone Miner. Metab.* 30 (1), 69–77. doi: 10.1007/s00774-011-0284-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Shi, Ren, Jia, Tang, Guo, Ni and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Transcriptome Analysis Identifies Piwi-Interacting RNAs as Prognostic Markers for Recurrence of Prostate Cancer

Yuanli Zuo¹, Yu Liang², Jiting Zhang¹, Yingyi Hao², Menglong Li², Zhining Wen^{2,3*} and Yun Zhao^{1*}

¹ Key Laboratory of Bio-Resource and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China, ² College of Chemistry, Sichuan University, Chengdu, China, ³ Medical Big Data Center, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:

Zhichao Liu,
National Center for Toxicological
Research (FDA), United States

Reviewed by:

James S. Sutcliffe,
Vanderbilt University, United States
Xiangwen Liu,
University of Arkansas at Little Rock,
United States
Yifan Zhang,
University of Arkansas at Little Rock,
United States

*Correspondence:

Yun Zhao,
zhaoyun@scu.edu.cn
Zhining Wen,
w_zhining@163.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 23 January 2019

Accepted: 24 September 2019

Published: 22 October 2019

Citation:

Zuo Y, Liang Y, Zhang J, Hao Y,
Li M, Wen Z and Zhao Y (2019)
Transcriptome Analysis Identifies
Piwi-Interacting RNAs as Prognostic
Markers for Recurrence
of Prostate Cancer.
Front. Genet. 10:1018.
doi: 10.3389/fgene.2019.01018

Prostate cancer remains the second leading cause of male cancer death, and there is an unmet need for biomarkers to identify patients with such aggressive disease. Piwi-interacting RNAs (piRNAs) have been classified as transcriptional and posttranscriptional regulators in somatic cells. In this study, we discovered three piRNAs as novel prognostic markers and their association with prostate cancer biochemical recurrence was confirmed in validation data set. To obtain a better understanding of piRNA expression patterns in prostate cancer and to find gene coexpression with piRNAs, we performed weighted gene coexpression network analysis. Target genes of three piRNAs have also been predicted based on base complementarity and expression correlativity. Functional analysis revealed the relationships between target genes and prostate cancer. Our work also identified differential expression of piRNAs between Gleason stage 3 + 4 and 4 + 3 prostate cancer. Overall, this study may explain the roles and demonstrate the potential clinical utility of piRNAs in prostate cancer in a way.

Keywords: piRNA, prostate cancer, biomarker, survival analysis, WGCNA

INTRODUCTION

Prostate cancer is one of the most common male malignant tumors and the second leading cause of cancer death in men worldwide (Mei et al., 2013; Damborský et al., 2017). Currently, the most common clinical diagnostic indices of prostate cancer prognosis include age, prostate-specific antigen level, tumor volume, perineural invasion, and Gleason grading (Carlsson and Roobol, 2017). However, using clinical parameters alone is not sufficient for accurate prognosis (Patel and Gnanapragasam, 2016). Thus, biomarkers that can provide more accurate risk stratification and help clinicians to make improved decision at the pretreatment stage are urgently needed (Long et al., 2014).

Noncoding RNAs (ncRNAs) have been gaining recognition for their involvement in genetic and epigenetic regulation (Cech and Steitz, 2014). Recent studies suggest that ncRNAs could be a promising hallmark in human diseases, particularly in cancer (Esteller, 2011). Studies have documented that expression patterns of some ncRNAs such as lncRNA and miRNA had relationships with clinic situation of prostate cancer (Jin et al., 2011; Ru et al., 2012; Zhu et al., 2014).

As important members in ncRNAs family, Piwi-interacting RNAs (piRNAs) are ~26- to 32-nt RNAs whose names derive from their association with the PIWI subfamily of Argonaute proteins. piRNAs are first identified in a genetic screen for mutants affecting asymmetric division of stem cells in the *Drosophila* germline (Aravin et al., 2006; Vagin et al., 2006), and they have also been found expressed in stem and other somatic cells (Juliano et al., 2011). piRNAs are best known for their roles in transposable element repression. But they may additionally regulate gene expression through an miRNA-like base complementary mechanism (Lee et al., 2011; Martinez et al., 2015). Recent studies revealed that expression patterns of piRNAs showed markedly different in human multiple myeloma, breast, lung, gastric, and other cancer tissues compared with their corresponding nontumor tissues (Mei et al., 2013; Li et al., 2015; Moyano and Stefani, 2015; Ng et al., 2016). Hence, more thorough analyses must be conducted before utilizing piRNAs as diagnostic and prognostic markers (Assumpcao et al., 2015; Lim and Kai, 2015).

MATERIALS AND METHODS

RNA-Seq Data Sets and Clinic Data

A total of 106 prostate cancer tissues RNA-seq data generated by the Department of Pathology & Laboratory Medicine, Emory University were obtained from NCBI SRA database (SRP036848). The corresponding clinical information was downloaded from NCBI Gene Expression Omnibus (Series GSE54460) (Long et al., 2014).

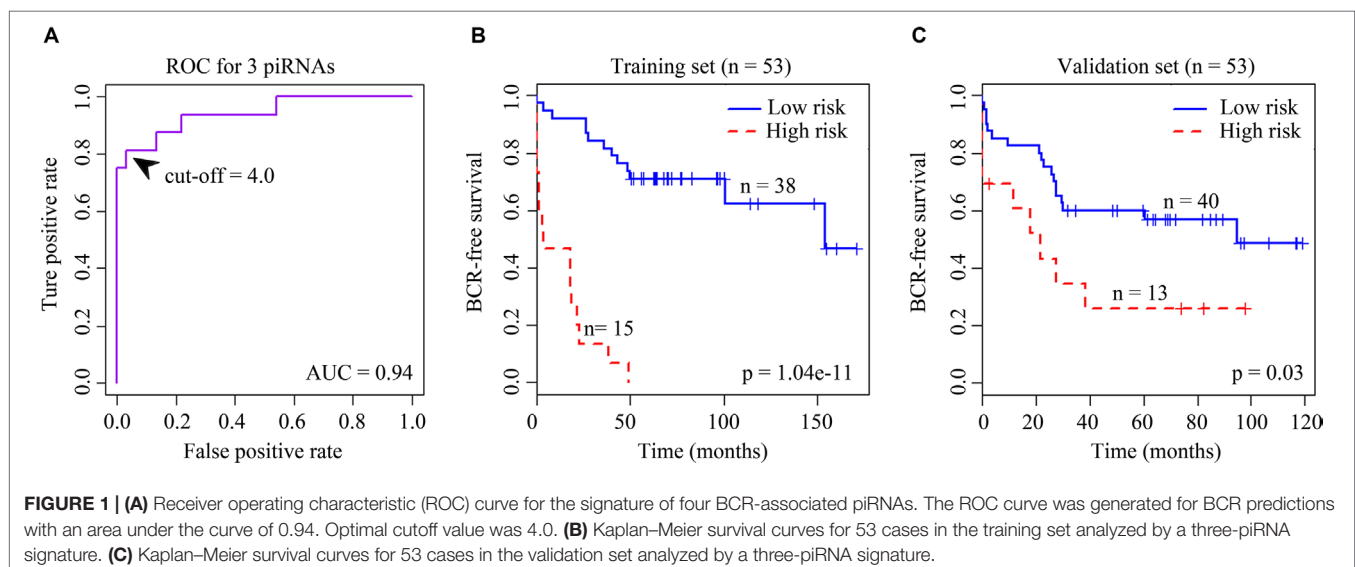
piRNA Expression Analysis

SRR files downloaded from NCBI were converted to FASTQ files using the “fastq-dump” tool in sratoolkit.2.8.1-win64. The unpaired reads were abandoned. Resulting FASTQ files were trimmed using Trimmomatic-0.36 to remove low-quality reads (Bolger et al., 2014). The remaining reads were aligned to human genome hg38 using the Spliced Transcripts Alignment

to a Reference (STAR-2.5.2b) software (Dobin et al., 2013). The piRNA reference transcriptome was generated for annotation and quantitation by using the information from the piRNABank database (<http://pirnabank.ibab.ac.in/>) (Sai Lakshmi and Agrawal, 2008). Expression counts of transcripts were quantitated using HTSeq package. Following TMM normalization, expression values were transformed to count per million mapped reads (CPM) using edgeR (Krishnan et al., 2017); piRNAs with CPM values ≥ 1 in at least 10% of samples were deemed as expressed and taken into further analyses.

Survival Analysis

Clinical information was downloaded from the NCBI GEO database. Patients' biochemical recurrence (BCR) information was used in survival analysis. We first performed univariate Cox regression analysis to identify candidates significantly associated with patient outcome ($p < 0.05$). Next, a robust likelihood-based survival modeling approach was used to select the piRNA signature. We implemented our analysis by using the “rbsurv” package in R (Cho et al., 2009). Then we built a multivariate Cox regression model by the selected piRNAs to find a final set of piRNAs that had a significant association with BCR of prostate cancer ($p < 0.01$). Both the univariate and multivariate Cox analyses were executed using “coxph” function in “survival” package. Significantly associated piRNAs were used to calculate each patient's BCR risk. Briefly, we first multiplied a piRNA's expression value by its corresponding Cox coefficient to obtain an individual piRNA weight. Then we summed all the individual piRNA weights to get the risk score (Firmino et al., 2016; Martinez et al., 2016). And then receiver operating characteristic curve was employed for estimating optimal cutoff points for the outcomes to stratify patients into low- and high-risk groups (Krishnan et al., 2016) (Figure 1). The risk score whose corresponding difference between the true-positive rate and the false-positive rate was the maximum was chosen to be the



optimal cutoff. Kaplan–Meier curves for two distinct groups of patients were plotted using “survfit” function in “survival” package. *P* value from log-rank test was computed using “survdiff” function.

Gene Coexpression Network Analysis

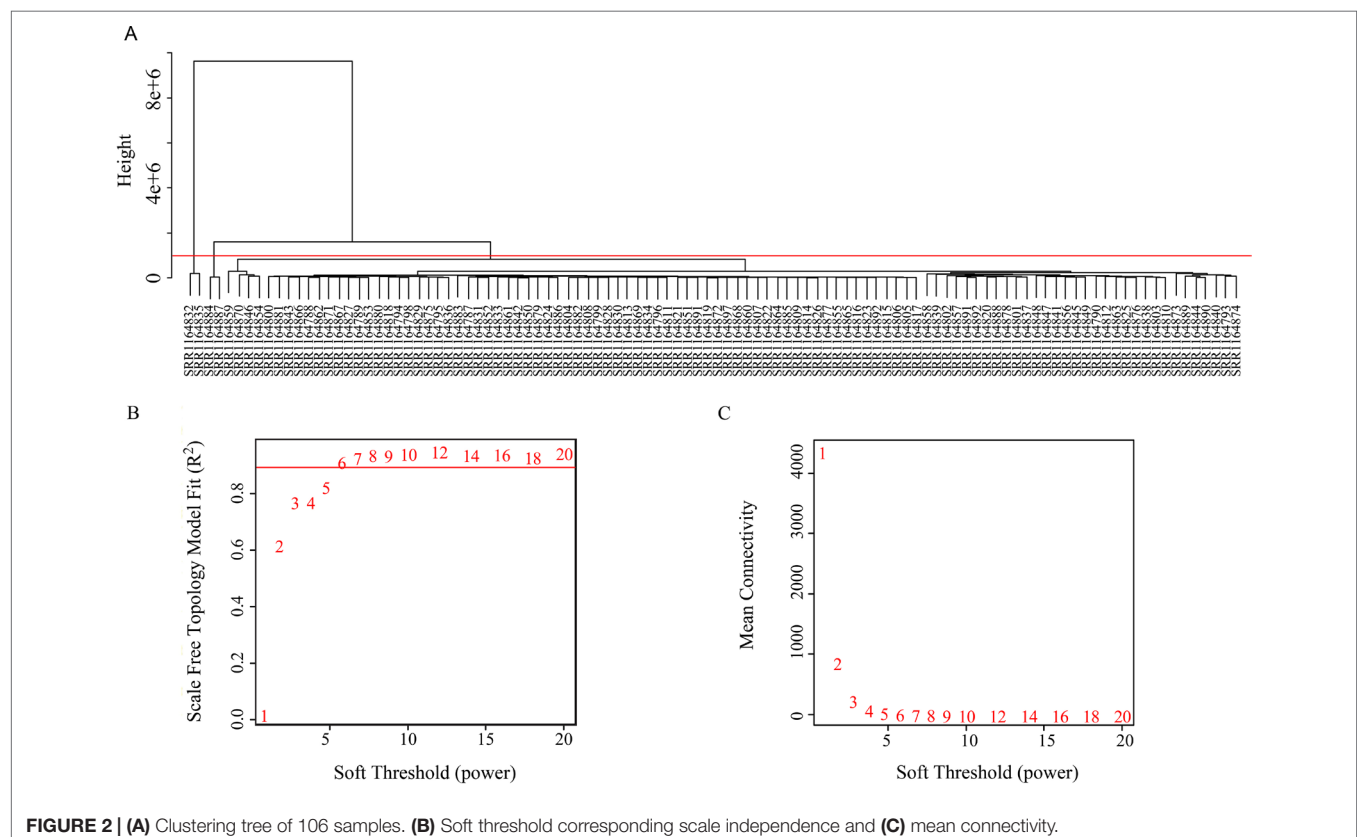
The coexpression analysis was performed using weighted gene coexpression network analysis (WGCNA) method based on the significantly variant genes ($SD \geq 2$) and the three survival-associated piRNAs expression data according to the protocols of WGCNA in an R environment (Langfelder and Horvath, 2008). Outlier samples were detected using hierarchical clustering. Setting the cut-height of 1,000,000, we removed four outliers. The remaining 102 samples were taken into the following analysis process (Figure 2). We then generated an adjacency matrix by calculating the Pearson correlation between all genes. The PickSoftThreshold function of WGCNA was used to choose the appropriate power for the network topology from various soft thresholding powers. The scale-free network was rendered by raising the soft thresholding power (β) to six, resulting in a scale-free topology index (R^2) of 0.9 (Figure 2) and a mean connectivity approximate of zero (Figure 2). The gene coexpression networks were constructed using the blockwiseModules function by a one-step method. Then a topological overlap matrix was calculated using the adjacency matrix, and Interaction networks were constructed for select modules. Cytoscape v 3.5.0 is used for network visualization.

piRNA Target Prediction

Recent evidence has suggested interaction between piRNAs and mRNAs through base-pair complementarity and a possible inverse correlation between piRNA expression and its corresponding mRNA targets (Hashim et al., 2014; Preethi Krishnan, 2016). But in this study, we did not exclude the possibility that piRNA might have interaction with other RNAs rather than only mRNAs. Fasta sequences of all the genes were obtained from GENCODE database (<http://www.genencodegenes.org/releases/26.html>) and fasta sequences of the piRNAs were obtained from piRNABank (hg 38). The targets of selected piRNAs were identified using miRanda against the RNA library of human genome with a mean free energy of maximum 20 kcal/mol and alignment score threshold of 140 (Rajan et al., 2016). The resulting RNAs had been taken intersected with the genes that had a coexpression pattern (topological overlap matrix weight ≥ 0.01) with the piRNAs to obtain the piRNA targets that meet both the requirement of base-pair complementarity (results from MiRanda target predicting) and expression pattern relevance (results from WGCNA coexpression analysis).

Functional Analysis

Since there was no means to annotate the function of piRNAs directly, we turned to analyze the potential functional insights of piRNAs by focusing on their target genes. Gene Ontology (GO) functional module enrichment, Kyoto Encyclopedia



of Genes and Genomes (KEGG) pathway analysis, database searching, and literature consulting were used. GO and KEGG classification of genes targeted by selected piRNA was performed using DAVID functional annotation tool (<https://david.ncifcrf.gov/>) (Huang et al., 2009). Redundancy in mRNA was removed before analysis. GO terms and KEGG pathways with $p < 0.05$ were taken into consideration to summarize the enriched functions.

Differential Expression Analysis

Fifty-six Gleason 3 + 4 and 24 Gleason 4 + 3 samples were taken into differential expression analysis. Pairwise comparisons were applied to identify significantly differentially expressed piRNA between the same Gleason stage 3 + 4 and stage 4 + 3 patient cohorts. Differential expression of expressed piRNAs was calculated using DESeq2 version 1.4.1 available in Bioconductor version 2.8. DESeq2 uses a negative binomial distribution model to test for differential expression in deep sequencing data sets. The piRNAs with the absolute value of fold change > 1.5 and adjusted p value with false discovery rate < 0.05 were considered significant.

RESULTS

piRNAs Are Associated With Prostate Cancer BCR

We developed a custom analysis pipeline to detect expression patterns of piRNAs in prostate cancer patients from high-throughput sequencing data. There were 7,630 piRNAs expressed with at least 1 CPM in 10% of the samples. Given the limitation of samples, a Holdout method cross-validation was applied to reduce the effect of data variability and avoid overfitting. Briefly, we randomly chose 53 samples as training set and used the rest as validation set.

We first selected an initial set of piRNAs by performing univariate survival analysis using Cox proportional hazards regression model. With the threshold of $p < 0.05$, a total of 808 piRNAs associated with the BCR were initially identified. Next, we screened the optimal survival-associated signature piRNAs based on a robust likelihood-based survival model. Six piRNAs were selected as signature piRNAs that can optimally predict the BCR of patients with prostate cancer. By fitting a multivariate Cox proportional hazards regression model, we finally get a prediction panel that comprised three piRNAs (Table 1). The risk scores weighting the BCR of prostate cancer were constructed using the three piRNAs. And then we used a receiver operating characteristic based estimation to get an optimal cutoff score and dichotomized the patients into two groups: low risk (risk score < 4.0) and high risk (risk score ≥ 4.0 ; Figure 1, see *Materials and Methods*). Thirty-eight patients (71.7%) were categorized to the

high-risk group, whereas 15 (28.3%) were categorized to the low-risk group. The Kaplan–Meier plot of piRNA risk scores shows that it can distinguish the patients with high risk of BCR from the low-risk patients (log-rank $p = 1.04 \times 10^{-11}$, Figure 1).

Validation of Three-piRNA Signature

To evaluate the robustness and effectiveness of the three piRNAs signature, we used the rest of the 53 samples as validation set. The BCR risk score of each patient was calculated based on expression values of three piRNAs signature. We further calculated the risk score of each sample and divided the patients into two risk groups based on the Cox coefficients and optimal cutoff risk scores obtained from the training data set.

For the validation data set, 40 (75.5%) and 13 (24.5%) patients were distinguished as the low- and high-risk groups, respectively. Kaplan–Meier plots indicated significant differences between BCRs of the two groups in the validation data set (log-rank $p = 0.03$, Figure 1). Similar to the results obtained in the training data set, the risk score showed promising prognostic power of prostate cancer BCR.

Coexpression Gene Analysis

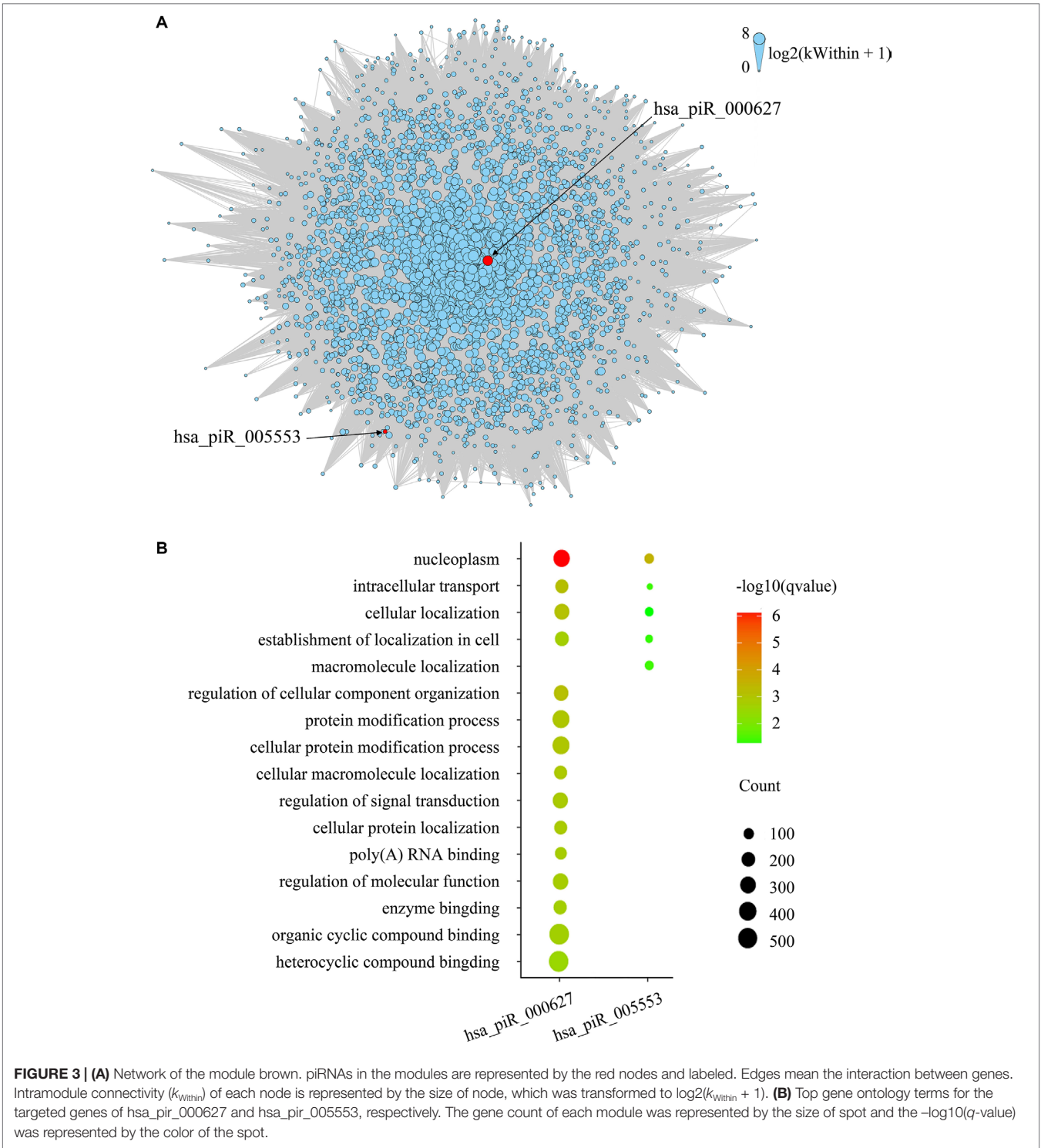
To evaluate gene expression from a network perspective and gain further insight into the mechanisms by which piRNA changes might influence gene expression, we performed WGCNA to build a gene coexpression network based on the three BCR-associated piRNAs and 37,316 genes whose expression values varied among all the samples (Langfelder and Horvath, 2008). A total of 127 modules were recognized, and two included piRNAs. Module brown consisted of hsa_piR_000627, hsa_piR_005553, and 2,721 other genes (Supplementary Table 1), and hsa_piR_019346 had been included in the module lightcyan1 with the other 111 genes (Supplementary Table 2). The gene expression networks of the two modules were visualized in Cytoscape. As we can see, hsa_piR_000627 is near the centric position of the network, and hsa_piR_005553 is on the periphery (Figure 3). Consistently, the intramodule connectivity (k_{within}) of hsa_piR_000627 is 47.46, and hsa_piR_005553 is only 1.71 (median k_{within} value of the whole module is 7.89). It indicates that hsa_piR_000627 might be a hub of this network. It has more coexpression genes and stronger interaction with these genes.

piRNAs Target Gene Prediction

Recent evidence suggests that piRNAs, in a mechanism similar to miRNAs, may regulate gene expression through base pair complementarity with their targets. However, few studies have identified the corresponding gene targets of specific piRNAs (Hashim et al., 2014; Chu et al., 2015). For this study, we only considered significantly prognosis-related piRNAs (three nonredundant piRNAs in total from BCR) and focused on the correlations between piRNA and its targets. Using MiRanda algorithm v3.3b and applying the cutoffs, we identified nonredundant gene targets of each piRNAs. In order to get targets with more authenticity, we took the intersection of results from MiRanda and the coexpression genes of the three piRNAs. The results are shown in Table 2 (target genes were listed in Supplementary Tables 3 and 4). Intriguingly, we found that 343 target genes (92.45%) of hsa_piR_005553 were

TABLE 1 | Three piRNAs significantly associated with BCR of prostate cancer patients.

	Cox coefficient	HR	95% CI	p
hsa_piR_005553	0.05357	1.055	1.0305–1.0802	8.14×10^{-6}
hsa_piR_019346	0.03015	1.0306	1.0098–1.0519	0.00381
hsa_piR_000627	0.01083	1.0109	1.0032–1.0186	0.00525



also targets of has_pir_000627. This consists with the results that these two piRNAs are in the same gene module and implied that there might be some interaction between has_pir_000627 and has_pir_005553. GO enrichment also was used for functional analysis of hsa_pir_000627 and hsa_pir_005553 targeting genes (Figure 3). As we can see, the GO modules of both hsa_pir_000627

TABLE 2 | Number of targeted genes of three BCR-associated piRNAs.

piRNA	Targeted gene	Coding gene	Noncoding gene
hsa_pir_000627	1,869	1,556	313
hsa_pir_005553	371	321	50
hsa_pir_019346	1	1	0

and hsa_pir_007316 had a very high similarity. For instance, both their first modules are nucleoplasm. This might imply their close correlation in biological function furthermore.

Since only one target gene of hsa_pir_019346 was found, we analyzed its function through literature consulting and database searching instead of GO enrichment. The protein coded by the target gene *PNPLA7* (patatin-like phospholipase domain containing 7) is a member of human patatin-like phospholipase domain containing proteins family, which worked as an insulin-regulated lysophospholipase (Kienesberger et al., 2009). Human *PNPLA7* is predominantly expressed in prostate and pancreas; it is involved in regulation of adipocyte differentiation and induced by metabolic stimuli (Wilson et al., 2006). Its related pathways are metabolism and glycerophospholipid biosynthesis. GO annotations related to this gene include lysophospholipase activity and hydrolase activity. Recent work revealed that its gene polymorphism correlated with menstrual disorder. But no work suggests that it is directly associated with human tumor so far.

Differential Expression of piRNAs Between Gleason Stage 3 + 4 and 4 + 3

Gleason score is known to be a powerful metric that can be used to stratify prostate cancer patients into different risk categories. The grading system for prostate cancer is unique in that the final pathologic grade is a Gleason sum of the primary Gleason patterns and the secondary pattern. It has been suggested that primary Gleason 4 pattern and Gleason 3 pattern tumors represent different disease states (Chan et al., 2000; Lavery and Droller, 2012), and several studies suggested that different primary Gleason patterns of patients with a Gleason score of 7 will result in different clinical outcomes (Herman et al., 2001; Berg et al., 2014).

To investigate the differences in gene expression, we performed DESeq2 differential expression analysis to explore piRNAs differentially expressed between 56 samples with Gleason 3 + 4 (primary pattern 3) and 24 samples with Gleason 4 + 3 (primary pattern 4). When setting the thresholds that the absolute value of fold

change is >1.5 and adjusted $p < 0.05$, we identified four differentially expressed piRNAs (Table 3, Figure 4A). Interestingly, all the four piRNAs were up-regulated in Gleason 4 + 3 compared with 3 + 4 cases. And we also found that three out of four differentially expressed piRNAs have very close locations in q21.1 of chromosome 2 (Figure 4B). This might imply that they came from the same piRNA cluster.

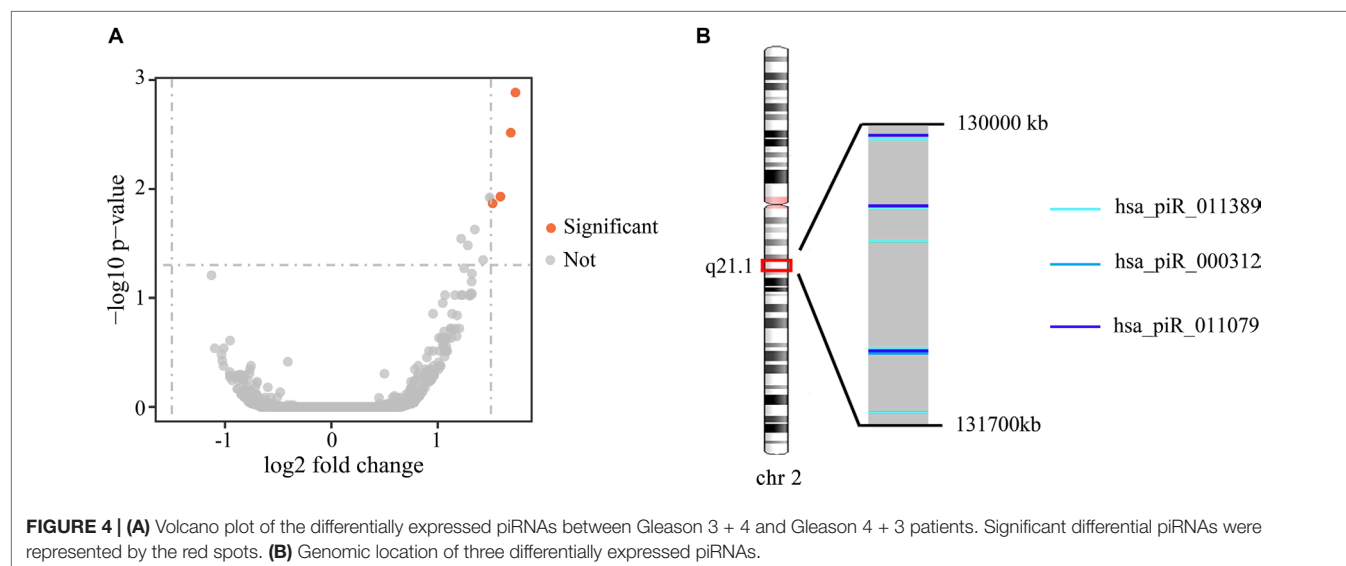
DISCUSSION

Prostate cancer remains the most common male cancer, and with over 28,000 deaths per year, it ranks second among tumor mortality (Long et al., 2011; Long et al., 2014). Nowadays, measures for diagnosis and prognosis of prostate cancer have improved a lot. However, major challenges about improvement of prognosis accuracy remain. The inconsistency between results of different methods usually made the physicians and patients disoriented. As a class of important small ncRNA, piRNA gained a growing concern. More and more studies focused on their correlation with human diseases, especially cancer.

In this study, we assessed piRNA expression in 106 prostate cancer samples from NCBI database using a custom piRNA analysis pipeline. Our findings revealed that piRNAs were expressed in human prostate cancer tissues. In addition, we identified three piRNAs (hsa_pir_000627, hsa_pir_005553, hsa_pir_019346) associated with prostate cancer BCR. We also successfully validated the piRNAs' prognostic significance through cross-validation. Using

TABLE 3 | Four piRNAs differentially expressed between stage 3 + 4 and stage 4 + 3 patients.

	log2FC	p	p _{adj}
hsa_pir_011389	1.730631697	2.70E-07	0.00130497
hsa_pir_000312	1.686705408	1.26E-06	0.003051624
hsa_pir_011079	1.589715335	7.30E-06	0.01174993
hsa_pir_012366	1.516241812	1.69E-05	0.01356796



WGCNA, we constructed the piRNA-correlated gene networks. The results indicated that hsa_pir_000627 and hsa_pir_005553 were in a same network module and had a close relation. Gene targets of three candidate piRNAs have also been identified. We found that hsa_pir_000627 and hsa_pir_005553 had 343 cotargeting genes, and they account for 92.45% of targets of hsa_pir_005553. Functional analysis indicated that both their target genes were mainly associated with nucleoplasm and intracellular transport. The little number of target genes of hsa_pir_019346 might explain why it had been assigned to a small module. Since its target gene *PNPLA7* is insufficiently studied so far, the biological function and association with human prostate cancer of hsa_pir_019346 needs a further investigation. Moreover, we found four piRNAs differentially expressed between Gleason stage 3 + 4 and 4 + 3 patients. This might be a helpful information to solve the puzzle of accurately distinguishing these two groups.

CONCLUSIONS

In conclusion, our data revealed that three candidate piRNAs, namely, hsa_pir_000627, hsa_pir_005553 and hsa_pir_019346, had significant correlation with BCR of prostate cancer and can be potential prognostic biomarkers. The comparison of Gleason stage 3 + 4 and 4 + 3 cases identified four differentially expressed piRNAs. This shows the utility of piRNAs in clinical classification. In a word, our study shows that piRNAs had potential to be prognosis biomarkers of prostate cancer.

REFERENCES

- Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., et al. (2006). A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* 442, 203–207. doi: 10.1038/nature04916
- Assumpcao, C. B., Calcagno, D. Q., Araujo, T. M., Santos, S. E., Santos, A. K., Riggins, G. J., et al. (2015). The role of piRNA and its potential clinical implications in cancer. *Epigenomics* 7, 975–984. doi: 10.2217/epi.15.37
- Berg, K. D., Röder, M. A., Brasso, K., Vainer, B., and Iversen, P. (2014). Primary Gleason pattern in biopsy Gleason score 7 is predictive of adverse histopathological features and biochemical failure following radical prostatectomy. *Scand. J. Urol.* 48, 168–176. doi: 10.3109/21681805.2013.821628
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Carlsson, S. V., and Roobol, M. J. (2017). Improving the evaluation and diagnosis of clinically significant prostate cancer in 2017. *Curr. Opin. Urol.* 27, 198–204. doi: 10.1097/MOU.0000000000000382
- Cech, T. R., and Steitz, J. A. (2014). The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* 157, 77–94. doi: 10.1016/j.cell.2014.03.008
- Chan, T. Y., Partin, A. W., Walsh, P. C., and Epstein, J. I. (2000). Prognostic significance of Gleason score 3 + 4 versus Gleason score 4 + 3 tumor at radical prostatectomy. *Urology* 56, 823–827. doi: 10.1016/S0090-4295(00)00753-6
- Cho, H., Yu, A., Kim, S., Kang, J., and Hong, S.-M. (2009). Robust likelihood-based survival modeling for microarray gene expression Data. *J. Stat. Software* 29 (1), 1–16.
- Chu, H., Hui, G., Yuan, L., Shi, D., Wang, Y., Du, M., et al. (2015). Identification of novel piRNAs in bladder cancer. *Cancer Lett.* 356, 561–567. doi: 10.1016/j.canlet.2014.10.004
- Damborský, P., Damborská, D., Belický, Š., Tkáč, J., and Katrlík, J. (2017). Sweet Strategies in Prostate Cancer Biomarker Research: Focus on a Prostate Specific Antigen. *BioNanoScience* 8 (2), 690–700. doi: 10.1007/s12668-017-0397-z

AUTHOR CONTRIBUTIONS

ZW designed the experiments. YZu, YL, JZ, and YH performed data analysis. YZu wrote the initial version of manuscript. YZu, YL and ML prepared all the figures. ZW and YZh discussed the results and revised the manuscript. All authors contributed to discussions regarding the results and the manuscript.

FUNDING

This project was supported by a grant from the National Natural Science Foundation of China (no. 21575094).

ACKNOWLEDGMENTS

We would like to thank Kegin Liu, Jiwei Xue and Suqing Li for assisting in this study, and Yulan Deng, Lei Zhu for critical reading and discussion of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01018/full#supplementary-material>

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21. doi: 10.1093/bioinformatics/bts635
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874. doi: 10.1038/nrg3074
- Firmino, N., Martinez, V. D., Rowbotham, D. A., Enfield, K. S., Bennewith, K. L., and Lam, W. L. (2016). HPV status is associated with altered PIWI-interacting RNA expression pattern in head and neck cancer. *Oral. Oncol.* 55, 43–48. doi: 10.1016/j.oraloncology.2016.01.012
- Hashim, A., Rizzo, F., Marchese, G., Ravo, M., Tarallo, R., Nassa, G., et al. (2014). RNA sequencing identifies specific PIWI-interacting small non-coding RNA expression patterns in breast cancer. *Oncotarget* 5, 9901–9910. doi: 10.18632/oncotarget.2476
- Herman, C., Kattan, M., Otori, M., Scardino, P., and Wheeler, T. (2001). Primary Gleason pattern as a predictor of disease progression in gleason score 7 prostate cancer: a multivariate analysis of 823 men treated with radical prostatectomy. *Am. J. Surg. Pathol.* 25, 657–660. doi: 10.1097/00000478-200105000-00014
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4, 44–57. doi: 10.1038/nprot.2008.211
- Jin, G., Sun, J., Isaacs, S. D., Wiley, K. E., Kim, S. T., Chu, L. W., et al. (2011). Human polymorphisms at long non-coding RNAs (lncRNAs) and association with prostate cancer risk. *Carcinogenesis* 32, 1655–1659. doi: 10.1093/carcin/bgr187
- Juliano, C., Wang, J., and Lin, H. (2011). Uniting germline and stem cells: the function of Piwi proteins and the piRNA pathway in diverse organisms. *Annu. Rev. Genet.* 45, 447–469. doi: 10.1146/annurev-genet-110410-132541
- Kienesberger, P. C., Oberer, M., Lass, A., and Zechner, R. (2009). Mammalian patatin domain containing proteins: a family with diverse lipolytic activities involved in multiple biological functions. *J. Lipid Res.* 50 Suppl, S63–S68. doi: 10.1194/jlr.R800082-JLR200
- Krishnan, A. R., Korrapati, A., Zou, A. E., Qu, Y., Wang, X. Q., Califano, J. A., et al. (2017). Smoking status regulates a novel panel of PIWI-interacting RNAs in

- head and neck squamous cell carcinoma. *Oral. Oncol.* 65, 68–75. doi: 10.1016/j.oraloncology.2016.12.022
- Krishnan, P., Ghosh, S., Graham, K., Mackey, J. R., Kovalchuk, O., and Damaraju, S. (2016). Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer. *Oncotarget* 7, 37944–37956. doi: 10.18632/oncotarget.9272
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 9, 559. doi: 10.1186/1471-2105-9-559
- Lavery, H. J., and Droller, M. J. (2012). Do Gleason patterns 3 and 4 prostate cancer represent separate disease states? *J. Urology* 188, 1667–1675. doi: 10.1016/j.juro.2012.07.055
- Lee, E. J., Banerjee, S., Zhou, H., Jammalamadaka, A., Arcila, M., Manjunath, B. S., et al. (2011). Identification of piRNAs in the central nervous system. *RNA* 17, 1090–1099. doi: 10.1261/rna.2565011
- Li, Y., Wu, X., Gao, H., Jin, J. M., Li, A. X., Kim, Y. S., et al. (2015). Piwi-Interacting RNAs (piRNAs) Are Dysregulated in Renal Cell Carcinoma and Associated with Tumor Metastasis and Cancer-Specific Survival. *Mol. Med.* 21, 381–388. doi: 10.2119/molmed.2014.00203
- Lim, R. S., and Kai, T. (2015). A piece of the pi(e): The diverse roles of animal piRNAs and their PIWI partners. *Semin. Cell Dev. Biol.* 47–48, 17–31. doi: 10.1016/j.semcdb.2015.10.025
- Long, Q., Johnson, B. A., Osunkoya, A. O., Lai, Y. H., Zhou, W., Abramovitz, M., et al. (2011). Protein-coding and microRNA biomarkers of recurrence of prostate cancer following radical prostatectomy. *Am. J. Pathol.* 179, 46–54. doi: 10.1016/j.ajpath.2011.03.008
- Long, Q., Xu, J., Osunkoya, A. O., Sannigrahi, S., Johnson, B. A., Zhou, W., et al. (2014). Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer Res.* 74, 3228–3237. doi: 10.1158/0008-5472.CAN-13-2699
- Martinez, V. D., Enfield, K. S., Rowbotham, D. A., and Lam, W. L. (2016). An atlas of gastric PIWI-interacting RNA transcriptomes and their utility for identifying signatures of gastric cancer recurrence. *Gastric Cancer* 19, 660–665. doi: 10.1007/s10120-015-0487-y
- Martinez, V. D., Vucic, E. A., Thu, K. L., Hubaux, R., Enfield, K. S., Pikor, L. A., et al. (2015). Unique somatic and malignant expression patterns implicate PIWI-interacting RNAs in cancer-type specific biology. *Sci. Rep.* 5, 10423. doi: 10.1038/srep10423
- Mei, Y., Clark, D., and Mao, L. (2013). Novel dimensions of piRNAs in cancer. *Cancer Lett.* 336, 46–52. doi: 10.1016/j.canlet.2013.04.008
- Moyano, M., and Stefani, G. (2015). piRNA involvement in genome stability and human cancer. *J. Hematol. Oncol.* 8, 38. doi: 10.1186/s13045-015-0133-5
- Ng, K. W., Anderson, C., Marshall, E. A., Minatel, B. C., Enfield, K. S., Saprunoff, H. L., et al. (2016). Piwi-interacting RNAs in cancer: emerging functions and clinical utility. *Mol. Cancer* 15, 5. doi: 10.1186/s12943-016-0491-9
- Patel, K. M., and Gnanapragasam, V. J. (2016). Novel concepts for risk stratification in prostate cancer. *J. Clin. Urol.* 9, 18–23. doi: 10.1177/2051415816673502
- Preethi Krishnan, S. G., Graham, K., John, R., Mackey, O., and Kovalchuk, S. D. (2016). Piwi-interacting RNAs and PIWI genes as novel prognostic markers for breast cancer. *Oncotarget* 7 (25), 37944–37956. doi: 10.18632/oncotarget.9272
- Rajan, K. S., Velmurugan, G., Gopal, P., Ramprasath, T., Babu, D. D., Krithika, S., et al. (2016). Abundant and altered expression of piwi-interacting rnas during cardiac hypertrophy. *Heart Lung Circ.* 25, 1013–1020. doi: 10.1016/j.hlc.2016.02.015
- Ru, P., Steele, R., Newhall, P., Phillips, N. J., Toth, K., and Ray, R. B. (2012). miRNA-29b suppresses prostate cancer metastasis by regulating epithelial-mesenchymal transition signaling. *Mol. Cancer Ther.* 11, 1166–1173. doi: 10.1158/1535-7163.MCT-12-0100
- Sai Lakshmi, S., and Agrawal, S. (2008). piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.* 36, D173–D177. doi: 10.1093/nar/gkm696
- Vagin, V. V., Sigova, A., Li, C., Seitz, H., Gvozdev, V., and Zamore, P. D. (2006). A distinct small RNA pathway silences selfish genetic elements in the germline. *Science* 313, 320–324. doi: 10.1126/science.1129333
- Wilson, P. A., Gardner, S. D., Lambie, N. M., Commans, S. A., and Crowther, D. J. (2006). Characterization of the human patatin-like phospholipase family. *J. Lipid Res.* 47, 1940–1949. doi: 10.1194/jlr.M600185-JLR200
- Zhu, M., Chen, Q., Liu, X., Sun, Q., Zhao, X., Deng, R., et al. (2014). lncRNA H19/miR-675 axis represses prostate cancer metastasis by targeting TGFBI. *FEBS J.* 281, 3766–3775. doi: 10.1111/febs.12902

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Zuo, Liang, Zhang, Hao, Li, Wen and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Sodium Taurocholate Cotransporting Polypeptide (NTCP) Deficiency Hidden Behind Citrin Deficiency in Early Infancy: A Report of Three Cases

Hui Lin, Jian-Wu Qiu, Yaqub-Muhammad Rauf, Gui-Zhi Lin, Rui Liu, Li-Jing Deng, Mei Deng and Yuan-Zong Song*

Department of Pediatrics, The First Affiliated Hospital, Jinan University, Guangzhou, China

OPEN ACCESS

Edited by:

Tielu Shi,
East China Normal University, China

Reviewed by:

Dirk Rudi De Waart,
Academic Medical Center (AMC),
Netherlands
Yaqiong Jin,
Capital Medical University, China

*Correspondence:

Yuan-Zong Song
songyuanzong@vip.tom.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 12 February 2019

Accepted: 16 October 2019

Published: 07 November 2019

Citation:

Lin H, Qiu J-W, Rauf Y-M, Lin G-Z,
Liu R, Deng L-J, Deng M and
Song Y-Z (2019) Sodium Taurocholate
Cotransporting Polypeptide (NTCP)
Deficiency Hidden Behind Citrin
Deficiency in Early Infancy:
A Report of Three Cases.
Front. Genet. 10:1108.
doi: 10.3389/fgene.2019.01108

Sodium taurocholate cotransporting polypeptide (NTCP), a carrier protein encoded by the gene *SLC10A1*, is expressed in the basolateral membrane of the hepatocyte to uptake bile acids from plasma. As a new inborn error of bile acid metabolism, NTCP deficiency remains far from being well understood in terms of the clinical and molecular features. Citrin deficiency is a well-known autosomal recessive disease arising from *SLC25A13* mutations, and in neonates or infants, this condition presents as transient intrahepatic cholestasis which usually resolves before 1 year of age. All the three patients in this paper exhibited cholestatic jaundice and elevated total bile acids in their early infancy, which were attributed to citrin deficiency by *SLC25A13* genetic analysis. In response to feeding with lactose-free and medium-chain triglycerides-enrich formula, their clinical and laboratory presentations disappeared gradually while the hypercholanemia persisted, even beyond 1 year of age. On subsequent *SLC10A1* analysis, they were all homozygous for the well-known pathogenic variant c.800C > T (p.Ser267Phe), and NTCP deficiency was thus definitely diagnosed. The findings in this paper indicated that NTCP deficiency could be covered up by citrin deficiency during early infancy; however, in citrin-deficient patients with intractable hypercholanemia following resolved cholestatic jaundice, NTCP deficiency should be taken into consideration.

Keywords: cholestasis, citrin deficiency, sodium taurocholate cotransporting polypeptide deficiency, *SLC25A13*, *SLC10A1*, variant, child

BACKGROUND

Sodium taurocholate cotransporting polypeptide (NTCP) is a carrier protein in the basolateral membrane of the hepatocyte to uptake bile acids from plasma, playing a crucial role in the enterohepatic circulation of bile acids (Hagenbuch and Dawson, 2004). Although the causative gene *SLC10A1* was cloned as early as in 1994 (Hagenbuch and Meier, 1994) and NTCP function has been studied extensively (Ho et al., 2004; Pan et al., 2011; Yan et al., 2012; Yan et al., 2014), NTCP deficiency, as an inborn error of bile acid metabolism, was just described in very recent years. It was in 2015 that the first patient with NTCP deficiency was reported by Vaz et al. (2015). Following that, some articles involving patients with NTCP deficiency have been published (Deng et al., 2016;

Liu et al., 2017; Qiu et al., 2017; Song and Deng, 2017; Van Herpe et al., 2017; Li et al., 2018), but the reported patients were rather limited in number, and the genotypic and phenotypic features of this condition remained far from being completely understood.

Citrin, a bipartite protein in the mitochondrial inner membrane, has been well-known as the aspartate-glutamate carrier isoform 2 (AGC2), playing a significant role in the malate shuttle, urea cycle as well as gluconeogenesis from lactate (Begum et al., 2002; Saheki et al., 2005; Palmieri, 2013; Palmieri, 2014). *SLC25A13*, the gene encoding citrin, was cloned in the year 1999 (Kobayashi et al., 1999), and citrin deficiency encompassed three age-dependent clinical phenotypes, i.e. Neonatal Intrahepatic Cholestasis caused by Citrin Deficiency (NICCD) in neonates or infants (Ohura et al., 2001; Tazawa et al., 2001; Tomomasa et al., 2001), adult-onset citrullinemia type II (CTLN2) in adolescents or adults (Kobayashi et al., 1999), and Failure to Thrive and Dyslipidemia caused by Citrin Deficiency (FTTDCD) at pediatric age beyond 1 year (Song et al., 2011; Saheki and Song, 2017). To the best of our knowledge, although the clinical and molecular characteristics of NICCD has been studied for years (Ohura et al., 2007; Chen et al., 2013; Song et al., 2013; Ricciuto and Buhas, 2014; Zeng et al., 2014; Wang et al., 2015; Lin et al., 2016; Zhang et al., 2017), patients with NTCP deficiency complicated by NICCD have never been reported thus far.

Very recently, our team diagnosed three pediatric patients suffering from citrin deficiency and NTCP deficiency as well, and their molecular and clinical findings were reported herein.

CASE PRESENTATION

Patient 1 was a 5-year-and-11-month-old female referred to the First Affiliated Hospital, Jinan University due to abnormal liver function discovered for 5 years and 7 months. When aged 4 months, she was admitted to a hospital in Guangzhou due to jaundice for 3 months, where physical examination revealed an enlarged liver 4.0 cm below the right costal margin, and a liver

function test revealed elevated serum levels of total bilirubin (TBIL), direct bilirubin (DBIL), indirect bilirubin (IBIL), alanine transaminase (ALT), aspartate transaminase (AST), γ -glutamyl transpeptidase (GGT), and alkaline phosphatase (ALP), indicating cholestatic jaundice (**Table 1**). Blood amino acid spectrum analysis by tandem mass spectrometry (MS-MS) revealed raised citrulline, methionine, arginine, and threonine, while large quantities of galactose, galactitol, galactonate, and 4-hydroxyphenyllactate (4HPL) were detected on urinary gas chromatography-mass spectrometry (GC-MS) analysis. Considering the above clinical and laboratory findings, NICCD was suspected, and breast-feeding was stopped while a lactose-free and medium-chain triglycerides (MCT)-enriched formula was introduced. When aged 4.9 months, *SLC25A13* genetic analysis in our hospital unveiled a homozygote of the c.852_855del4 mutation (**Figure 1A**) and the diagnosis of NICCD was hence made. Thereafter, besides feeding with the therapeutic formula, supplemental foods rich in protein were encouraged. As a result, her liver function indices got improved gradually and returned to normal by age 10.2 months. However, the hypercholanemia was refractory, with total bile acid (TBA) levels fluctuating from 27.6 $\mu\text{mol/L}$ to 340.2 $\mu\text{mol/L}$ (reference range: 0–10 $\mu\text{mol/L}$) (**Table 1**). After the age 2 years, the patient showed a fondness for foods rich in protein and fat while an aversion to carbohydrate-rich diets.

As the first product of a non-consanguineous couple, the child was delivered spontaneously at the gestational age of 38 weeks and 3 days after an uneventful pregnancy, with a birth weight of 3.0 kg and body length 50 cm. The parents were healthy, and there was no family history of any genetic or infectious diseases.

Physical examination revealed a body temperature (T) 36.5°C, heart rate (HR) 115 beats/min (bpm), respiratory rate (RR) 20 bpm, weight (WT) 18 kg, height 107.0 cm. No jaundice was observed in the skin and sclera. The lungs were clear. No murmurs or abnormal heart sounds were heard. There was no abdominal distention, and the liver and spleen were not enlarged. Physiological reflexes were normal and no pathological reflexes could be found on nervous system examination. The extremities were warm, and the distal perfusion was excellent.

TABLE 1 | Biochemical alterations over time in patient 1 and the parents.

Indices (reference range)	Ages of patient 1														Father	Mother
	4 M	4.9 M	6 M	8 M	10.2 M	1Y1M	1Y4M	1Y7M	1Y10M	2Y3M	2Y8M	3Y8M	4Y8M	5Y11M		
ALT (5–40 U/L)	54	52	81	79	38	29	19	28	22	39	18	19	14	20	28	18
AST (5–40 U/L)	124	98	113	74	33	32	40	32	33	53	32	61	29	24	19	20
GGT (8–50 U/L)	150	138	42	23	19	14	—	10	14	13	10	13	11	12	30	12
ALP (40–500 U/L)	644	353	355	259	341	327	—	388	232	227	287	224	277	257	77	65
TP (60.0–83.0 g/L)	50.6	63.4	61.2	64.2	63.4	69.4	—	69	71.9	70.2	74.6	73.7	73.9	73.9	75.1	75.4
Alb (35.0–55.0 g/L)	38.2	43.0	47.4	48.4	47.6	49.6	—	47.2	47.7	43.5	49.7	50.7	48.1	49.5	46.1	40.6
Glb (20.0–30.0 g/L)	12.4	20.4	13.5	15.8	15.8	19.8	—	21.8	24.2	31.3	20.5	23.9	25.6	24.4	29	34.8
Tbil (2–19 μmol/L)	97.3	28.2	7.2	5	4.4	4.5	5.1	7.1	4.9	5.5	5.0	7.7	5.5	4.8	7.6	6.6
Dbil (0–6 μmol/L)	73.9	21.8	4.0	2.3	1.4	1.4	1.4	2.3	0.6	2.1	1.5	2.6	1.7	1.7	1.9	1.3
Ibil (2.56–20.9 μmol/L)	23.4	6.4	3.2	2.7	3.0	3.1	3.7	4.8	4.3	3.4	3.5	5.1	3.8	3.1	5.7	5.3
TBA (0–10 μmol/L)	—	340.2	174	135.2	132.8	123.2	82.1	109.3	165.4	111.1	105.8	173.6	27.6	48.7	3.2	3.7

ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, γ -glutamyl transpeptidase; ALP, alkaline phosphatase; TP, total protein; Alb, albumin; Glb, globulin; Tbil, total bilirubin; Dbil, direct bilirubin; Ibil, indirect bilirubin; TBA, total bile acids; —, not tested. For the ages, Y represents years; M, months.

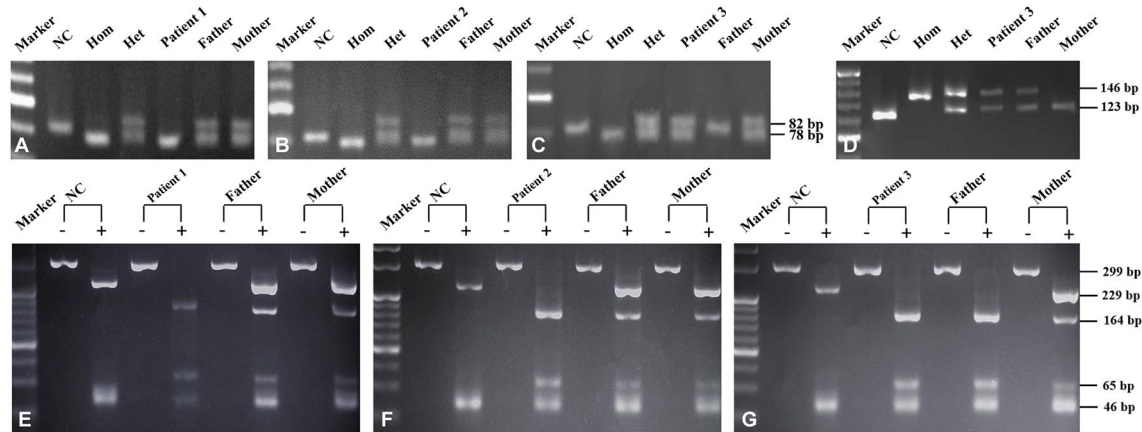


FIGURE 1 | *SLC25A13* and *SLC10A1* genetic analysis of the three families. *SLC25A13* analysis revealed that patients 1 (A) and 2 (B) were both homozygotes of the c.852_855del4 mutation while patient 3 (C and D), a compound heterozygote of the mutations c.852_855del4 and c.1638_1660dup. On *SLC10A1* analysis, patients 1 (E), 2 (F), and 3 (G) as well as her father were all homozygotes, while the parents of patients 1 and 2 and the mother of patient 3, all carriers, of the variant c.800C>T(p.Ser267Phe). NC (normal control), Hom (homozygous control), Het (heterozygous control). The “-” and “+” over every lane in panels E–G represented with and without enzymatic digestion by using the HphI enzyme, respectively.

Laboratory investigation showed otherwise normal biochemical indices but a TBA level of 48.7 $\mu\text{mol/L}$. In view of the intractable hypercholanemia, NTCP deficiency was highly suspected, and *SLC10A1* genetic analysis was performed. As a result, the patient was a homozygote, while the parents, carriers, of the reportedly pathogenic variant c.800C > T(p.Ser267Phe) (Figure 1E). NTCP deficiency was thus definitely diagnosed. No specific therapy was given but close clinic follow-up was underway.

Patient 2 was a 1-year-and-1-month-old male visiting our clinic due to hypercholanemia discovered for 11 months. At the age 2 months, he was referred to a local hospital because of prolonged jaundice for about 1 month. On biochemistry analysis, elevated serum levels of AST, GGT, ALP, TBIL, DBIL, and IBIL, together with decreased level of albumin were detected, and notably, the TBA level reached 268.7 $\mu\text{mol/L}$ (Table 2). Subsequent urinary GC-MS analysis detected elevated

4-hydroxyphenylpyruvate (4-HPPV) and 4HPL, while raised levels of citrulline, methionine, and threonine were detected on MS-MS analysis of blood sample. When aged 2.1 months, the infant undertook *SLC25A13* analysis in our hospital, and proved to be a homozygote of the c.852_855del4 mutation (Figure 1B), and the diagnosis of NICCD was thus made. Then breast-feeding was stopped and a lactose-free and MCT-enriched formula was given. Then his jaundice disappeared rapidly, and serum bilirubin levels returned to normal at his age 5.5 months (Table 2). However, the hypercholanemia persisted, even beyond 1 year of age (Table 2). No steatorrhea or acholic stool was observed during the course of the disease.

As the first child of a non-consanguineous couple, the patient was delivered vaginally at the gestational age of 37 weeks and 4 days with the birth weight 2700 g. The Apgar score was 9 points at 1 min and 10 points at 5 min after umbilical ligation. Parents were both hepatitis B virus (HBV) carriers, who were apparently

TABLE 2 | Biochemical changes over time in patient 2 and the parents.

Indices (reference range)	Ages of patient 2							Father	Mother
	2M	2.1M	3.2M	5.5M	8.5M	1Y 1M	1Y 6M		
ALT (5–40 U/L)	23	15	23	21	31	35	11.5	35	13
AST (5–40 U/L)	105	45	48	34	41	42	31.6	34	18
GGT (8–50 U/L)	281	281	162	18	16	16	18	24	10
ALP (20–220 U/L)	1,681	422	312	247	209	214	170.5	79	78
TP (60.0–83.0 g/L)	49.1	45.5	57.8	60.2	61.9	68	70.2	76.9	73.2
Alb (35.0–55.0 g/L)	32.4	30	40.4	46.3	46.2	45.6	43.7	46.1	43.1
Glb (20.0–30.0 g/L)	—	15.5	17.4	13.9	15.7	22.4	26.5	30.8	30.1
Tbil (2–19 $\mu\text{mol/L}$)	204.4	79.1	15.5	4.7	7.3	6.7	6.0	8.8	8.1
Dbil (0–6 $\mu\text{mol/L}$)	93.3	43.2	12.7	2.2	1.4	1.5	1.1	1.9	2.2
Ibil (2.56–20.9 $\mu\text{mol/L}$)	111.1	35.9	2.8	2.5	5.9	5.2	4.9	6.9	5.9
TBA (0–10 $\mu\text{mol/L}$)	268.7	110.8	233.8	143.3	208.7	234.5	148	31.5	78.8

Abbreviations as in Table 1.

healthy but with slightly raised serum TBA levels (Table 2). Family history of any genetic diseases was denied.

Physical examination revealed a body T 36.6°C, weight 10.5 kg, HR 126 bpm, and RR 32 bpm. No jaundiced skin and sclera was observed. On auscultation, no abnormal sounds were heard on the lungs and heart. There was no abdominal distention, and the liver and spleen were non-palpable. Primitive reflexes were normal and pathological reflexes could not be found on nervous system examination.

Laboratory test at visiting revealed a serum TBA level of 234.5 μmol/L and otherwise normal indices. *SLC10A1* genetic analysis demonstrated that the patient was a homozygote, and his parents, carriers, of the variant c.800C > T (p.Ser267Phe) (Figure 1F). The diagnosis of NTCP deficiency was thus made. No specific therapy was given but clinic follow up was suggested. His serum TBA level was 148 μmol/L (Table 2) when aged 1 year and 6 months, and a fondness for low-carbohydrate and high-protein foods was noticed since the age of 1 year.

Patient 3 was a 1-year-and-2-month-old female referred to our hospital because of abnormal liver function discovered for 12.7 months. At the age 1.3 months, she went through a liver function test because of prolonged jaundice for 1 month, which showed raised levels of AST, GGT, ALP, TBIL, DBIL, and IBIL (Table 3). When aged 1.8 months, her TBA level was found to be as high as 172.0 μmol/L besides the cholestatic alterations (Table 3), and the MS-MS analysis revealed increased levels of tyrosine, citrulline, and methionine while large quantities of urinary 4HPPV and 4HPL were detected on GC-MS analysis. NICCD was consequently suspected, and breast-feeding was stopped while a lactose-free and MCT-enriched formula was suggested. Following that, her cholestatic jaundice got alleviated rapidly and the laboratory alterations recovered to normal levels by the age 5 months, while the hypercholanemia was intractable, even beyond 1 year of age (Table 3).

As the first child of a non-consanguineous couple, the infant was delivered by cesarean section at the gestation age of 38 weeks and 2 days with the birth weight 2,750 g. Her father was clinically healthy with an elevated serum TBA level of 21.1 μmol/L (0–10 μmol/L), and her mother was physically and biochemically healthy (Table 3). There was no family history of any genetic diseases.

Physical examination at referral revealed a body weight 10.1 kg, length 80 cm and head circumference 46 cm. No jaundice was observed in the skin and sclera. Examinations of the heart, the lungs, the abdomen, and nervous system were all normal.

Biochemical test at referral revealed a TBA level 50.9 μmol/L with otherwise normal indices (Table 3). On genetic analysis, the patient was a compound heterozygote of the *SLC25A13* mutations c.852_855del4 and c.1638_1660dup, which was inherited from the father and mother, respectively (Figures 1C, D); moreover, the patient and her father were both homozygous for the *SLC10A1* variant c.800C > T (p.Ser267Phe), while her mother was a carrier (Figure 1G). Hence, citrin deficiency and NTCP deficiency were definitely diagnosed for the infant. No specific therapy was given, and his TBA level tended downward to 25.6 μmol/L at the age of 3 years and 4 months (Table 3), still remaining beyond the upper limit. The patient also had a fondness of protein-rich foods while an aversion to carbohydrate-rich foods from the age 1 year.

The molecular findings above were further confirmed by Sanger sequencing (Figure 2) and illustrated as family tree diagrams (Figure 3). The clinical and molecular features of all the 3 patients were summarized in Supplementary Table 1.

DISCUSSION

All the three NICCD patients in this paper exhibited typical biochemical and clinical presentations of intrahepatic cholestasis, which were corrected by uptake of lactose-free and MCT-enriched formulas. Increased NADH/NAD⁺ ratio in the plasma of the hepatocyte was a critical pathophysiologic alteration of citrin deficiency, leading to energy shortage in the liver due to the impaired glycolysis (Saheki and Kobayashi, 2002; Saheki et al., 2010). MCTs were better absorbed as medium chain free fatty acids (MCFA) and transported *via* the portal vein, and then more quickly oxidized compared with long chain triglycerides (LCTs). MCFA oxidation within mitochondria produced acetyl-CoA, FADH₂, and NADH to yield energy; and excess acetyl-CoA could enhance malate–citrate shuttle activity, generating more cytosolic NAD⁺, thus

TABLE 3 | Biochemical indices over time in patient 3 and the parents.

Indices (reference range)	Ages of patient 3									Father	Mother
	1.3M	1.8M	5M	7.2M	1Y2M	1Y10M	2Y	2Y8M	3Y4M		
ALT (5–40 U/L)	—	29.3	23	23	21.3	14.4	18	21.5	14.7	14	12
AST (5–40 U/L)	112	92.1	37	42	35	32.8	38	39.2	31.3	19	14
GGT (8–50 U/L)	299	284.6	47	20	11.8	11.6	14	15	15.6	34	14
ALP (20–220 U/L)	1,124	731.0	267.0	168.0	225.1	174.7	226	190.2	136.6	75	38
TP (60.0–83.0 g/L)	38.89	44.2	61.8	65.3	61.3	68.1	70.5	78.5	71	77.1	74.7
Alb (35.0–55.0 g/L)	26.6	32.5	40.2	40.5	44.7	43.7	47.4	50.1	47.2	48.6	48.2
Glb (20.0–30.0 g/L)	12.29	11.7	21.6	24.8	16.6	24.4	23.1	28.4	23.8	28.5	26.49
Tbil (2–19 μmol/L)	266.17	255.4	8.1	8.2	5.4	8.7	10.1	3.3	5.1	19.9	10.4
Dbil (0–6 μmol/L)	69.26	89.3	3.0	2.4	0.9	2.0	1.7	1.6	1.7	3.8	2.1
Ibil (2.56–20.9 μmol/L)	196.91	166.1	5.1	5.8	4.5	6.8	8.4	1.7	3.4	16.1	8.3
TBA (0–10 μmol/L)	—	172.0	94.0	72.0	50.9	19.7	10.7	36.1	25.6	21.1	1.5

Abbreviations as in Table 1.

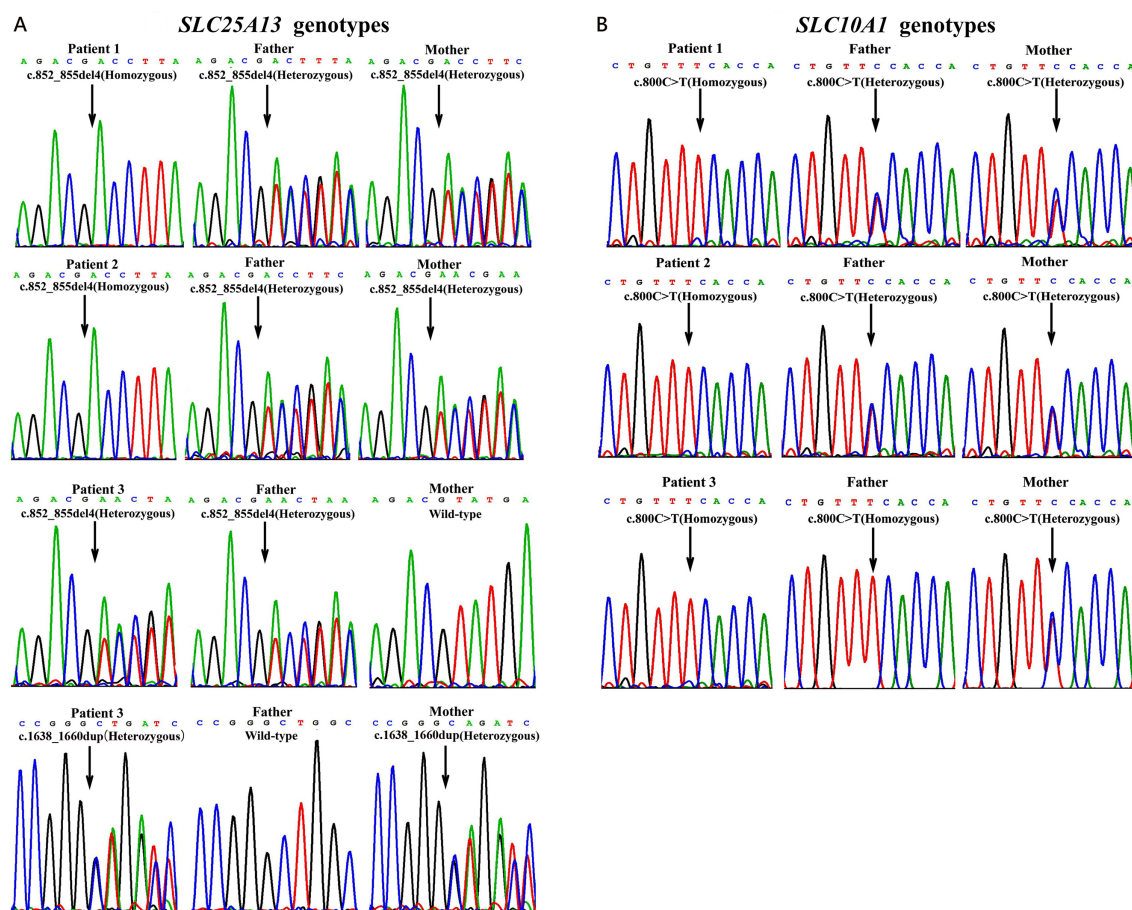


FIGURE 2 | Genotyping results by Sanger sequencing analysis in the three unrelated families. On *SLC25A13* analysis (A), patients 1 and 2 were both homozygotes of the c.852_855del4 mutation while patient 3, a compound heterozygote of the mutations c.852_855del4 and c.1638_1660dup. The parents of patients 1 and 2 as well as the father of patient 3 were all carriers of the c.852_855del4 mutation, while the mother of patient 3, a carrier of c.1638_1660dup. On *SLC10A1* analysis (B), patients 1, 2, and 3 as well as her father were all homozygotes, while the parents of patients 1 and 2 and the mother of patient 3, all carriers, of the variant c.800C > T(p.Ser267Phe).

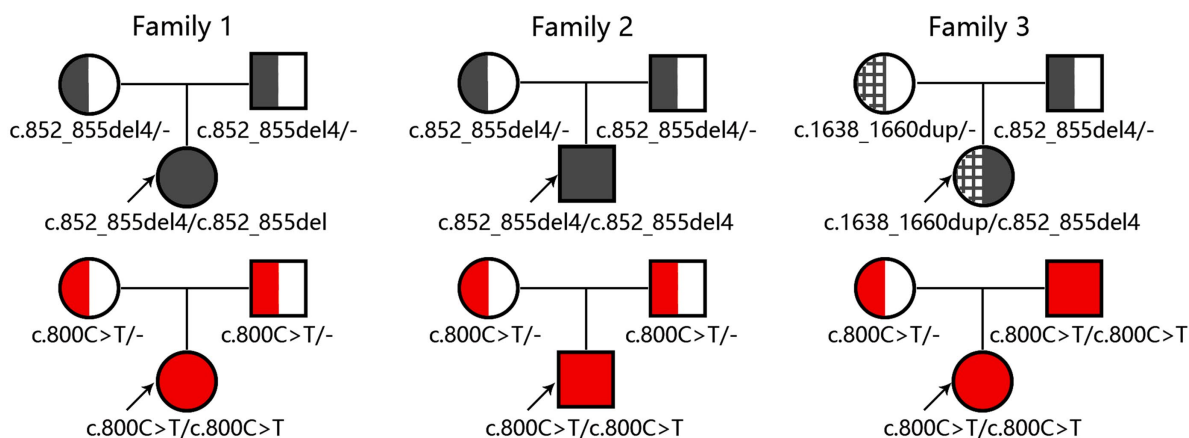


FIGURE 3 | Diagrams for family trees with the causative genes *SLC25A13* and *SLC10A1*. The upper and lower panel illustrated the *SLC25A13* and *SLC10A1* genotyping findings in the three families, respectively.

decreasing the NADH/NAD⁺ ratio (Hayasaka et al., 2012; Hayasaka and Numakura, 2018). On the other hand, the lactose was digested in the gut into glucose as well as galactose, and the latter was then absorbed into blood and converted into glucose by way of Leloir pathway in the liver (Frey, 1996). The galactose metabolism in the hepatocyte increased cytosolic NADH/NAD⁺ ratio and inhibited the activity of uridine diphosphate (UDP)-galactose-4-epimerase (Maxwell, 1957; Saheki et al., 2002), and the resultant secondary galactosemia injured the hepatocyte and led to hepatic dysfunction (Ning et al., 2000; Bosch, 2006; Song et al., 2010). Therefore, the lactose-free and MCTs-enriched formulas exhibited therapeutic effectiveness in the NICCD patients in this paper. Moreover, the peculiar food preferences in the citrin-deficient children beyond NICCD stage might be a self-saving dietary behavior to avoid raising the NADH/NAD⁺ ratio by too much carbohydrate uptake (Saheki et al., 2004; Saheki et al., 2005; Saheki et al., 2008).

The prominent evidence suggestive of NTCP deficiency in the three patients was persistent hypercholanemia after age 1 year. As the major carrier protein in the enterohepatic circulation of bile salts, NTCP uptakes conjugated bile salts from the plasma compartment into the hepatocyte in a sodium-dependent way (Hagenbuch and Meier, 1994). The *SLC10A1* p.Ser267Phe variant has been proved to be pathogenic functionally, bioinformatically and clinically, rendering NTCP without any function to uptake bile acids (Ho et al., 2004; Yan et al., 2014; Deng et al., 2016; Liu et al., 2017). The impaired NTCP function might be partially compensated by other transporters to uptake bile acids from the plasma, such as Organic Anion Transporting Polypeptide (OATP) 1B1 and 1B3 in the basolateral membrane of hepatocytes; however, in the absence of NTCP, they could just played a limited role in bile acid clearance and were unable to compensate for loss of NTCP (Karpen and Dawson, 2015). As such, it was not surprising for the four patients with NTCP deficiency, including three children and the father in family 3, to present with refractory hypercholanemia in this study (Tables 1–3). It was noteworthy that, although hypercholanemia was the unique clinical presentation for NTCP deficiency, this biochemical change itself was just a nonpathognomonic marker suggestive cholestatic liver disease, making NTCP deficiency be covered up by NICCD at early infancy, as in the three pediatric patients reported in this study.

Although molecular techniques and genetic data have significantly improved the understanding of rare diseases in the recent years (Jia and Shi, 2017; Ni and Shi, 2017), it was rather rare for two genetic diseases of the liver to affect the same individual. In this paper, however, citrin deficiency and NTCP deficiency were found to affect three pediatric patients simultaneously. This rare finding might be explained by the relatively high prevalence of the two genetic conditions in south China, especially in Guangdong province where the three patients were located. The allele frequency of *SLC10A1* variant c.800C > T varied in different populations, with the highest incidence occurring in Southern China (8% and 12% in Chinese Han and Dai respectively), suggesting that this

hypercholanemia affected 0.64% of the Southern Han as well as 1.44% of the Dai Chinese population (Liu et al., 2017). On the other hand, molecular epidemiological survey showed that the carrier rate of *SLC25A13* mutations was 1/940 in the north but 1/48 in the south of Yangtze River of mainland China (Lu et al., 2005). In particular, the carrier rate of five prevalent *SLC25A13* mutations (including c.851_854del, c.1638_1660dup, c.615+5G > A, IVS16ins3kb, and c.1399C > T) was about 1/47 in Guangdong province, with an estimated morbidity of 1/8,800 for patients with citrin deficiency (Zhang et al., 2014).

Interestingly, the parents of patient 2, two carriers of the p.Ser267Phe variant, also exhibited slightly elevated TBA levels (Table 2). However, this finding did not constitute a challenge against NTCP deficiency as an autosomal recessive disorder. As a reasonable explanation, their HBV carrier status might be responsible for their mild hypercholanemia. Actually, besides functioning as a carrier protein to uptake bile acids from plasma, NTCP had proven to be a functional receptor for HBV to cross the basolateral membrane, entering into the hepatocyte (Yan et al., 2012). It was reported that the NTCP residues between 157 to 165 were important for pre-S1 lipopeptide binding of the HBV large envelope protein, and contributed to HBV infections on HepG2 cells (Yan et al., 2012; Yan et al., 2013). Moreover, Yan et al. identified the HBV L-protein derived lipopeptides as inhibitors of NTCP, indicating that the specific pre-S1 lipopeptide binding might inhibit NTCP from transporting bile salts (Yan et al., 2014). In a word, being HBV carriers might block the function of NTCP to uptake bile acids from plasma.

In conclusion, this paper reported three pediatric patients with NTCP deficiency complicated by citrin deficiency. The findings indicated that NTCP deficiency could be covered up by citrin deficiency during early infancy; however, in citrin-deficient patients with intractable hypercholanemia following resolved cholestatic jaundice, NTCP deficiency should be taken into consideration.

DATA AVAILABILITY STATEMENT

All datasets generated and analyzed for this study are included in the article/Supplementary Material.

ETHICS STATEMENT

This study has been approved by the Committee for Medical Ethics, the First Affiliated Hospital, Jinan University. The authors declare that this study was performed after written informed consent had been obtained from the parents of the three families, which permitted publication of this case report.

AUTHOR CONTRIBUTIONS

HL, Y-ZS, and RL performed data collection and drafted the initial manuscript. Y-ZS conceptualized and designed the study,

critically reviewed and revised the manuscript. MD, L-JD, Y-MR, G-ZL, and J-WQ carried out the genetic analyses and reviewed the manuscript. Y-ZS managed and followed up the pediatric patients. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

The present study was supported by National Natural Science Foundation (NSFC) of China (Nos. 81570793, 81741080, and 81974057).

REFERENCES

- Begum, L., Jalil, M. A., Kobayashi, K., Iijima, M., Li, M. X., Yasuda, T., et al. (2002). Expression of three mitochondrial solute carriers, citrin, aralar1 and ornithine transporter, in relation to urea cycle in mice. *Biochim. Biophys. Acta* 1574, 283–292. doi: 10.1016/0167-4781(01)00376-1
- Bosch, A. M. (2006). Classical galactosaemia revisited. *J. Inherit. Metab. Dis.* 29, 516–525. doi: 10.1007/s10545-006-0382-0
- Chen, R., Wang, X. H., Fu, H. Y., Zhang, S. R., Abudouxikuer, K., Saheki, J. T., et al. (2013). Different regional distribution of SLC25A13 mutations in Chinese patients with neonatal intrahepatic cholestasis. *World J. Gastroenterol.* 19, 4545–4551. doi: 10.3748/wjg.v19.i28.4545
- Deng, M., Mao, M., Guo, L., Chen, F. P., Wen, W. R., and Song, Y. Z. (2016). Clinical and molecular study of a pediatric patient with sodium taurocholate cotransporting polypeptide deficiency. *Exp. Ther. Med.* 12, 3294–3300. doi: 10.3892/etm.2016.3752
- Frey, P. A. (1996). The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *J. FASEB* 4, 461–470.
- Hagenbuch, B., and Dawson, P. (2004). The sodium bile salt cotransport family slc10. *Pflugers Arch.* 447, 566–570. doi: 10.1007/s00424-003-1130-z
- Hagenbuch, B., and Meier, P. J. (1994). Molecular cloning, chromosomal localization, and functional characterization of a human liver Na⁺/bile acid cotransporter. *J. Clin. Invest.* 93, 1326–1331. doi: 10.1172/JCI117091
- Hayasaka, K., and Numakura, C. (2018). Adult-onset type II citrullinemia: Current insights and therapy. *Appl. Clin. Genet.* 11, 163–170. doi: 10.2147/TACG.S162084
- Hayasaka, K., Numakura, C., Toyota, K., and Kimura, J. T. (2012). Treatment with lactose(galactose)-restricted and medium-chain triglyceride-supplemented formula for neonatal intrahepatic cholestasis caused by citrin deficiency. *JIMD Rep.* 2, 37–44. doi: 10.1007/8904-2011-42
- Ho, R. H., Leake, B. F., Roberts, R. L., Lee, W., and Kim, R. B. (2004). Ethnicity-dependent Polymorphism in Na⁺-taurocholate Cotransporting Polypeptide (SLC10A1) Reveals a Domain Critical for Bile Acid Substrate Recognition. *J. Biol. Chem.* 279, 7213–7222. doi: 10.1074/jbc.m305782200
- Jia, J. M., and Shi, J. T. L. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Karpen, S. J., and Dawson, P. A. (2015). Not all (bile acids) who wander are lost: The first report of a patient with an isolated NTCP defect. *Hepatology* 61, 24–27. doi: 10.1002/hep.27294
- Kobayashi, K., Sinasac, D. S., Iijima, M., Bortright, A. P., Bequm, L., Lee, J. R., et al. (1999). The gene mutated in adult-onset type II citrullinaemia encodes a putative mitochondrial carrier protein. *Nat. Genet.* 22, 159–163. doi: 10.1038/9667
- Li, H., Qiu, J. W., Lin, G. Z., Deng, M., Lin, W. X., and Song, Y. Z. (2018). Clinical and genetic analysis of a pediatric patient with sodium taurocholate cotransporting polypeptide deficiency. *Zhongguo Dang Dai Er Ke Za Zhi* 20, 279–284. doi: 10.7499/j.issn.1008-8830.2018.04.005
- Lin, W. X., Zeng, H. S., Zhang, Z. H., Mao, M., Zheng, Q. Q., Zhao, S. T., et al. (2016). Molecular diagnosis of pediatric patients with citrin deficiency in

ACKNOWLEDGMENTS

We appreciate all the research subjects for their cooperation as well as the financial support of National Natural Science Foundation (NSFC) of China (Nos. 81570793, 81741080, and 81974057).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01108/full#supplementary-material>

- China: SLC25A13 mutation spectrum and the geographic distribution. *Sci. Rep.* 6, 29732. doi: 10.1038/srep29732
- Liu, R., Chen, C., Xia, X., Liao, Q., Wang, Q., Newcombe, P. J., et al. (2017). Homozygous p.Ser267Phe in SLC10A1 is associated with a new type of hypercholanemia and implications for personalized medicine. *Sci. Rep.* 7, 9214. doi: 10.1038/s41598-017-07012-2
- Lu, Y. B., Kobayashi, K., Ushikai, M., Tabata, A., Iijima, M., Li, M. X., et al. (2005). Frequency and distribution in East Asia of 12 mutations identified in the SLC25A13 gene of Japanese patients with citrin deficiency. *J. Hum. Genet.* 50, 338–346. doi: 10.1007/s10038-005-0262-8
- Maxwell, E. (1957). The enzymatic interconversion of uridine diphosphogalactose and uridine diphosphoglucose. *J. Biol. Chem.* 229, 139–151.
- Ni, X., and Shi, T. L. (2017). The Challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60, 681–685. doi: 10.1007/s11427-017-9100-1
- Ning, C., Reynolds, R., Chen, J., Yager, C., and Berry, G. T. (2000). Galactose metabolism by the mouse with galactose-1-phosphate uridylyltransferase deficiency. *Pediatr. Res.* 48, 211–217. doi: 10.1203/00006450-200008000-00015
- Ohura, T., Kobayashi, K., Tazawa, Y., Abukawa, D., Sakamoto, O., Tsuchiya, S., et al. (2007). Clinical pictures of 75 patients with neonatal intrahepatic cholestasis caused by citrin deficiency (NICCD). *J. Inherit. Metab. Dis.* 30, 139–144. doi: 10.1007/s10545-007-0506-1
- Ohura, T., Kobayashi, K., Tazawa, Y., Nishi, I., Abukawa, D., Sakamoto, O., et al. (2001). Neonatal presentation of adult-onset type II citrullinemia. *Hum. Genet.* 108, 87–90.
- Palmieri, F. (2013). The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol. Aspects. Med.* 34, 465–484. doi: 10.1016/j.mam.2012.05.005
- Palmieri, F. (2014). Mitochondrial transporters of the SLC25 family and associated diseases: a review. *J. Inherit. Metab. Dis.* 37, 565–575. doi: 10.1007/s10545-014-9708-5
- Pan, W., Song, I. S., Shin, H. J., Kim, M. H., Choi, Y. L., Lim, S. J., et al. (2011). Genetic polymorphisms in Na⁺-taurocholate co-transporting polypeptide (NTCP) and ileal apical sodium-dependent bile acid transporter (ASBT) and ethnic comparisons of functional variants of NTCP among Asian populations. *Xenobiotica* 41, 501–510. doi: 10.3109/00498254.2011.555567
- Qiu, J. W., Deng, M., Cheng, Y., Atif, R. M., Lin, W. X., Gou, L., et al. (2017). Sodium taurocholate cotransporting polypeptide (NTCP) deficiency: Identification of a novel SLC10A1 mutation in two unrelated infants presenting with neonatal indirect hyperbilirubinemia and remarkable hypercholanemia. *Oncotarget* 8, 106598–106607. doi: 10.18632/oncotarget.22503
- Ricciuto, A., and Buhas, D. (2014). A novel citrin deficiency mutation in a cholestatic infant. *J. Pediatr. Gastroenterol. Nutr.* 59, e52. doi: 10.1097/MPG.0000000000000556
- Saheki, T., and Kobayashi, K. (2002). Mitochondrial aspartate glutamate carrier (citrin) deficiency as the cause of adult-onset type II citrullinemia (CTLN2) and idiopathic neonatal hepatitis (NICCD). *J. Hum. Genet.* 47, 333–341. doi: 10.1007/s100380200046
- Saheki, T., and Song, Y. Z. (2017). “Citrin Deficiency,” in *GeneReviews [Internet]*. Eds. Adam, M. P., HH, Ardinger, RA, Pagon, SE, Wallace, Bean,

- L. J. H., Stephens, K., and Amemiya, A. (Seattle (WA): University of Washington, Seattle), 1993–2019.
- Saheki, T., Inoue, K., Tushima, A., Mutoh, K., Kobayashi, K., Saheki, T., et al. (2010). Citrin deficiency and current treatment concepts. *Mol. Genet. Metab.* 10, S59–S64. doi: 10.1016/j.yimgme.2010.02.014
- Saheki, T., Kobayashi, K., Iijima, M., Horiuchi, M., Begum, L., Jalil, M. A., et al. (2004). Adult-onset type II citrullinemia and idiopathic neonatal hepatitis caused by citrin deficiency: involvement of the aspartate glutamate carrier for urea synthesis and maintenance of the urea cycle. *Mol. Genet. Metab.* 81, S20–S26. doi: 10.1016/j.yimgme.2004.01.006.
- Saheki, T., Kobayashi, K., Iijima, M., Moriyama, M., Yazaki, M., Takei, Y., et al. (2005). Metabolic derangements in deficiency of citrin, a liver-type mitochondrial aspartate-glutamate carrier. *Hepatol. Res.* 33, 181–184. doi: 10.1016/j.hepres.2005.09.031
- Saheki, T., Kobayashi, K., Iijima, M., Nishi, I., Yasuda, T., Yamaguchi, N., et al. (2002). Pathogenesis and Pathophysiology of Citrin (a Mitochondrial Aspartate Glutamate Carrier) Deficiency. *Metab. Brain Dis.* 17, 335–346.
- Saheki, T., Kobayashi, K., Terashi, M., Ohura, T., Yanagawa, Y., Okano, Y., et al. (2008). Reduced carbohydrate intake in citrin-deficient subjects. *J. Inherit. Metab. Dis.* 31, 386–394. doi: 10.1007/s10545-008-0752-x
- Song, Y. Z., and Deng, M. (2017). Sodium taurocholate cotransporting polypeptide deficiency manifesting as cholestatic jaundice in early infancy: a complicated case study. *Zhongguo Dang Dai Er Ke Za Zhi* 19, 350–354. doi: 10.7499/j.issn.1008-8830.2017.03.020
- Song, Y. Z., Deng, M., Chen, F. P., Wen, F., Guo, L., Cao, S. L., et al. (2011). Genotypic and phenotypic features of citrin deficiency: Five-year experience in a Chinese pediatric center. *Int. J. Mol. Med.* 28, 33–40. doi: 10.3892/ijmm.2011.653
- Song, Y. Z., Wen, F., Chen, F. P., Kobayashi, K., and Saheki, T. (2010). Neonatal intrahepatic cholestasis caused by citrin deficiency: efficacy of therapeutic formulas and update of clinical outcomes. *Jpn. J. Inherit. Metab. Dis.* 26, 57–69.
- Song, Y. Z., Zhang, Z. H., Lin, W. X., Zhao, X. J., Deng, M., Ma, Y. L., et al. (2013). SLC25A13 gene analysis in citrin deficiency: sixteen novel mutations in East Asian patients, and the mutation distribution in a large pediatric cohort in China. *PLoS One* 8, e74544. doi: 10.1371/journal.pone.0074544
- Tazawa, Y., Kobayashi, K., Ohura, T., Abukawa, D., Nishinomiya, F., Hosoda, Y., et al. (2001). Infantile cholestatic jaundice associated with adult-onset type II citrullinemia. *J. Pediatr.* 138, 735–740. doi: 10.1067/mpd.2001.113264
- Tomomasa, T., Kobayashi, K., Kaneko, H., Shimura, H., Fukusato, T., Tabata, M., et al. (2001). Possible clinical and histologic manifestations of adult-onset type II citrullinemia in early infancy. *J. Pediatr.* 138, 741–743. doi: 10.1067/mpd.2001.113361
- Van Herpe, F., Waterham, H. R., Adams, C. J., Mannens, M., Bikker, H., Vaz, F. M., et al. (2017). NTCP deficiency and persistently raised bile salts: an adult case. *J. Inherit. Metab. Dis.* 40, 313–315. doi: 10.1007/s10545-017-0031-9
- Vaz, F. M., Paulusma, C. C., Huidekoper, H., de Ru, M., Lim, C., Koster, J., et al. (2015). Sodium taurocholate cotransporting polypeptide (SLC10A1) deficiency: Conjugated hypercholanemia without a clear clinical phenotype. *Hepatology* 61, 260–267. doi: 10.1002/hep.27240
- Wang, H., Shu, S., Chen, C., Huang, Z., and Wang, D. (2015). Novel mutations in the SLC25A13 gene in a patient with NICCD and severe manifestations. *J. Pediatr. Endocrinol. Metab.* 28, 471–475. doi: 10.1515/jpem-2014-0278
- Yan, H., Peng, B., He, W., Zhong, G., Qi, Y., Ren, B., et al. (2013). Molecular determinants of hepatitis B and D virus entry restriction in mouse sodium taurocholate cotransporting polypeptide. *J. Virol.* 88, 7977–7991. doi: 10.1128/JVI.03540-12
- Yan, H., Peng, B., Liu, Y., Xu, G., He, W., Ren, B., et al. (2014). Viral entry of Hepatitis B and D viruses and bile salts transportation share common molecular determinants on Sodium Taurocholate Cotransporting Polypeptide. *J. Virol.* 88, 3273–3284. doi: 10.1128/jvi.03478-13
- Yan, H., Zhong, G., Xu, G., He, W., Jing, Z., Gao, Z., et al. (2012). Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. *Elife* 1, e00049. doi: 10.7554/eLife.00049
- Zeng, H. S., Zhao, S. T., Deng, M., Zhang, Z. H., Cai, X. R., Chen, F. P., et al. (2014). Inspissated bile syndrome in an infant with citrin deficiency and congenital anomalies of the biliary tract and esophagus: identification and pathogenicity analysis of a novel SLC25A13 mutation with incomplete penetrance. *Int. J. Mol. Med.* 34, 1241–1248. doi: 10.3892/ijmm.2014.192
- Zhang, Z. H., Yang, Z. G., Chen, F. P., Kikuchi, A., Liu, Z. H., et al. (2014). Screening for five prevalent mutations of SLC25A13 gene in Guangdong, China: a molecular epidemiologic survey of citrin deficiency. *Tohoku J. Exp. Med.* 233, 275–281. doi: 10.1371/journal.pone.0089267
- Zhang, Z. H., Lin, W. X., Zheng, Q. Q., Guo, L., and Song, Y. Z. (2017). Molecular diagnosis of citrin deficiency in an infant with intrahepatic cholestasis: identification of a 21.7kb gross deletion that completely silences the transcriptional and translational expression of the affected SLC25A13 allele. *Oncotarget* 3, 87182–87193. doi: 10.18632/oncotarget.19901

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lin, Qiu, Rauf, Lin, Liu, Deng, Deng and Song. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Clinical Assessments and EEG Analyses of Encephalopathies Associated With Dynamin-1 Mutation

Hua Li, Fang Fang^{*}, Manting Xu, Zhimei Liu, Ji Zhou, Xiaohui Wang, Xiaofei Wang and Tongli Han

Department of Neurology, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing, China

OPEN ACCESS

Edited by:

Tielu Shi,

East China Normal University, China

Reviewed by:

Alberto Spalice,

Policlinico Umberto I, Italy

Li Zhang,

East China Normal University, China

*Correspondence:

Fang Fang

13910150389@163.com

Specialty section:

This article was submitted to
Translational Pharmacology,
a section of the journal
Frontiers in Pharmacology

Received: 03 January 2019

Accepted: 13 November 2019

Published: 04 December 2019

Citation:

Li H, Fang F, Xu M, Liu Z, Zhou J, Wang X, Wang X and Han T (2019)
Clinical Assessments and EEG
Analyses of Encephalopathies
Associated With Dynamin-1 Mutation.
Front. Pharmacol. 10:1454.
doi: 10.3389/fphar.2019.01454

Epileptic encephalopathy, caused by mutations in the dynamin-1 (*DNM1*; NM_004408) gene, is a newly identified neurologic disorder in children. Thus far, the full clinical and electroencephalographic features of children with *DNM1* mutation-related epileptic encephalopathy have not been established. The aim of this study is to characterize the phenotypic, genetic, and electroencephalographic features of children with *DNM1* mutation-related epileptic encephalopathy. Here, we investigated a patient with a novel pathogenic *DNM1* variant, who received treatment in Beijing Children's Hospital and had detailed clinical, EEG, and genetic information. Conversely, we performed an extensive literature search in PubMed, EMBASE, Cochrane Central Register of Controlled Trials, Chinese BioMedical Literature Database, China National Knowledge Infrastructure, and Wanfang Database using the term “*DNM1*” and were able to find 32 cases reported in nine articles (in English) from January 2013 to December 2018. The clinical features of 33 cases with pathogenic *DNM1* variants were analyzed and the results showed that patients carrying pathogenic variants in the GTPase or middle domains present with epileptic encephalopathy and severe neurodevelopmental symptoms. Patients carrying pathogenic variants in both domains exhibited comparable phenotypes.

Keywords: epileptic encephalopathy, dynamin-1, mutation, electroencephalogram, children

INTRODUCTION

Epileptic encephalopathy, caused by dynamin-1 (*DNM1*) mutations, is a newly characterized neurologic disorder in children (Kolnikova et al., 2018). The *DNM1* gene codes for the *DNM1* protein is involved in the synaptic vesicle cycle that facilitates the exocytosis of neurotransmitters during receptor-mediated endocytosis, which is necessary for signaling pathway function and central nervous system development. Pathogenic *DNM1* variants affect brain development and function and cause epileptic encephalopathy associated with severe neurodevelopmental complications (Allen et al., 2013; Appenzellar, 2014; Allen et al., 2016; Deng et al., 2016; Nakashima et al., 2016). Previously reported pathogenic variants of *DNM1* have been associated with early onset of epileptic encephalopathy (including West and Lennox-Gastaut syndromes) and are present in up to 2% of patients with infantile spasms or Lennox-Gastaut syndrome (Appenzellar, 2014). For many years, considerable attention has been paid to the genetic studies, patients carrying pathogenic variants in

the GTPase or middle domains of *DNM1* exhibit epileptic encephalopathy and severe neurodevelopmental complications. With no clear and effective treatment, antiepileptic medications, as a limited manner, are often insufficient for seizure control in patients with earlier onset and higher frequency of seizures. Thus far, the specific clinical and electroencephalographic features of children with *DNM1* mutation-related epileptic encephalopathy have not been clearly established. Here, we characterized the phenotypic, genetic, and electroencephalographic features of children with *DNM1* mutation-related epileptic encephalopathy.

MATERIALS AND METHODS

Patients

We reported a patient with a novel pathogenic *DNM1* variant, who received treatment in Beijing Children's Hospital and had detailed clinical, EEG and genetic information. In addition, we performed an extensive literature search in PubMed, EMBASE, Cochrane Central Register of Controlled Trials, Chinese BioMedical Literature Database, China National Knowledge Infrastructure and Wanfang Database using the term "*DNM1*," and were able to find 32 cases reported in nine articles with complete clinical data (in English) from January 2013 to December 2018 (Table 1). We then analyzed the clinical features of 33 cases with pathogenic *DNM1* variants, including gender, age at seizure onset, seizure types, development, number of antiepileptic drugs (AEDs) administrated, EEG, and mutations. Seizures and epilepsy syndromes were classified in accordance with the guidelines of the International League Against Epilepsy (Belousova et al., 2017). The analysis of the pathogenic effects of these variations on *DNM1* gene was conducted mainly by using three prediction algorithms: SIFT, PolyPhen 2 Hvar, Mutation taster (Table 2).

Statistical Analysis

The continuous variables were described by mean or median with range, and the categorical variables by number or percentage. Chi-square tests were used for comparison of phenotypic difference with mutations in different gene domains. *P*-value below 0.05 was considered significant. Statistical analysis was performed using SPSS 22.0.

Ethics Statement

The present study was approved by the Ethics Committee of Beijing Children's Hospital and informed consents were collected from the participant's parents.

RESULTS

Case Report

A 3-year-old girl presented with severe psychomotor developmental delay; nonverbal and non-ambulatory. She had been delivered at full term after 40 weeks of gestation (birth weight 3,150 g, Apgar score 10), and was the first child of the family. The patient has normal personal and family history. She

exhibited limb shaking at 2 months after birth and EEG analysis revealed no epileptiform discharges. Over the following 4 months, she exhibited shaking limbs intermittently upon waking. At 6 months of age, the patient exhibited "binocular vision and tongue vomiting" which gradually increased over time, although the time of onset was unclear. Video-EEG monitoring was performed multiple times, based on the suspicion of non-epileptic seizures. Levetiracetam was administrated during observation. Infantile spasms manifested at 8 months after birth. EEG analysis showed prime spike waves, a small number of multiple spike waves, spike-slow waves, and synchronous or non-synchronous discharge (atypical hypsarrhythmia) onto the bilateral rear head during seizures. Subsequent treatment (beginning at 8 months of age) constituted administration of levetiracetam and topiramate. Changes in EEG are shown in Figures 1–4. The bilateral ventricles were slightly widened on magnetic resonance images beginning at 2 months of age.

After obtaining informed consent from the parents, whole exome sequencing (WES) of the patient and parental samples were analyzed by a trio-based analysis, which identified a variant, c.135C > A, in the *DNM1* gene (NM_004408), with the amino acid changes of p.Ser45Arg (Figure 5), which was confirmed by Sanger sequencing. To the best of our knowledge, this was an unreported *de novo* mutation. This missense mutation is absent in gnomAD, ExAC, 1000 Genomes, and ESP 6500 databases; moreover, it is predicted to be a disease variant by Polyphen-2 (score of 0.988), MutationTaster (disease causing), SIFT (score of 0.006), and M-CAP (score of 0.979). According to sequence alignment, the Ser45 residue is highly conserved across species, indicating evolutionary importance (Figure 6).

Clinical Characteristics of Patients With Pathogenic *DNM1* Variants

Our data showed that patients carrying pathogenic variants in the GTPase or middle domains present with epileptic encephalopathy and severe neurodevelopmental symptoms. For the analysis of *DNM1*-related encephalopathy, 31 out of the 33 patients were included (9 females, 21 males, the sex of one patient was not available). The age range of the patients was 0.6–24 years, the median age was, at inclusion, 8 years. Pregnancy and delivery were unremarkable in all patients with normal birth parameters. Patient 25 died at 2 years of age, before enrolment in this study. The clinical characteristics of *DNM1* mutation-related epileptic encephalopathy patients were analyzed as follows:

Seizures

From the 31 patients with *DNM1* mutation-related epileptic encephalopathy, 29 (93.5%) experienced epileptic seizures. Patient 11 did not have seizures, and patient 24 showed only subcortical, nonepileptic myoclonic jerks. Seizures began at a median age of 5 months (range 1 day to 4.5 years). Patient 26 was an outlier with onset at 4.5 years with a febrile infection related epilepsy syndrome phenotype. During the course of the disease, 23 (74.2%) patients had spasm seizures, 12 (38.7%) patients had absence seizures, 9 (29.0%) patients had tonic seizures, 12 (38.7%) patients had myoclonic seizures, 5 (16.1%) patients had atonic

TABLE 1 | Clinical characteristics of patients with pathogenic DNM1 variants.

Case Sequence (reference number)	Sex, age (years)	Seizure onset (months)	Seizure type at onset	Course of seizures	Development	EEG feature
1 (Deng et al., 2016)	M, 5	7	Twitches: Nodded, struggled, strangled	AtS, MS	NVNA; severe ID; hypotonia	SSW, Hyps
2 (Nakashima et al., 2016)	M, 15	11	Neck antelexion, rolling of the eyes, and elevating upper limbs	AtS, MS	NVNA; severe ID; hypotonia	Partial hyps
3 (Nakashima et al., 2016)	M, 6	10	Atonic type with head dropping	AS	NVNA; severe ID; hypotonia	MFED
4 (Appenzeller, 2014)	M, 6	13	ES	Atypical-As, TS, FS	NVNA; profound ID; hypotonia	Hyps
5 (Appenzeller, 2014)	F, 15	7	ES	Atypical-As with eyelid fluttering, drop attacks, GTCS	Nonverbal; severe ID; hypotonia	Slow bg, MFED
6 (Appenzeller, 2014)	M, 8	6	ES	AtS	NVNA; severe ID; hypotonia	Hyps
7 (Appenzeller, 2014)	F, 13	12	ES	MS, atypical-As, FS, GTCS	NVNA; profound ID; hypotonia	Modified Hyps
8 (Appenzeller, 2014)	M, 6	2	ES	SZ-free since age 3.5 years	NVNA; severe ID; hypotonia	High voltage bilateral SSW
9 (Allen et al., 2016)	M, N/A	2	Nodded, Ms	N/A	Profound ID	MFED
10 (von Spiczak et al., 2017)	F, 8	0.75	N/A	AS, MS	NVNA; profound ID; hypotonia	Slow bg
11 (von Spiczak et al., 2017)	M, 2	None	None	None	NVNA; profound ID; hypotonia	Normal
12 (von Spiczak et al., 2017)	M, 8	13	IS	AS, TS, FS, SE	NVNA; profound ID; hypotonia	Hyps, MFED, GPFA, slow bg
13 (von Spiczak et al., 2017)	M, 18	4	IS	AS, TS, GTCS, SE	NVNA; profound ID; hypotonia	Hyps, MFED, SSW, GSW, slow bg
14 (von Spiczak et al., 2017)	F, 15	7	IS	AS, ATs, GTCS	Nonverbal; severe ID; hypotonia	MFED, SSW, slow bg
15 (von Spiczak et al., 2017)	M, 9	2	IS	AS, MS, TS	Nonverbal; severe ID; hypotonia	MFED
16 (von Spiczak et al., 2017)	M, 7	6	IS	AtS, GTCS	Nonverbal; severe ID; hypotonia	Hyps, SSW, FED
17 (von Spiczak et al., 2017)	M, 24	3	IS	AS, GTCS, FS	NVNA; profound ID; hypotonia	MFED, SSW, slow bg
18 (von Spiczak et al., 2017)	M, 3	5	IS	AS, ATs, GTCS	NVNA; profound ID; hypotonia	Hyps, MFED, GSW, slow bg
19 (von Spiczak et al., 2017)	M, 2 (slb of 21)	4	IS	MS, GTCS, FS	NVNA; profound ID; hypotonia	Hyps, MFED, FED, slow bg
20 (von Spiczak et al., 2017)	F, 5 (slb of 20)	N/A	IS	MS, TS, GTCS, FS	NVNA; profound ID; hypotonia	MFED, FED, slow bg
21 (von Spiczak et al., 2017)	M, 12	5	IS	AS, MS, ATs, GTCS	NVNA; profound ID; hypotonia	Hyps, MFED, SSW, GSW, GPFA, slow bg
22 (von Spiczak et al., 2017)	M, 19	8	N/A	GTCS	NVNA; profound ID; hypotonia	N/A
23 (von Spiczak et al., 2017)	M, 13	6	IS	MS, TS	Nonverbal; profound ID; hypotonia	Hyps, MFED, FED, slow bg
24 (von Spiczak et al., 2017)	M, 1	3	None	None	NVNA; profound ID; hypotonia	MFED, slow bg
25 (von Spiczak et al., 2017)	F, 1	1	IS	MS, TS, GTCS, FS, SE	NVNA; profound ID; hypotonia	Hyps, MFED, SSW, GSW, GPFA
26 (von Spiczak et al., 2017)	F, 5	4.5 years	N/A	MS, TS, GTCS, FS, rSE	NVNA; profound ID	GSW, slow bg
27 (von Spiczak et al., 2017)	M, 2	3	IS	MS	NVNA; hypotonia	Hyps, MFED, slow bg
28 (Kohikova et al., 2018)	F, 8	1 day	MS, TS	IS, MS	Nonverbal, unable to roll over or sit alone, profound ID	Abnormal bg, SSW over right hemisphere;
29 (Allen et al., 2013)	N/A	Infancy	N/A	N/A	N/A	N/A
30 (Lazzara et al., 2018)	M, 12	5 days	Neonatal seizures	FS, CE, SE	N/A; severe ID; hypotonia	Mild slow bg and MFED
31 (Brereton et al., 2018)	F, 8	N/A	N/A	N/A	Sit and walk without support; minimal speech; mild-moderate ID; hypotonia	Normal
32 (Brereton et al., 2018)	F, 8	N/A	N/A	Age 5, an episode of unresponsiveness and eye deviation to the left side, characterized by back arching and leaning to the left side	Sit and walk without support; limited speech; mild-moderate ID; hypotonia	Normal
The present case	F, 0.6	8	IS	IS	NVNA; severe ID; hypotonia	Sharp wave, MFED

DNM1, dynamin-1; F, female; M, male; DD, developmental delay; N/A, not available or applicable; IS, infantile spasms; ES, epileptic spasms; AS, absence seizures; FS, focal seizures; TS, tonic seizures; GTCS, generalized tonic-clonic seizures; CE, convulsive seizures; SE, status epilepticus; MS, myoclonic seizure; SZ, seizure; CLB, clobazam; VCB, vigabatrin; VPA, valproic acid; Hyps, hypsarrhythmia; KD, ketogenic diet; bg, background; GPFA, generalized paroxysmal fast activity; GSW, generalized spike-wave or poly spike-wave discharges; MFED, multifocal epileptiform discharges; FED, focal discharges; SSW, slow spike-wave discharges; CC, corpus callosum; MCM, mega cisterna magna; ID, intellectual disability; FL, frontal lobe; NVNA, nonverbal, non-ambulatory.

TABLE 2 | Details of the DNMI mutations variants of 33 cases.

Case Sequence (reference number)	Domain involved	Mutation	MutationTaster	Polyphen-2	SIFT	SIFT SCORE	rs. number
1 (Deng et al., 2016)	GTPase	c.443A > G (p.Gln148Arg)	Disease causing	Probably damaging	Damaging	0	NO rs
2 (Nakashima et al., 2016)	GTPase	c.127G > A (p.Gly43Ser)	Disease causing	Probably damaging	Damaging	0.001	NO rs
3 (Nakashima et al., 2016)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
4 (Appenzellar, 2014)	GTPase	c.194C > A (p.Thr65Asn)	Disease causing	Probably damaging	Damaging	0	NO rs
5 (Appenzellar, 2014)	GTPase	c.529G > C (p.Ala177Pro)	Disease causing	Probably damaging	Damaging	0.001	rs587777860
6 (Appenzellar, 2014)	GTPase	c.618G > C (p.Lys206Asn)	Disease causing	Probably damaging	Damaging	0	rs587777861
7 (Appenzellar, 2014)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs76020633
8 (Appenzellar, 2014)	Middle	c.1076G > C (p.Gly359Ala)	Disease causing	Probably damaging	Damaging	0.001	rs587777862
9 (Allen et al., 2016)	GTPase	c.865A > T (p.Ile289Phe)	Disease causing	Probably damaging	Damaging	0.01	NO rs
10 (von Spiczak et al., 2017)	GTPase	c.127G > A (p.Gly43Ser)	Disease causing	Probably damaging	Damaging	0	NO rs
11 (von Spiczak et al., 2017)	GTPase	c.134G > A (p.Ser45Asn)	Disease causing	Probably damaging	Damaging	0	NO rs
12 (von Spiczak et al., 2017)	GTPase	c.194C > A (p.Thr65Asn)	Disease causing	Probably damaging	Damaging	0	NO rs
13 (von Spiczak et al., 2017)	GTPase	c.416G > T (p.Gly139Val)	Disease causing	Probably damaging	Damaging	0	NO rs
14 (von Spiczak et al., 2017)	GTPase	c.529G > C (p.Ala177Pro)	Disease causing	Probably damaging	Damaging	0.001	rs587777860
15 (von Spiczak et al., 2017)	GTPase	c.616A > G (p.Lys206Glu)	Disease causing	Probably damaging	Damaging	0	NO rs
16 (von Spiczak et al., 2017)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
17 (von Spiczak et al., 2017)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
18 (von Spiczak et al., 2017)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
19 (von Spiczak et al., 2017)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
20 (von Spiczak et al., 2017)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
21 (von Spiczak et al., 2017)	GTPase	c.709C > T (p.Arg237Trp)	Disease causing	Probably damaging	Damaging	0	rs760270633
22 (von Spiczak et al., 2017)	GTPase	c.713G > A * (p.Ser238Ile)	N/A	N/A	Damaging	0	NO rs
23 (von Spiczak et al., 2017)	Middle	c.1037G > T (p.Gly346Val)	Disease causing	Probably damaging	Damaging	0	rs1064794903
24 (von Spiczak et al., 2017)	Middle	c.1075G > A (p.Gly359Arg)	Disease causing	Probably damaging	Damaging	0	NO rs
25 (von Spiczak et al., 2017)	Middle	c.1075G > A (p.Gly359Arg)	Disease causing	Probably damaging	Damaging	0	NO rs
26 (von Spiczak et al., 2017)	Middle	c.1117G > A (p.Glu373Lys)	Disease causing	Probably damaging	Damaging	0.046	NO rs
27 (von Spiczak et al., 2017)	Middle	c.1190G > A (p.Gly397Asp)	Disease causing	Probably damaging	Damaging	0.002	NO rs
28 (Kolnikova et al., 2018)	Middle	c.1089_1090insCTTCCA (p.Asn363_Arg364insLeuPro)	polymorphism	N/A	N/A		NO rs
29 (Allen et al., 2013)	Middle	c.1190G > A (p.Gly397Asp)	Disease causing	Probably damaging	Damaging	0.002	NO rs
30 (Lazzara et al., 2018)	Middle	c.796C > T (p.Arg266Cys)	Disease causing	Probably damaging	Damaging	0	rs138053929
31 (Brereton et al., 2018)	PH	c.1603A > G (p.Lys535Glu)	Disease causing	possibly damaging	Damaging	0.002	NO rs
32 (Brereton et al., 2018)	PH	c.1603A > G (p.Lys535Glu)	Disease causing	possibly damaging	Damaging	0.002	NO rs
Present case	GTPase	c.135C > A (p.Ser45Arg)	Disease causing	Probably damaging	Damaging	0.006	NO rs

Transcript ID: ENST00000372923.

*Transcription is unknown.

N/A, not available.

seizures, 13 (41.9%) patients had generalized tonic-clonic seizures, and 9 (29.0%) patients had focal seizures (**Table 1**). Fourteen out of 15 patients (93.3%) presented with infantile spasms initially, whereas 1 patient presented with myoclonic seizures, tonic seizures, generalized tonic-clonic seizures (GTCS), and focal seizures. Information was not available for one patient.

Development

All patients with *DNMI* mutation-related epileptic encephalopathy were nonverbal except for two patients, which were not mentioned in literature, with severe to profound intellectual disability. In 24 out of the 31 patients (77.4%), the developmental delay was apparent before seizure onset. Except for six patients who had normal development until the onset of refractory seizures, all patients had considerable developmental delays in the first year of life. Twenty-eight out of the 31 patients (90.3%) were non-ambulatory.

Response to Treatment

Seizure outcome was assessed in 31 patients: 24 out of 31 patients (77.4%) had refractory seizures. Three patients (9.7%)

became seizure-free post treatment. Patient 8 became seizure-free after placement on ketogenic diet at the age of 3.5 years while patient 6 had some response to ketogenic diet. Seizures in patients 1, 2, and 5 were controlled with valproic acid, clobazam, or vigabatrin over a period of 5 years between the ages 3 to 8 years.

In addition, patients 31 and 32, carrying pathogenic variants in the pleckstrin homology domain exhibited milder phenotypes without epilepsy. The two girls, 8 years old, were monozygotic triplet sisters who presented for evaluation of developmental delay, autism spectrum disorder, some dysmorphic features, and hypotonia without repeated seizures.

EEG Results

The EEG results of patients 31 and 32 were normal. EEG results were abnormal in 30 out of 31 (96.7%) patients with *DNMI* mutation-related epileptic encephalopathy. The EEG patterns of five patients reveal varied epileptiform discharges initiating as hypsarrhythmia, and evolving from slow generalized spike-wave discharges to paroxysmal fast activity. One patient exhibited non-specific background activity. Among



FIGURE 1 | A sample EEG recording of 4-month-old child in this study. The interictal EEG recording shows a sharp slow wave discharge in the left anterior temporal region.

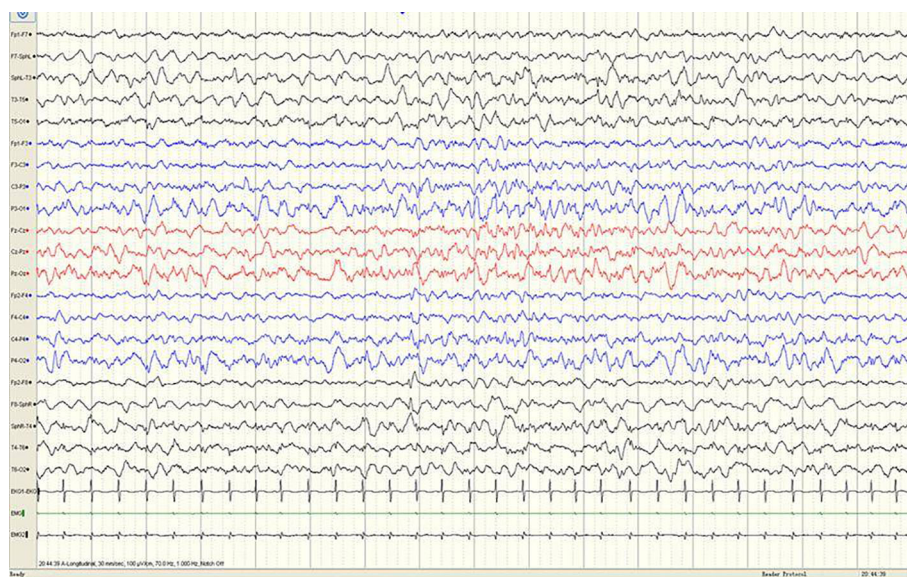


FIGURE 2 | A sample EEG recording of 4-month-old child in this study. The interictal EEG recording shows a sharp slow wave discharge in the right anterior temporal region.

the 31 patients, 15 (48.4%) had epileptiform discharge and background slowing. Twenty-eight (90.3%) patients had epileptiform discharge, of which multifocal discharge was the most common. There were 18 (58.1%) patients with multifocal discharge and 14 (45.2%) patients with hypsarrhythmia (Figure 3). Regarding other epileptiform discharges, there were nine (29.0%) patients with slow-spike wave complex, three (9.7%) patients with fast-wave activity, three (9.7%) patients

with extensive spike activity, and four (12.9%) patients with focal epileptiform discharge. In addition, there were two (6.5%) patients with multifocal discharge or hypsarrhythmia and slow-spike wave complex.

DNM1 Mutation Results

The current study reviewed data from 33 patients, including 31 sporadic patients and a sibling pair (patients 20 and 21), resulting

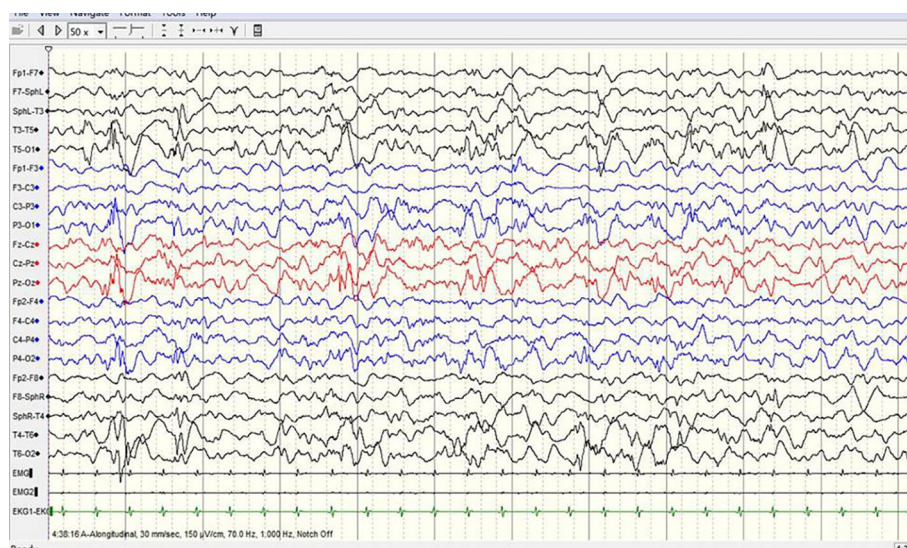


FIGURE 3 | A sample EEG recording of 8-month-old child in this study. The interictal EEG recording shows prime spike waves, a small number of multiple spike waves, spike-slow waves, and multiple spike-slow wave synchronous or non-synchronous discharge (atypical hypersarrhythmia) onto the bilateral rear head.

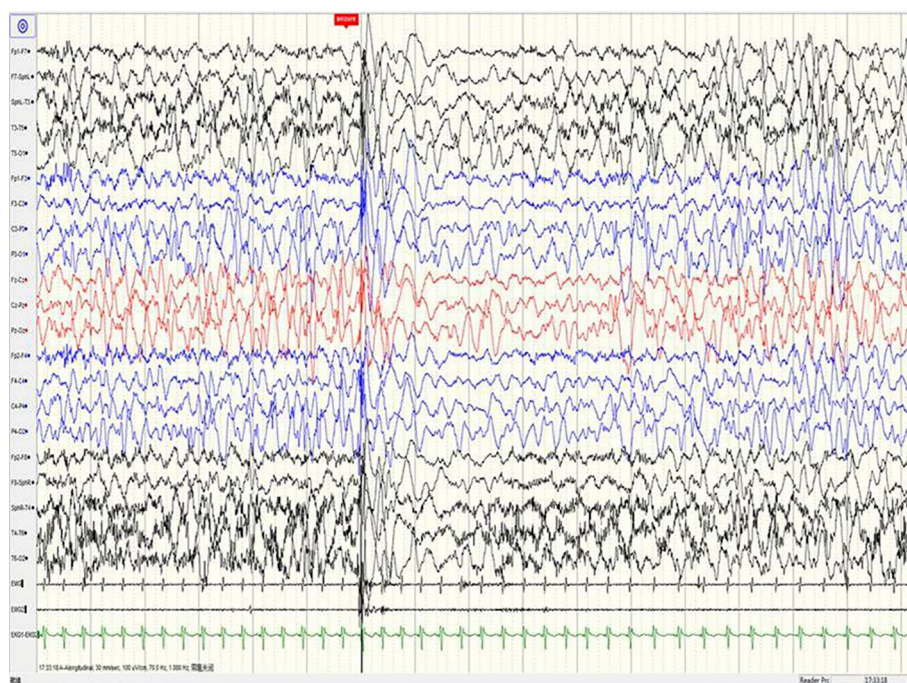


FIGURE 4 | A sample EEG recording of 8-month-old child in this study. The seizures are characterized by left oblique movement of both eyes and limb flexion uplift; the same period of the EEG signals show bilateral lead extensive spike-slow wave and myoelectric outbreak in the bilateral deltoid muscle.

in a total of 20 independent mutations (**Table 2**). The most common mutation was c.709C.T (p.Arg237Trp), which was found in 8 out of the 33 independent patients (24.2%). All mutations were confirmed to be *de novo*, except for the

affected sibling pair. Twelve of 20 mutations (60.0%), including the recurrent c.709C.T (p.Arg237Trp) mutation, occurred in the GTPase domain of *DNM1*. Seven out of the 20 mutations (35.0%) occurred in the middle domain of *DNM1*. One (5.0%) occurred

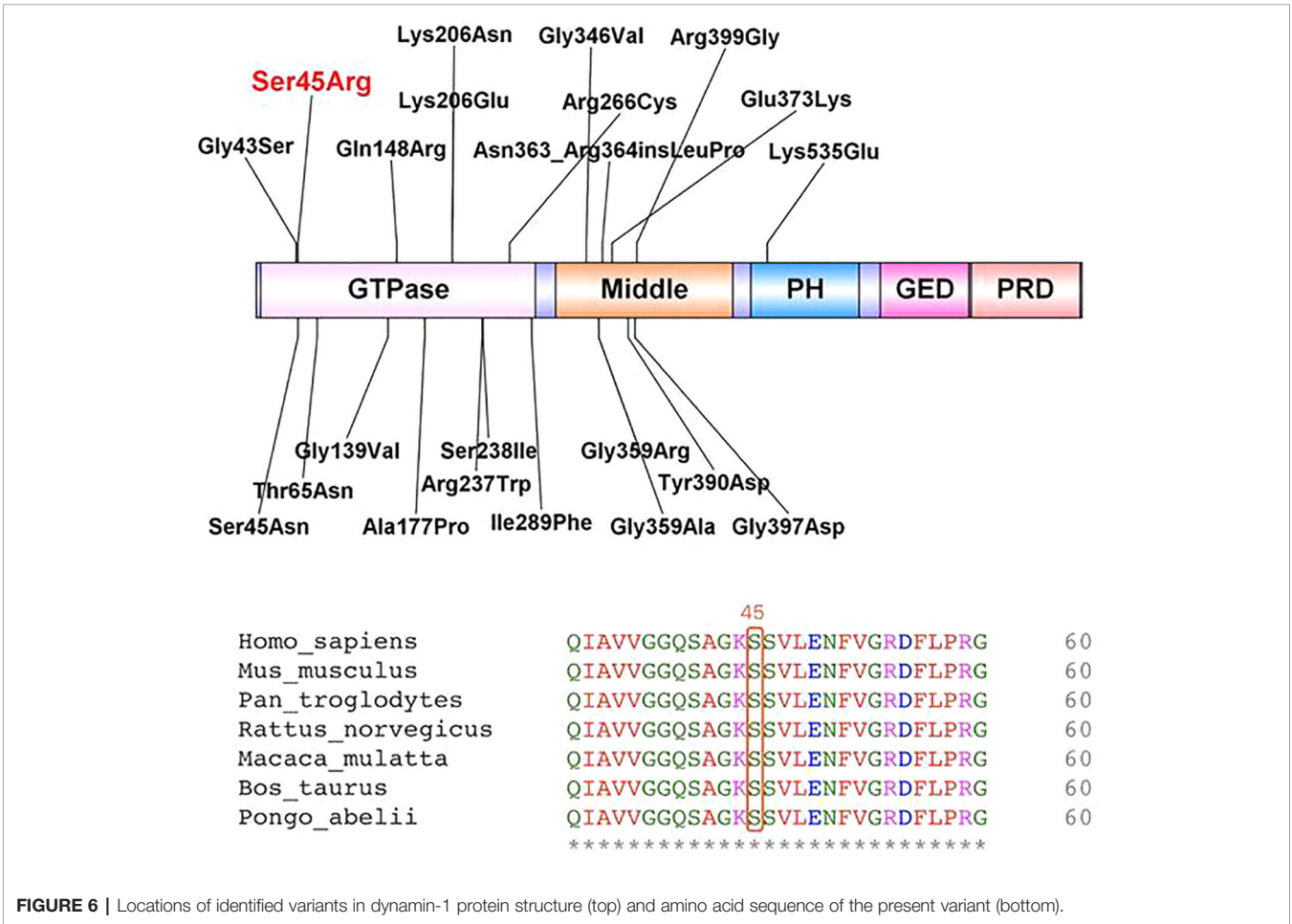
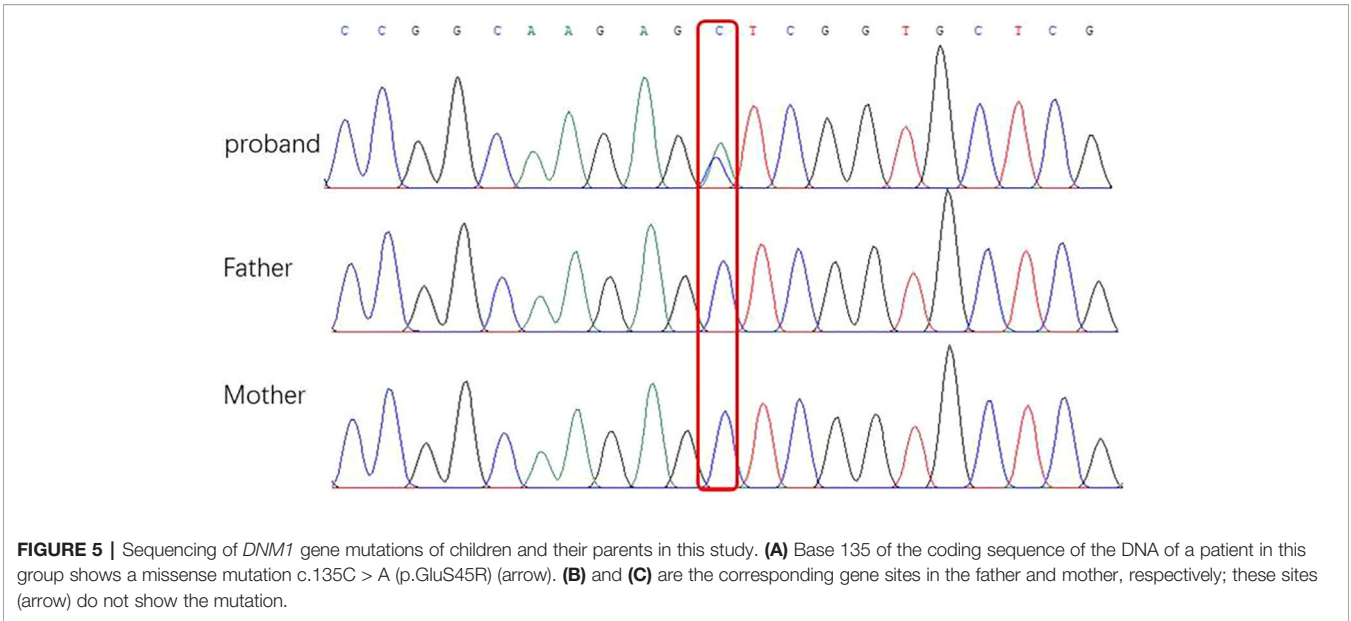


TABLE 3 | Comparison of gene domains and clinical features of 31 cases with mutation-related epileptic encephalopathy.

	Domain involved				χ^2	<i>P</i>
	GTPase		Middle			
	No.	Percent	No.	Percent		
Sex					0.285	0.5935
Female	6	27.27%	3	37.50%		
Male	16	72.73%	5	62.50%		
Age at seizure onset					1.829	0.4007
<6 months	9	42.86%	6	66.67%		
>12 months	2	9.52%	1	11.11%		
6–12 months	10	47.62%	2	22.22%		
Seizure type at onset					0.359	0.5491
IS/ES	15	78.95%	4	66.67%		
Other type	4	21.05%	2	33.33%		
Seizure outcome					—	0.1145
Intractable	20	95.24%	4	66.67%		
Seizure-free	1	4.76%	2	33.33%		
ID					0.353	0.5523
Profound	13	59.09%	5	71.43%		
Severe	9	40.91%	2	28.57%		

IS, infantile spasms; ES, epileptic spasms; ID, intellectual disability.

in the PH domain of *DNM1*. Twenty novel missense/frame-insertion mutations were predicted as pathogenic using the *in silico* prediction tools Mutation Taster Server, Polyphen-2, and SIFT.

Comparison of Genetic and Clinical Phenotypes of Children With *DNM1* Mutation-Related Encephalopathy

As shown in **Table 3**, sex (female vs. male, $P = 0.5935$), age at seizure onset (< 6 months vs. > 12 months vs. 6–12 months, $P = 0.4007$), seizure type at onset (infantile/epileptic spasms vs. other type, $P = 0.5491$), seizure outcome (intractable vs. seizure-free, $P = 0.1145$), and intellectual disability (profound vs. severe, $P = 0.5523$) showed no significant associations with the GTPase or middle domains.

DISCUSSION

Neurotransmission in the central system relies on synaptic vesicle transport. *DNM1* is a protein involved in the synaptic vesicle cycle, which facilitates the exocytosis of neurotransmitters necessary for normal signaling pathways and development in the central nervous system. Dynamin proteins have five domains; the GTPase domain is the largest and best understood, followed by a middle domain, a pleckstrin homology domain, a GTPase effector domain, and a proline-rich domain (McNiven et al., 2000). The pleckstrin homology domain is thought to interact directly with the lipid bilayer. The *DNM1* gene is mainly expressed in the central nervous system (Romeu and Arola, 2014), which explains the neurological phenotypes in *DNM1*-related disorders. Next-generation sequencing has been rapidly

implemented into routine clinical practice, where it has improved the diagnostic rate of patients with neuromuscular diseases. The widespread application of next-generation sequencing has greatly facilitated the understanding of the underlying mechanisms of epileptic encephalopathy (Fang et al., 2017; Ni and Shi, 2017). Previous publications have characterized the functional consequences of *DNM1* mutations and found that the seizure phenotype is largely due to the deleterious effects of *DNM1* mutations in GABAergic interneurons, while behavioral locomotor phenotypes may be due to the effect of the mutation in pyramidal cells (Asinof et al., 2015; Asinof et al., 2016).

Based on the collected mutation pattern and clinical information, we analyzed the relationship between genotypes and phenotypes. Previous research interpreted the molecular mechanisms of *DNM1* mutations and inferred the connection between genotypes and phenotypes to certain extent. It has been reported that mutations in different domains lead to distinct clinical phenotypes. Patients carrying pathogenic variants in the GTPase or middle domains present with epileptic encephalopathy and severe neurodevelopmental symptoms. These mutations have been reported in association with early onset epileptic encephalopathy (Appenzellar, 2014), intractable seizures (seizure onset in *DNM1* patients ranges from 2–13 months of and usually presents with infantile spasms. The seizure type manifests in various forms as the patient ages, ranging from absence seizures to generalized tonic-clonic seizures.), motor impairments, and severe to profound intellectual disability. In this study, 24 patients (77.4%) had refractory seizures. During the course of the disease, 23 (74.2%) patients initially had spasm seizures; all patients had severe to profound intellectual disability and considerable developmental delay in the first year of life. Other clinical features reported in some affected individuals included hypotonia, developmental regression, movement disorder, autism, cortical visual impairment, behavioral concerns, and microcephaly. Patients carrying pathogenic variants in both domains exhibited comparable phenotypes (**Table 3**), although the mechanism of protein disruption was distinct from that of patients with variants in a single domain. Most variants in the GTPase domain were predicted to impair hydrolysis of GTP, but not its binding to the synaptic vesicle; this resulted in integrated oligomeric assembly and impaired vesicle scission. However, middle domain variants were predicted to impair the ability of the *DNM1* protein to form larger oligomeric assemblies. In the case of the patients, the dominant negative effect of *DNM1* results in a generally similar overall phenotype suggesting a similar pathway (von Spiczak et al., 2017). Patients carrying pathogenic variants in the pleckstrin homology domain exhibited milder phenotypes without epilepsy. These patients were 8-year-old identical twin sisters who had no seizures and exhibited mild-to-moderate developmental delay/intellectual disability and autism spectrum disorder (Brereton et al., 2018). The *de novo* p.Lys535Glu mutation is a likely pathogenic novel variant in exon 15 of *DNM1*. However, this was reported in a single patient without an epilepsy phenotype. Therefore, reports

of additional patients are needed to define the relationship between genotype and phenotype in *DNM1* mutation-related epileptic encephalopathy (Jia and Shi, 2017).

EEG is an important tool for assessment of the diagnosis and prognosis of epileptic encephalopathy in patients carrying *DNM1* mutations. The EEG patterns are consistent with changes in the electrical activities of the brain in patients with infantile epileptic encephalopathy. The EEG patterns reveal varied epileptiform discharges initiating as hypsarrhythmia and evolving from slow generalized spike-wave discharges to paroxysmal fast activity. In this study of epileptic encephalopathies in 31 patients carrying *DNM1* mutations (Table 1), approximately 96.7% of patients' recordings portrayed abnormal EEG; multifocal discharge was most common (58.1%), followed by hypsarrhythmia (45.2%). Other epileptiform discharges were characterized by slow-spike and slow-wave complex, fast-wave activity, extensive spike activity, and focal epileptic discharge. The results of this study were consistent with those of a retrospective study published in 2017. The specific EEG pattern remains the basis for the diagnosis. Series EEG with video and video EEG with electromyogram electrodes is also recommended. The association of characteristic multiple seizure types and intellectual disability represents the classic hallmark of Lennox-Gastaut syndrome. This diagnostic triad may not be completely present at the onset of seizures; therefore, an accurate diagnosis of Lennox-Gastaut syndrome often requires further disease development over time (Markand, 2003; Arzimanoglou et al., 2009; Camfield, 2011; Bourgeois et al., 2014). The patient in the present case underwent a series of video EEG monitoring, which initially showed focal discharge, followed by atypical hypsarrhythmia and infantile spasms; this suggested evolution of the disease and provided clues for diagnosis and treatment.

The long-term outcomes of patients with *DNM1* mutation-related epileptic encephalopathy were often disappointing. The choice of AEDs at the onset of seizures was tailored to seizure type, clinical presentation, and EEG pattern. Thus far, there are no international guidelines for the pharmacological treatment of *DNM1* mutation-related epileptic encephalopathy because of the limited efficacy of antiepileptic medications. In the present study, three patients had been given sodium valproate and became seizure-free; however, as the disease progressed, they developed drug-refractory epilepsy. Therefore, we analyzed differences in the response of the same gene mutation to drug treatment. First, we speculated that the choice of treatment time or the natural process of disease might influence the response. Then, we investigated whether the type and site distributions of *DNM1* gene mutations were associated with clinical phenotype, potentially providing clues for clinical diagnosis and treatment. Eight patients carried the p.Arg237Trp mutation (Tables 1 and 2). Given that *DNM1* mutations are present in up to 2% of patients with severe epilepsy (Kolnikova et al., 2018), this mutation is particularly frequent in patients with epileptic encephalopathy. The relatively homogeneous phenotype and predicted dominant-negative mechanism of this mutation

make *DNM1*-associated encephalopathy has the potential of being an effective therapeutic target. Gene therapy might also be an effective means to restore *DNM1* function (Kolnikova et al., 2018). We presume that treatment methods and strategies will be further refined with additional studies involving more patients and investigations into the molecular basis of the disease.

In conclusion, to the best of our knowledge, this is the first integrated analysis of the phenotypic, genetic, and electroencephalographic features of children with *DNM1* mutation-related encephalopathy. Our study highlighted the role of series EEG and video EEG of children with *DNM1* mutation-related encephalopathy; EEG patterns may aid in providing clues for treatment.

There were several limitations to this study, such as the fact that it is a retrospective and summary study. Due to the small number of cases, we have not come to a definite conclusion; the pathogenic variant in this study needs to be confirmed by functional experiments. To determine the association of phenotype and genotype of children with *DNM1* mutation-related encephalopathy, further analysis of additional patients is needed.

ETHICS STATEMENT

The present study was approved by the Ethics Committee of Beijing Children's Hospital and the patient have gave written informed consent.

AUTHOR CONTRIBUTIONS

All authors contributed to the study design, critically reviewed the manuscript, and approved the final version. HL performed literature search and analysis, and wrote the manuscript. MX, ZL, JZ, XiaohW, XiaofW, and TH performed literature search and analysis. FF revised the manuscript.

FUNDING

This work was supported by National Natural Science Foundation of China(81541115), the Capital Health Research and Development Fund(2018-2-2096) and Beijing Municipal Administration of Hospitals Incubating Program(PX2017065).

ACKNOWLEDGMENTS

We thank Editage (<http://editage.com/frontiers/>) for editing a draft of this manuscript. We also acknowledge the financial support of the Open Access Publication Fund of Beijing Children's Hospital.

REFERENCES

- Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N., Dlugos, D., Eichler, E. E., et al. (2013). De novo mutations in epileptic encephalopathies. *Nature* 501, 217–221. doi: 10.1038/nature12439
- Allen, N. M., Conroy, J., Shahwan, A., Lynch, B., Correa, R. G., Pena, S. D., et al. (2016). Unexplained early onset epileptic encephalopathy: exome screening and phenotype expansion. *Epilepsia* 57, e12–e17. doi: 10.1111/epi.13250
- Appenzeller (2014). De novo mutations in synaptic transmission genes including *DNM1* cause epileptic encephalopathies. *Am. J. Hum. Genet.* 95, 360–370. doi: 10.1016/j.ajhg.2014.08.013
- Arzimanoglou, A., French, J., Blume, W. T., Cross, J. H., Ernst, J. P., Feucht, M., et al. (2009). Lennox-Gastaut syndrome: a consensus approach on diagnosis, assessment, management, and trial methodology. *Lancet Neurol.* 8, 82–93. doi: 10.1016/S1474-4422(08)70292-8
- Asinof, S. K., Sukoff Rizzo, S. J., Buckley, A. R., Beyer, B. J., Letts, V. A., Frankel, W. N., et al. (2015). Independent neuronal origin of seizures and behavioral comorbidities in an animal model of a severe childhood genetic epileptic encephalopathy. *PLoS Genet.* 11, e1005347. doi: 10.1371/journal.pgen.1005347
- Asinof, S., Mahaffey, C., Beyer, B., Frankel, W. N., and Boumil, R. (2016). Dynamin 1 isoform roles in a mouse model of severe childhood epileptic encephalopathy. *Neurobiol. Dis.* 95, 1–11. doi: 10.1016/j.nbd.2016.06.014
- Belousova, E. D., Zavadenko, N. N., Kholin, A. A., and Sharkov, A. A. (2017). New classifications of epilepsies and seizure types created by the International League against Epilepsy (2017). *Zh. Nevrol. Psikiatr. Im. S. S. Korsakova* 117, 99–106. doi: 10.17116/jnevro20171177199-106
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Bourgeois, B. F., Douglass, L. M., and Sankar, R. (2014). Lennox-Gastaut syndrome: a consensus approach to differential diagnosis. *Epilepsia* 55 Suppl 4, 4–9. doi: 10.1111/epi.12567
- Brereton, E., Fassi, E., Araujo, G. C., Dodd, J., Telegrafi, A., Pathak, S. J., et al. (2018). Mutations in the PH Domain of *DNM1* are associated with a nonepileptic phenotype characterized by developmental delay and neurobehavioral abnormalities. *Mol. Genet. Genomic Med.* 6, 294–300. doi: 10.1002/mgg3.362
- Camfield, P. R. (2011). Definition and natural history of Lennox-Gastaut syndrome. *Epilepsia* 52 Suppl 5, 3–9. doi: 10.1111/j.1528-1167.2011.03177.x
- Chappie, J. S., Acharya, S., Leonard, M., Schmid, S. L., and Dyda, F. (2010). G domain dimerization controls dynamin's assembly-stimulated GTPase activity. *Nature* 465, 435–440. doi: 10.1038/nature09032
- Deng, X. L., Yin, F., Zhang, C. L., Ma, Y. P., He, F., Wu, L. W., et al. (2016). [Dynamin-1-related infantile spasms: a case report and review of literature]. *Zhonghua Er. Ke. Za. Zhi.* 54, 856–859. doi: 10.3760/cma.j.issn.0578-1310.2016.11.014
- Fang, F., Liu, Z., Fang, H., Wu, J., Shen, D., Sun, S., et al. (2017). The clinical and genetic characteristics in children with mitochondrial disease in China. *Sci. China Life Sci.* 60, 746–757. doi: 10.1007/s11427-017-9080-y
- Farwell, K. D., Shahmirzadi, L., El-Khechen, D., Powis, Z., Chao, E. C., Tippin, D. B., et al. (2015). Enhanced utility of family-centered diagnostic exome sequencing with inheritance model-based analysis: results from 500 unselected families with undiagnosed genetic conditions. *Genet. Med.* 17, 578–586. doi: 10.1038/gim.2014.154
- Jia, J., and Shi, T. (2017). Towards efficiency in rare disease research: what is distinctive and important? *Sci. China Life Sci.* 60, 686–691. doi: 10.1007/s11427-017-9099-3
- Kolnikova, M., Skopkova, M., Ilencikova, D., Foltan, T., Payerova, J., Danis, D., et al. (2018). *DNM1* encephalopathy—atypical phenotype with hypomyelination due to a novel de novo variant in the *DNM1* gene. *Seizure* 56, 31–33. doi: 10.1016/j.seizure.2018.01.020
- Lazzara, A., Asghar, S., Zacharia, T., and Byler, D. (2018). *DNM1* Mutation in a child associated with progressive bilateral mesial temporal sclerosis. *Clin. Case Rep.* 6, 2037–2039. doi: 10.1002/ccr3.1793
- Markand, O. N. (2003). Lennox-Gastaut syndrome (childhood epileptic encephalopathy). *J. Clin. Neurophysiol.* 20, 426–441. doi: 10.1097/00004691-200311000-00005
- McNiven, M. A., Cao, H., Pitts, K. R., and Yoon, Y. (2000). The dynamin family of mechanoenzymes: pinching in new places. *Trends Biochem. Sci.* 25, 115–120. doi: 10.1016/S0968-0004(99)01538-8
- Nakashima, M., Kouga, T., Lourenco, C. M., Shiina, M., Goto, T., Tsurusaki, Y., et al. (2016). De novo *DNM1* mutations in two cases of epileptic encephalopathy. *Epilepsia* 57, e18–e23. doi: 10.1111/epi.13257
- Ni, X., and Shi, T. (2017). The challenge and promise of rare disease diagnosis in China. *Sci. China Life Sci.* 60, 681–685. doi: 10.1007/s11427-017-9100-1
- Reubold, T. F., Faelber, K., Plattner, N., Posor, Y., Ketel, K., Curth, U., et al. (2015). Crystal structure of the dynamin tetramer. *Nature* 525, 404–408. doi: 10.1038/nature14880
- Romeu, A., and Arola, L. (2014). Classical dynamin *DNM1* and *DNM3* genes attain maximum expression in the normal human central nervous system. *BMC Res. Notes* 7, 188. doi: 10.1186/1756-0500-7-188
- von Spiczak, S., Helbig, K. L., Shinde, D. N., Huether, R., Pendziwiat, M., Lourenco, C., et al. (2017). *DNM1* encephalopathy: a new disease of vesicle fission. *Neurology* 89, 385–394. doi: 10.1212/WNL.0000000000004152

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Fang, Xu, Liu, Zhou, Wang, Wang and Han. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Neurologic Manifestations as Initial Clinical Presentation of Familial Hemophagocytic Lymphohistiocytosis Type2 Due to *PRF1* Mutation in Chinese Pediatric Patients

Wei-xing Feng, Xin-ying Yang, Jiu-wei Li, Shuai Gong, Yun Wu, Wei-hua Zhang, Tong-li Han, Xiu-wei Zhuo, Chang-hong Ding and Fang Fang*

Neurology Department, National Center for Children's Health China, Beijing Children Hospital affiliated to Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:

Ruth Angela Roberts,
Apconix, United Kingdom

Reviewed by:

Fan Jin,
Zhejiang University, China
Yifan Zhang,
University of Arkansas at Little Rock,
United States

*Correspondence:

Fang Fang
13910150389@163.com

Specialty section:

This article was submitted to
Genetic Disorders,
a section of the journal
Frontiers in Genetics

Received: 07 March 2019

Accepted: 03 February 2020

Published: 04 March 2020

Citation:

Feng W-x, Yang X-y, Li J-w, Gong S,
Wu Y, Zhang W-h, Han T-l, Zhuo X-w,
Ding C-h and Fang F (2020)
Neurologic Manifestations as Initial
Clinical Presentation of Familial
Hemophagocytic Lymphohistiocytosis
Type2 Due to *PRF1* Mutation in
Chinese Pediatric Patients.
Front. Genet. 11:126.
doi: 10.3389/fgene.2020.00126

Familial hemophagocytic lymphohistiocytosis Type 2 (FHL2) associated central nervous system (CNS) involvement is less understood in children, especially when considering neurologic manifestations as part of the initial presentation. We conducted a retrospective review of the clinical manifestations and genetic abnormality of four Han Chinese children with FHL2 who were patients at the neurology department of Beijing Children's Hospital from November 2015 to October 2018. These four patients initially manifested CNS symptoms in their disease presentation, and all four patients were misdiagnosed as having demyelinating disease, such as acute disseminated encephalomyelitis and multiple sclerosis. Given these misdiagnoses, it is important that general physicians and pediatricians maintain awareness of the possibility of FHL2 as a differential diagnosis. These four cases included neurologic manifestations including seizures, ataxia, spasticity, gait disorder, and coma. Bilateral abnormal signals in the cerebrum, including in white matter, gray matter, and junctions were discovered. Enhanced magnetic resonance imaging (MRI) in these patients showed spot or ring enhancement and/or hemorrhage. These patients all possessed a compound heterozygote mutation *PRF1* gene. Whole exome sequencing analysis revealed seven different mutations (three novel mutations) spread over the *PRF1* gene and a heterozygous missense mutation c.1349C > T [p.T450M] that was present in two patients. Three novel mutations, c.634T > C [p.Y212H], c.1083_1094del [p.361_364del], and c.1306G > T [p.D436Y], were discovered and through *in silico* analysis were discovered to be deleterious. Neurologic manifestations were the initial symptoms of FHL2 in these patients in addition to the expected leukopenia and hepatosplenomegaly. Whole exome sequencing of *PRF1* for patients with similar presentations would facilitate prompt and accurate diagnosis and treatment.

Keywords: neurological manifestations, familial hemophagocytic lymphohistiocytosis Type-2, pediatric patients, perforin 1, mutation

BACKGROUND

FHL is a severe clinical condition that typically presents with fever, pancytopenia, hepatomegaly, and/or splenomegaly, which can progress to hypertriglyceridemia, hypofibrinogenemia, hepatitis, and/or neurological manifestations (Osinska et al., 2014). Known pathological gene mutations and their associated familial HLH (FHL) subtypes include *PRF1* (FHL2), *UNC13D* (FHL3), *STX11* (FHL4), and *STXBP2* (FHL5) (Dufourcq-Lagelouse et al., 1999; Feldmann et al., 2003). Perforin is the protein product of *PRF1*, and it affects cellular cytotoxicity mechanisms.

Decreased production or activity level of perforin can result in impaired immune defense systems and dysregulation of the apoptotic mechanisms. FHL2 is believed to be invariably fatal during infancy or early childhood unless hematopoietic stem cell transplantation (HSCT) is performed. CNS involvement can be apparent at initial presentation, or it can occur at any time during the course of FHL2. A Chinese single-center study also indicated that twelve patients (13%) had neurological symptoms, including seizures, ataxia, coma, cranial nerve palsy, and hemiplegia (Yang et al., 2010). While the characteristics of FHL2 are well known, FHL2-associated CNS initial involvement is less understood.

We are reporting on four FHL2 cases in which neurological changes appeared as early clinical symptoms. These subjects were inpatients in our neurology department and initially

displayed CNS symptoms but had no effects of other systems. They were misdiagnosed as having CNS demyelination prior to having clinical, radiological, and cerebrospinal fluid (CSF) cytology data analyzed. The purpose of this report is to summarize the clinical and genetic characteristics of these cases.

CASE PRESENTATION

We conducted a retrospective review of the clinical manifestations and genetics of four Han Chinese patients with FHL2 at the neurology department of Beijing Children's Hospital from November 2015 to October 2018. The clinical findings and mutations of the cases are illustrated and listed in **Table 1**.

Case 1

In Case 1, an 18-month-old female was admitted to our institution for intermittent fever, weakness, and ataxia over a 15 day period. Brain MRI was performed and multiple irregular lesions were observed in both hemispheres, basal ganglia, and the cerebellum. The focal lesions revealed spot or ring enhancements. The patient developed signs of CNS involvement including seizures and left peripheral facial paralysis. A repeat MRI indicated enlargement of multiple lesions along with an increased number of lesions, and a more defined boundary of abnormal signals was noted in the

TABLE 1 | Clinical Findings and Mutations of *PRF1* in cases.

Patients	Case1	Case2	Case3	Case4
Gender	Female	Female	Female	Male
Age onset	1y 6m	4y 11m	9y	1y 9m
Age diagnosis	12m	14m	37m	2m
Family history	No	Yes	No	No
Clinical presentation				
Fever	Yes	Yes	Yes	No
Convulsion	Yes	Yes	Yes	Yes
Splenomegaly	Yes	Yes	No	No
Hepatomegaly	Yes	Yes	No	Yes
Neurological symptoms	gait disorder; seizure and facial paralysis; encephalopathy	headache; seizure; disturbance of consciousness; facial paralysis; ataxia; slurred speech	headache and vomiting and convulsions; ataxia and slurred speech; spasticity; coma	Gait disturbances; convulsion; irritability; somnolence
Laboratory investigations				
CSF cytology	$10 \times 10^6/L$	$6 \times 10^6/L$	$8 \times 10^6/L$; $50 \times 10^6/L$	$4 \times 10^6/L$
CSF Protein	850 mg/L	normal	530 mg/L; 11200mg/L	684 mg/L
Leukocyte	$2.68(10^9/L)$	$3.41(10^9/L)$	$1.55(10^9/L)$	$3.42(10^9/L)$
Neutrophils	$0.46(10^9/L)$	$0.67(10^9/L)$	$0.97(10^9/L)$	$0.27(10^9/L)$
Erythrocyte	normal	$3.21(10^{12}/L)$	normal	normal
Thrombocytes	normal	$68(10^9/L)$	normal	normal
Fibrinogen	normal	normal	normal	normal
Triglyceride	normal	normal	normal	normal
Hemophagocytosis in bone marrow smear	No	No	No	No
Ferritin	—	229.4 ng/ml	485.8 ng/ml	—
EBV	negative	negative	CSF EB-IgM positive	negative
<i>PRF1</i> paternal origin	c.634T > C	c.1349C > T	c.1349C > T	c.65delC
Protein change	p.Y212H ^a	p.T450M	p.T450M	p.P22Rfs*2
<i>PRF1</i> maternal origin	c.1083_1094del	c. 853_855del	c.1306G > T	c.148G > A
Protein change	p.361-364del ^a	p.285delK	p.D436Y ^a	p.V50M

1. "—" indicated no data.

2. ^a indicated the novel mutation.

cerebellum, even though intravenous immunoglobulin (IVIG) and corticosteroids were administered (**Figures 1** and **2**). Abnormalities noted on brain imaging appeared to be roughly proportional to the severity of the clinical manifestations. The laboratory results revealed leukopenia and leukocyte levels at $2.68(10^9/L)$ with an absolute neutrophil count of $0.46(10^9/L)$. Her condition worsened to include encephalopathy and convulsions. On day 31, she was released from our hospital and admitted to a local hospital. The MRI showed serious encephalatrophy. The girl died from multisystem organ failure 5 months later. About 6 months after Case 1's death, we diagnosed Case 2, and we observed clinical similarities. We obtained consent from Case 1's parents and

proceeded with whole exome sequencing. The presence of both the c.634T > C[p.Y212H] mutation (a novel mutation of paternal origin) and the c.1083_1094del[p.361_364del] mutation (a novel mutation of maternal origin) confirmed a compound heterozygous state in the subject.

The 2 novel mutations (p.Y212H and p.361-364del) were found to be deleterious *via in silico* analysis by SIFT, PolyPhen and Mutation_Taster software programs. SIFT (<http://blocks.fhcrc.org/sift/SIFT.html>), which uses evolutionary information from homologous proteins, showed damage, providing a SIFT score 0 (Deleterious sift ≤ 0.05). The PolyPhen tool (<http://www.bork.embl-heidelberg.de/PolyPhen/>), which incorporates

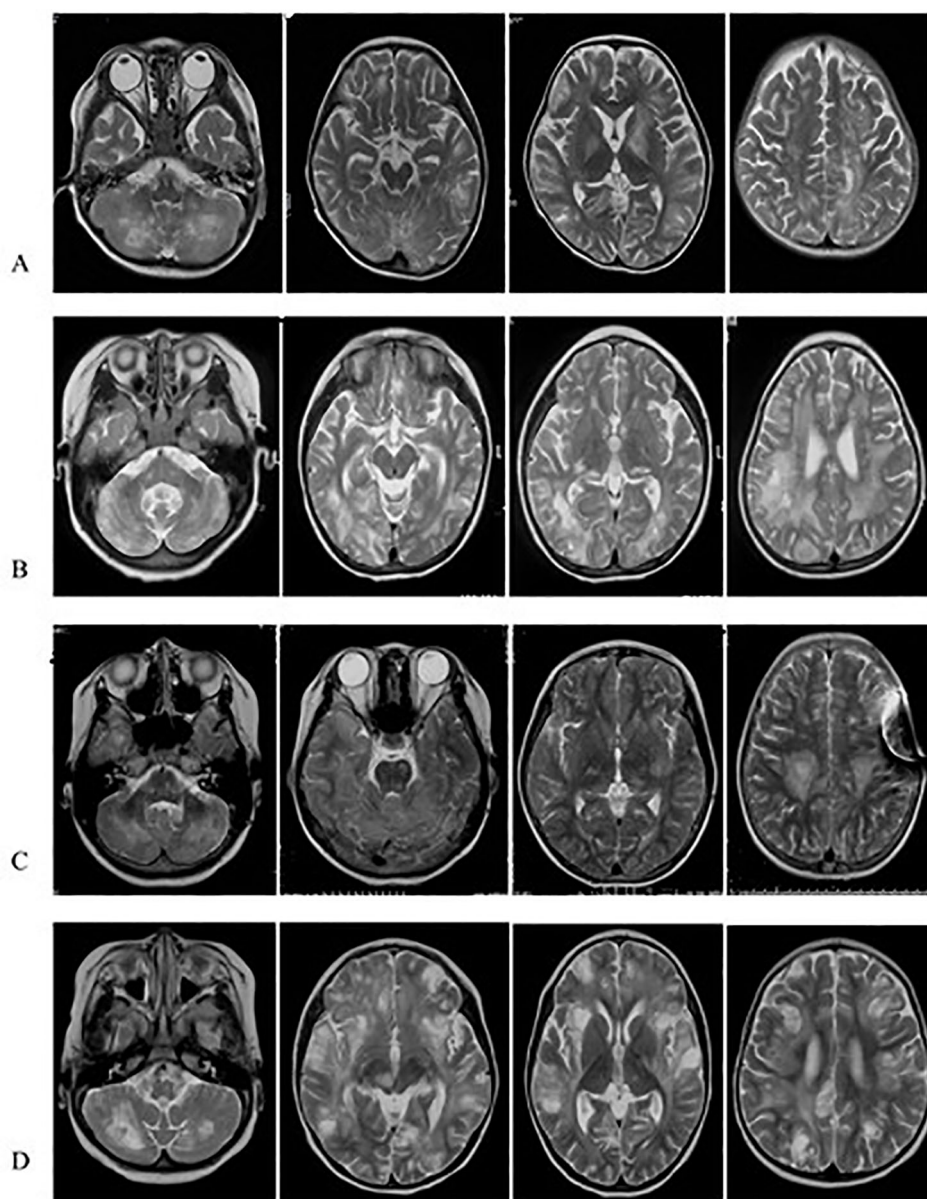


FIGURE 1 | TR-MRIs of Case1 (**A**), Case2 (**B**), Case3 (**C**) and Case4 (**D**). The brain T2-MRIs revealed numerous irregular shape hypersignal areas involving cerebral hemispheres, cerebellar hemisphere, basal ganglia, mostly located in the corticomedullary junction and deep white matter.

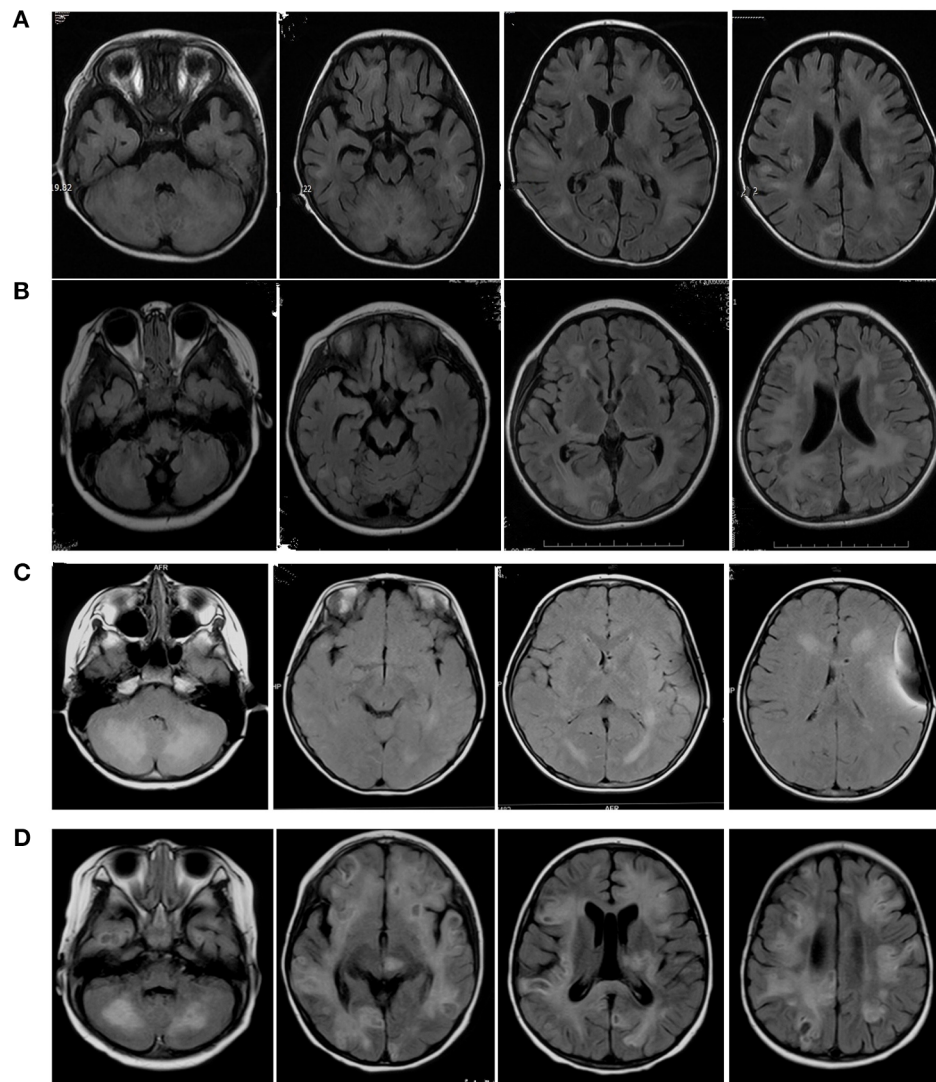


FIGURE 2 | Axial Flair sequences of brain MRIs of Case1 (A), Case2 (B), Case3 (C), Case4 (D). The FLAIR-MRI revealed numerous variable size in the cerebral hemisphere, basal ganglia, cerebellum and brain stem as well as hypersignal intensities on T2-MRI.

structural information into classification rules, showed probable damage with score of 1 (Probably damaging ≥ 0.909). The Mutation_Taster (<http://www.mutationtaster.org/>) prediction indicated “disease causing”. The position of the p.Y212H residue was highly conserved among different species (Appendix 1A). The red arrow indicates the difference in the p361-364del PRF1 protein in the three-dimensional model as predicted by the SWISS MODEL (<https://swissmodel.expasy.org/>) (Appendix 1B).

Case 2

Case 2 was a 4-year-11-month-old female who was admitted to the hospital for symptoms including fever, headache, seizure, and disturbance of consciousness for two days. Primary MRI findings included multiple bilateral abnormal signals in the cerebellar

hemispheres and cerebellum, in the posterior extremity of the right inner capsule, in the right brachium pontis, and in the brain stem. Diffusion-weighted imaging (DWI) revealed restricted diffusion in the lesions. Three months later, additional MRI imaging showed scattered, patchy, nodular, and enhanced bilateral abnormal lesions in the cerebrum, cortex, subcortex, periventricular area, basal ganglia, thalamus, midbrain, cerebellum, and pons, in addition to a possible slight hemorrhage without clinical symptoms. Susceptibility-weighted imaging (SWI) showed multifocal low signals. About 6 months after the first hospitalization, she experienced facial paralysis, ataxia, irritability, and slurred speech. Another MRI of the brain indicated enlargement of the lesions (Figures 1 and 2). The patient was diagnosed with multiple sclerosis (MS) and

prescribed methylprednisolone and IVIG for treatment. The patient then suffered from facial paralysis and gait disturbance. Subsequent brain MRI indicated lesion enlargement and progression. Fourteen months after her first hospitalization, the patient's younger brother developed leukopenia. The family history included a younger brother with cytopenia of two lineages (platelet and erythrocyte) and hepatosplenomegaly. This information led us to perform whole exome sequencing, which revealed that the patient and her brother carried the same mutations, in *PRF1*, c.1349C > T [p.T450M] (of paternal origin) and c.853_855del [p.285del] (of maternal origin). These two mutations were responsible for FHL2, and the patient and her brother are receiving chemotherapy, awaiting HSCT.

Case 3

Case 3 was a 12-year-old female admitted to our hospital due to intermittent headache, vomiting, convulsions, ataxia, and slurred speech for three years. Three years prior she was admitted to a local hospital with vomiting and headache. A brain MRI confirmed abnormal lesions in the bilateral cerebrum and cerebellar cortex, abnormal signals in white matter lesions of the cerebellar hemispheres, ventriculomegaly, and cerebellar tonsil herniation. CSF total protein was increased (530 mg/L) and CSF pressure was 300 mmH₂O. Cell count analysis was normal. The patient was treated for cerebral hernia with external ventricle drainage. The follow-up treatment was ventricle-peritoneal (V-P) shunt operation. During her admission 3 years later, a brain MRI demonstrated new bilateral lesions in the cerebral hemisphere, basal ganglia region, dorsal thalamus, and brainstem (**Figures 1 and 2**). IVIG and corticosteroids were administered to treat MS; however, the patient experienced insufficient symptom relief after treatment. On admission to our hospital, the patient suffered from headache, high fever, and dysphoria. The CSF leukocyte reading was 50×10^6 /L and total CSF protein was extremely increased (11,200 mg/L). CSF glucose was low (1.64 mmol/L) and CSF EB-IgM was positive. On the 4th day, her condition deteriorated to a coma with convulsions and respiratory failure. The patient then regained consciousness after several days. Given her three year history and gradual progression, we considered HLH for the differential diagnosis and proceeded with whole exome sequencing. The genetic testing indicated that the patient had two point mutations in *PRF1*, c.1349C > T [p.T450M] (of paternal origin) and c.1306G > T [p.D436Y] (a novel mutation of maternal origin), which was found to be deleterious using *in silico* analysis (these were the same scores seen in Case 1). The p.D436Y position residue was highly conserved among different species.

Five months later the patient could speak but was still unable to walk. A subsequent brain MRI did not indicate any improvement. The patient was receiving chemotherapy in the hematology department of the local hospital.

Case 4

Case 4 was 21-month-old male who was admitted to our hospital with gait disturbances for 2 weeks. Brain MRI indicated abnormally long T1 and T2 signals in the bilateral left basal

ganglia, thalamus, brainstem, and corpus callosum (**Figures 1 and 2**). DWI revealed restricted diffusion within the brain lesions. Enhanced MRI showed scattered, enhanced, and abnormal lesions in the bilateral cerebrum and cerebellum, left basal ganglia, thalamus, brainstem, and corpus callosum. The patient was diagnosed with CNS demyelination, and treated with corticosteroid and IVIG. During the patient's admission, he suffered from convulsions, irritability, and somnolence. A computed tomography (CT) brain scan displayed lamellar low density bilaterally in the white matter of the cerebrum and cerebellum. Hemorrhagic high density nodules were observed in the temporal and frontal lobes, and mannitol and oxcarbazepine were prescribed as treatment. The patient was discharged in better condition, but with continued gait disturbances. FHL2 was in the differential diagnosis and we proceeded with whole exome sequencing. This confirmed two point mutations in the *PRF1* gene, c.148G > A [p.V50M] (of maternal origin) and c.65delC [p.P22Rfs*2] (of paternal origin). The patient is undergoing chemotherapy in our hematology department and is awaiting HSCT.

DISCUSSION

FHL2 is a rare autosomal-recessive disorder with a poor prognosis, characterized by fever, hepatosplenomegaly, and pancytopenia, which generally presents in infancy or early childhood. In familial cases with a known genetic abnormality (FHL with mutations), a diagnosis can be made without consideration of acquired HLH criteria (Henter et al., 1991; Janka and Schneider, 2004; Henter et al., 2007). In these four cases, the neurologic manifestations were the initial clinical presentation. Three of four cases had fever which was neglected. The expected symptoms such as leukopenia and hepatosplenomegaly appeared as the disease progressed. Laboratory tests revealed leukopenia and neutropenia/agranulocytosis in all four cases. Hepatosplenomegaly was observed in Cases 1 and 2. No patient suffered from hypertriglyceridemia, hypofibrinogenemia, or hemophagocytosis.

The systemic manifestations of FHL typically evolve over a period of time, and neurological manifestations are reported in 20 to 73% of FHL patients (Horne et al., 2008; Dias et al., 2013). These four patients initially manifested CNS symptoms in disease presentation. These neurologic manifestations included seizures, ataxia, facial palsies, spasticity, irritability, gait disorders, and coma. These were consistent with other reports (Henter et al., 1991). Determining a diagnosis is more difficult in patients without a positive family history. In patients with FHL2, neurologic manifestations presenting as the initial clinical indications may delay accurate diagnosis since the symptoms were similar to other neurological diseases, such as acute disseminated encephalomyelitis (ADEM), meningitis, encephalopathy, multiple sclerosis (MS), and CNS vasculitis. All four patients were initially misdiagnosed with a demyelinating disease, such as ADEM or MS. The corticosteroid administration can provide temporary improvement. The time from symptom onset to accurate diagnosis was more than a year

in three cases. To identify the clinical features in these cases, the symptoms were often worse than at onset, although transitional symptomatic relief was observed after treatment. Since gray matter dysfunction is relatively common in these patients, all of the patients experienced changes in mental status, described variably as irritability, somnolence, disturbance of consciousness, and encephalopathy. Patients with FHL2 and CNS symptoms often have abnormal CSF findings including mildly elevated cell and/or protein levels. In three cases, the CSF protein content was moderately elevated, although the CSF cell count analysis was normal. The CSF cell count was high in Case 3, which may be associated with a CNS infection.

MRI findings significantly help in the assessment and monitoring of neurological involvement in patients with FHL2 as they reveal abnormal signals in the cerebral hemispheres, basal ganglia, cerebellum, and brain stem as well as hypersignal intensities on T2 and FLAIR images. MRIs also indicate multifocal and bilateral abnormalities with symmetric involvement in T2-weighted imaging. Abnormal signals in the bilateral cerebrum including white and gray matter and junctions were discovered. Abnormal cerebellar lesions were observed in all four cases (**Figures 1 and 2**). Abnormal spot or ring enhancements and/or hemorrhage, especially in the cerebellar hemisphere, were also observed in the brain MRIs (**Appendix 2**). Furthermore, large, ill-defined, confluent lesions were, as the disease was progressive and new lesions were observed subsequent to the original lesion(s). Chronic changes such as

atrophy were noted in Case 1. Abnormalities on MRIs appeared to be roughly proportional the severity of the clinical manifestations.

In contrast to typical early-onset FHL, our patient initially demonstrated isolated CNS involvement. Nonspecific clinical and neuroradiological findings with initial isolated CNS involvement can result in misdiagnosis and delayed diagnosis in these cases. The diagnosis was established with clinical and genetic testing. An early molecular diagnosis can improve the prognosis of FHL2 patients with prominent CNS involvement. Next-generation gene sequencing greatly aids in diagnosing FHL2. In 1999, *PRF1* gene was first identified as a cause of FHL (Stepp et al., 2015). In four patients, seven different compound heterozygous mutations (of which three were novel mutations) were identified: four missense mutations and three deletion mutations (**Figure 3**). The pathogenic probability of missense mutations was analyzed using different prediction programs. All of these novel mutations were present in the highly conserved region across the species. Cases 2 and 4 had previously reported mutations p.T450M and p.285del; and p.P22Rfs*2 and p.V50M. Cases 1 and 3 had novel mutations p.Y212H, p.361-364del, and p.D436Y. A heterozygous missense mutation c.1349C > T [p.T450M] was present in two patients, Cases 2 and 3.

Case 1 had two novel mutations, which had not been previously published or reported. Ueda *et al.* identified homozygosity for c.1090_1091delCT in exon 3 of the *PRF1* gene, resulting in a frameshift and premature termination (Ueda et al., 2003). This mutation has also been reported in

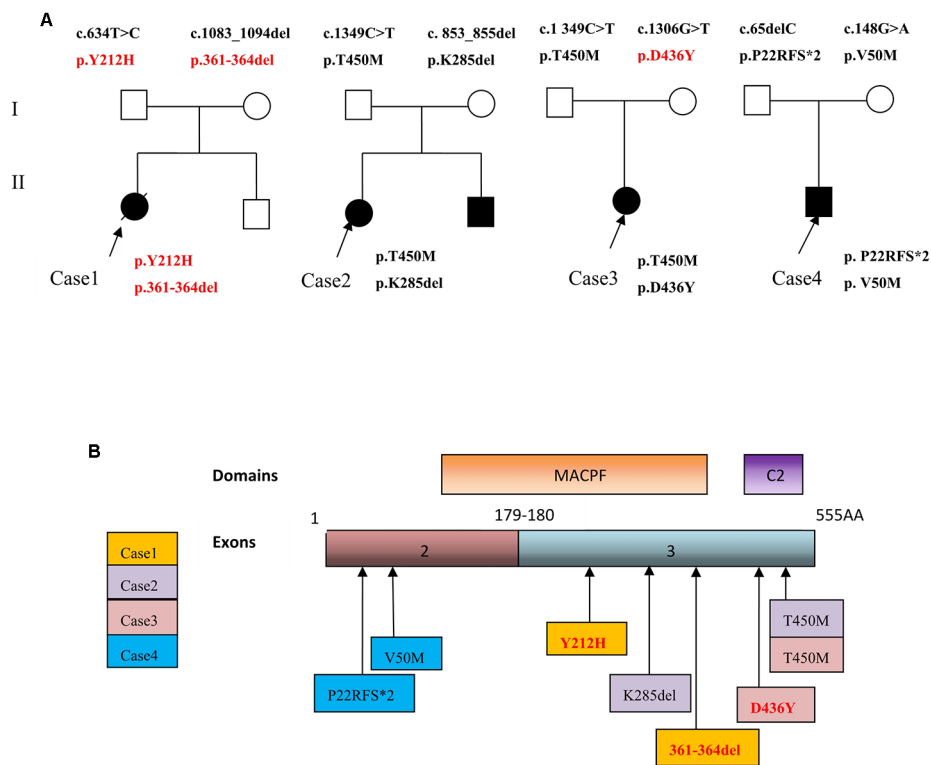


FIGURE 3 | (A) Family trees of four FHL2 patients with perforin gene mutations. **(B)** Schematic illustration showing mutation sites of the perforin gene in the FHL2 patients.

Korea (Kim et al., 2014). In our study, the c.1083_1094del mutation resulted in the loss of the Arg361 Arg362 Ala363 Leu364 [p.361_364del] without a frameshift. These mutations were found to be deleterious *via in silico* analysis. The patient's condition deteriorated rapidly, and death occurred within several months. This may be due to the mutations, and a further study (such as functional research) may be important.

The mutation in Case 2 c.1349C > T [p.T450M] has also been described in a Japanese case (Ueda et al., 2007) in which a female patient developed HLH at age 7 with an Epstein-Barr Virus (EBV) infection. She had no neurological symptoms; however, brain MRI scans showed high T2 lesions (Ueda et al., 2007). Case2 did not have EBV infection, and the neurologic manifestations were the initial clinical presentation with multiple abnormal signals in the cerebellar hemispheres seen on the MRI. The 3-bp (853_855) deletion in exon 3, resulting in loss of the lys285 residue [c. 853_855del; p.285del] was noted in two reports in two children from Turkey (Goransdotter Ericson et al., 2001; Zur Stadt et al., 2006). For these two patients, no more clinical information was provided, and one patient died before treatment (Goransdotter Ericson et al., 2001).

For Case 3, the patient had two heterozygous mutations, of which c.1349C > T [p.T450M] was also observed in Case 3 as a single heterozygous mutation of paternal origin. This was also reported in two half-brothers in China (Liu et al., 2018). The main clinical manifestations of the older brother were immunodeficiency and epilepsy, and CSF analysis showed increased WBCs. The younger brother did not suffer from seizures, but brain effusion cytology showed fewer lymphocytes. An 8-year-5-month-old Chinese girl with recurrent HLH and severe CNS disease was also analyzed with this mutation (Ding et al., 2019). The T450M mutation was also reported in a Chinese study (Zhang et al., 2016) in which the patient was a 2-year-old girl; however, no more clinical information was provided. In our study we discovered two patients with this mutation, which may be an important mutation in China, especially in CNS involvement of FHL2. The c.1306G > T [p.D436Y] is a novel mutation which was determined *via in silico* analysis to be deleterious.

For Case 4, the patient had compound heterozygous mutations. The c.148G > A has been described in a compound heterozygote case in Turkey (Goransdotter Ericson et al., 2001), in which the patient was a 4-month-old female with a fever, splenomegaly, cytopenia, and hemophagocytosis, and without CNS symptoms. She had a compound heterozygous mutation in *PRF1*, Tyr219stop, and Val50met and died after the treatment protocol HLH-94 (Henter et al., 1997). The c.148G > A [V50M] missense mutation is conserved in human, mouse and rat (Lowin et al., 1995). The c.65delC mutation [p.P22Rfs*2] was previously reported in two patients from Korea and one from Hong Kong in a compound heterozygous manner (Chiang et al., 2014; Kim et al., 2014; Kim et al., 2017). Two of the patients (one from Korea and one from Hong Kong) were diagnosed with CNS involved FHL2 (Chiang et al., 2014; Kim et al., 2017). The Hong Kong patient's brain MRI revealed diffuse parenchymal and leptomeningeal enhancing lesions. CSF analysis also showed pleocytosis and a lymphohistiocytic infiltrate. The other Korean patient had

progressive multiple organ failure (Kim et al., 2014). The c.65delC [p.Pro22Argfs*2] is a deleterious mutation from the introduction of a premature stop codon due to a frameshift. The three patients reported died after chemotherapy. For our case the boy is undergoing chemotherapy using the HLH-2004 protocol. The long term follow-up is necessary.

The most detrimental *PRF1* mutations associated with minimal or no protein expression, present during early infancy, with a mean onset at 2 months (Trizzino et al., 2008). In our study, four patients with FHL2 were older than one year old at age of onset. One possible reason for this delay is that compound heterozygous *PRF1* missense mutations encode for partially active perforin which might enable patients to survive for a significant period.

Corticosteroid treatment provided improvement in three of four patients; however, Case 1 did not improve, and her diagnosis was made after death *via* whole exome sequencing. There is potentially a strong correlation between genetic defects and the function of perforin. The compound heterozygosity of Case 1 may seriously influence the function of perforin, resulting in death. The severity of the disease depends on the residual activity of perforin. Additional studies could further explain this.

CONCLUSIONS

This study considers neurological manifestations such as seizures, ataxia, and facial palsies, which may provide different clinical presentations than the typical FHL2 symptoms. Thus it is important that pediatricians are aware of the potential diagnosis of FHL2. These symptoms should indicate testing for FHL2 markers, such as fever, leukopenia, hepatosplenomegaly, CSF protein elevation, or brain MRI showing multilobar and widespread lesions in bilateral cerebral hemispheres and the cerebellum. Spot or ring enhancements or hemorrhage were also observed. Whole exome sequencing or genetic analysis of *PRF1* should be performed earlier during the differential diagnosis evaluation to help with prompt, accurate diagnosis, and treatment. Pathogenic mutations in the *PRF1* gene were identified in our patients with FHL2. Three novel mutations were discovered in our study, which may play an important role in the diagnosis of new cases of FHL2.

DATA AVAILABILITY STATEMENT

The data included in this study are available upon request to the corresponding author of this article.

ETHICS STATEMENT

For research involving human participants, informed consent has been obtained from the patients or the guardian of the patients. The research has been approved by the Ethics Committee of the Beijing Children's Hospital.

AUTHOR CONTRIBUTIONS

Conceived and designed the manuscript: FF, W-XF, C-HD, and T-LH. Clinical data acquisition: W-HZ, J-WL, SG, and X-YY. Analyzed the clinical and genetic data: W-XF and YW. Wrote the paper: W-XF, X-WZ, FF, and YW.

FUNDING

This study was funded by the Beijing Municipal administration of Hospitals incubating Program (PX2017065).

REFERENCES

- Chiang, G. P., Li, C. K., Lee, V., Cheng, F. W., Leung, A. W., Imashuku, S., et al. (2014). Perforin gene mutation in familial haemophagocytic lymphohistiocytosis: the first reported case from Hong Kong. *H. K. Med. J. Xianggang yi xue za zhi* 20 (4), 339–342. doi: 10.12809/hkmj134041
- Dias, C., McDonald, A., Sincan, M., Rupps, R., Markello, T., Salvarinova, R., et al. (2013). Recurrent subacute post-viral onset of ataxia associated with a PRF1 mutation. *Eur. J. Hum. Genet. EJHG* 21 (11), 1232–1239. doi: 10.1038/ejhg.2013.20
- Ding, Q., Guo, X., and Li, Q. (2019). [Analysis of PRF1 gene variant in a child with late-onset familial hemophagocytic lymphohistiocytosis type 2 and severe central nervous system disease]. *Zhonghua Chuan Zhonghua Yixue yichuanxue zazhi Chin. J. Med. Genet.* 36 (6), 592–594. doi: 10.3760/cma.j.issn.1003-9406.2019.06.015
- Dufourcq-Lagelouse, R., Jabado, N., Le Deist, F., Stephan, J. L., Souillet, G., Bruin, M., et al. (1999). Linkage of familial hemophagocytic lymphohistiocytosis to 10q21-22 and evidence for heterogeneity. *Am. J. Hum. Genet.* 64 (1), 172–179. doi: 10.1086/302194
- Feldmann, J., Callebaut, I., Raposo, G., Certain, S., Bacq, D., Dumont, C., et al. (2003). Munc13-4 is essential for cytolytic granules fusion and is mutated in a form of familial hemophagocytic lymphohistiocytosis (FHL3). *Cell* 115 (4), 461–473. doi: 10.1016/S0092-8674(03)00855-9
- Goransdotter Ericson, K., Fadeel, B., Nilsson-Ardnor, S., Soderhall, C., Samuelsson, A., Janka, G., et al. (2001). Spectrum of perforin gene mutations in familial hemophagocytic lymphohistiocytosis. *Am. J. Hum. Genet.* 68 (3), 590–597. doi: 10.1086/318796
- Henter, J. I., Elinder, G., Soder, O., and Ost, A. (1991). Incidence in Sweden and clinical features of familial hemophagocytic lymphohistiocytosis. *Acta Paediatrica Scand.* 80 (4), 428–435. doi: 10.1111/j.1651-2227.1991.tb11878.x
- Henter, J. I., Arico, M., Egeler, R. M., Elinder, G., Favara, B. E., Filipovich, A. H., et al. (1997). HLH-94: a treatment protocol for hemophagocytic lymphohistiocytosis. HLH study Group of the Histiocyte Society. *Med. Pediatr. Oncol.* 28 (5), 342–347. doi: 10.1002/(SICI)1096-911X(199705)28:5<342::AID-MPO3>3.0.CO;2-H
- Henter, J. I., Horne, A., Arico, M., Egeler, R. M., Filipovich, A. H., Imashuku, S., et al. (2007). HLH-2004: Diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. *Pediatr. Blood Cancer* 48 (2), 124–131. doi: 10.1002/pbc.21039
- Horne, A., Trottestam, H., Arico, M., Egeler, R. M., Filipovich, A. H., Gadner, H., et al. (2008). Frequency and spectrum of central nervous system involvement in 193 children with haemophagocytic lymphohistiocytosis. *Br. J. Haematol.* 140 (3), 327–335. doi: 10.1111/j.1365-2141.2007.06922.x
- Janka, G. E., and Schneider, E. M. (2004). Modern management of children with haemophagocytic lymphohistiocytosis. *Br. J. Haematol.* 124 (1), 4–14. doi: 10.1046/j.1365-2141.2003.04726.x
- Kim, J. Y., Shin, J. H., Sung, S. I., Kim, J. K., Jung, J. M., Ahn, S. Y., et al. (2014). A novel PRF1 gene mutation in a fatal neonate case with type 2 familial hemophagocytic lymphohistiocytosis. *Korean J. Pediatr.* 57 (1), 50–53. doi: 10.3345/kjp.2014.57.1.50
- Kim, M. S., Cho, Y. U., Jang, S., Seo, E. J., Im, H. J., and Park, C. J. (2017). Familial hemophagocytic lymphohistiocytosis Type 2 in a Korean infant with compound heterozygous PRF1 defects involving a PRF1 mutation, c.1091T >

ACKNOWLEDGMENTS

We would like to thank the patients and the many clinicians and the clinical laboratory scientists who contributed to this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.00126/full#supplementary-material>

- G. *Ann. Lab. Med.* 37 (2), 162–165. doi: 10.3343/alm.2017.37.2.162
- Liu, C., Li, M., Wu, X., Yao, X., and Zhao, L. (2018). Type 2 familial hemophagocytic lymphohistiocytosis in half brothers: a case report. *Medicine* 97 (30), e11577 1–7. doi: 10.1097/MD.00000000000011577
- Lowin, B., Peitsch, M. C., and Tschopp, J. (1995). Perforin and granzymes: crucial effector molecules in cytolytic T lymphocyte and natural killer cell-mediated cytotoxicity. *Curr. Topics In Microbiol. Immunol.* 198, 1–24. doi: 10.1007/978-3-642-79414-8_1
- Osinska, I., Popko, K., and Demkow, U. (2014). Perforin: an important player in immune response. *Central-European J. Immunol.* 39 (1), 109–115. doi: 10.5114/ceji.2014.42135
- Stepp, S. E., Dufourcq-Lagelouse, R., and Kumar, V. (2015). Pillars article: Perforin gene defects in familial hemophagocytic lymphohistiocytosis. *Science*. 1999. 286: 1957–1959. *J. Immunol.* 194 (11), 5044–5046. doi: 10.1126/science.286.5446.1957
- Trizzino, A., Stadt, U., Ueda, I., Risma, K., Janka, G., Ishii, E., et al. (2008). Genotype-phenotype study of familial haemophagocytic lymphohistiocytosis due to perforin mutations. *J. Med. Genet.* 45 (1), 15–21. doi: 10.1136/jmg.2007.052670
- Ueda, I., Morimoto, A., Inaba, T., Yagi, T., Hibi, S., Sugimoto, T., et al. (2003). Characteristic perforin gene mutations of haemophagocytic lymphohistiocytosis patients in Japan. *Br. J. Haematol.* 121 (3), 503–510. doi: 10.1046/j.1365-2141.2003.04298.x
- Ueda, I., Kurokawa, Y., Koike, K., Ito, S., Sakata, A., Matsumura, T., et al. (2007). Late-onset cases of familial hemophagocytic lymphohistiocytosis with missense perforin gene mutations. *Am. J. Hematol.* 82 (6), 427–432. doi: 10.1002/ajh.20878
- Yang, S., Zhang, L., Jia, C., Ma, H., Henter, J. I., and Shen, K. (2010). Frequency and development of CNS involvement in Chinese children with hemophagocytic lymphohistiocytosis. *Pediatr. Blood Cancer* 54 (3), 408–415. doi: 10.1002/pbc.22239
- Zhang, J., Wang, Y. N., Wang, J. S., Wu, L., Wei, N., Fu, L., et al. (2016). [The significance of pedigree genetic screening and rapid immunological parameters in the diagnosis of primary hemophagocytic lymphohistiocytosis]. *Zhonghua xue za zhi Zhonghua xueyexue zazhi* 37 (7), 565–570. doi: 10.3760/cma.j.issn.0253-2727.2016.07.005
- Zur Stadt, U., Beutel, K., Kolberg, S., Schneppenheim, R., Kabisch, H., Janka, G., et al. (2006). Mutation spectrum in children with primary hemophagocytic lymphohistiocytosis: molecular and functional analyses of PRF1, UNC13D, STX11, and RAB27A. *Hum. Mutat.* 27 (1), 62–68. doi: 10.1002/humu.20274

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Feng, Yang, Li, Gong, Wu, Zhang, Han, Zhuo, Ding and Fang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership