Identification of copy number variants contributing to hallux valgus

Wentao Zhou[1†], Jun Jia[2†], Hui-Qi Qu[3†], Feier Ma[1], Junyi Li[1], Xiaohui Qi[1], Xinyi Meng[1], Zhiyong Ding[4], Gang Zheng[5], Hakon Hakonarson[3,6,7], Xiantie Zeng[2], Jin Li[1] and Qianghua Xia[8,1]*

[1]Department of Cell Biology, The Province and Ministry Co-sponsored Collaborative Innovation Center for Medical Epigenetics, Key Laboratory of Immune Microenvironment and Disease (Ministry of Education), School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China, [2]Department of Surgery of Foot and Ankle, Tianjin Hospital, Tianjin, China, [3]Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, United States, [4]Mills Institute for Personalized Cancer Care, Fynn Biotechnologies Ltd., Jinan, China, [5]National Supercomputer Center in Tianjin (NSCC-TJ), Tianjin, China, [6]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, United States, [7]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States, [8]Department of Bioinformatics, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China

Hallux valgus is a common form of foot deformity, and genetic factors contribute substantially to the pathogenesis of hallux valgus deformity. We conducted a genetic study on the structural variants underlying familial hallux valgus using whole exome sequencing approach. Twenty individuals from five hallux valgus families and two sporadic cases were included in this study. A total of 372 copy number variations were found and passed quality control filtering. Among them, 43 were only present in cases but not in controls or healthy individuals in the database of genomic variants. The genes covered by these copy number variations were enriched in gene sets related to immune signaling pathway, and cytochrome P450 metabolism. The hereditary CNVs demonstrate a dominant inheritance pattern. Two candidate pathogenic CNVs were further validated by quantitative-PCR. This study suggests that hallux valgus is a degenerative joint disease involving the dysregulation of immune and metabolism signaling pathways.

Introduction

Hallux valgus (HV) deformity refers to the lateral deflection of the great toe at the first metatarsophalangeal joint (Perera et al., 2011; Hecht and Lin, 2014), which is the most common forefoot deformity often requires surgery with a prevalence rate of 23% in people aged 18–65 (CI: 16.3%–29.6%) (Nix et al., 2010). African Americans are more likely to have HV than people of European ancestry [adjusted odds ratio (aOR) = 2.01, 95% confidence interval [CI] = 1.39–2.92] (Golightly et al., 2012), and no significant difference has been found in the incidence of HV in other populations. Conservative treatment is feasible for patients with deformity but without symptoms or mild symptoms (Bayar et al., 2011). Although the symptoms can be alleviated, they can not be completely reversed. If the patient's pain persists, surgery is necessary. However, many surgical complications and poor

prognosis have brought great burden to individuals and their families (Sammarco and Idusuyi, 2001). Prevention and early intervention are therefore important.

The pathogenesis of HV deformity is complex (Mann and Coughlin, 1981). HV may be associated with inappropriate footwear. HV is 15 times more common in people who wear shoes than in those who don't (Perera et al., 2011), and shoes that tighten the front foot appear to be one of the leading causes of HV. Heredity is another important risk factor for HV development, especially in adolescents. A positive family history of HV has been reported in many studies (Lieberson and Mendes, 1991; Pique-Vidal et al., 2007), and study has found that HV susceptibility is related to genetic polymorphisms associated with arthritis (Perera et al., 2011; Hecht and Lin, 2014). Linear bone arrangement or static stabilizer relaxation due to heredity may also lead to HV deformity (Perera et al., 2011).

Genetic factors make substantial contribution to the pathogenesis of HV deformity. In a study of 350 patients with three generations family trees, 90% had at least one affected relative, which is consistent with autosomal dominant inheritance pattern (Pique-Vidal et al., 2007). The heritability of HV in European ancestry populations is between 0.29 and 0.89 (Hannan et al., 2013), while the rate of HV in Korea is ~0.51 (Lee et al., 2014). In a genome-wide association study (GWAS) on European ancestry population including 1786 cases of HV deformity and 2623 controls, genome-wide SNPs accounted for 50% of the phenotypic variance in males and 48% of the phenotypic variance in females (Hsu et al., 2015). The missing heritability may lie in the contribution of rare variants and structural variants, which have been underexplored in previous studies.

Structural variation is generally defined as a region of DNA of approximately 1 kb or larger that includes changes in copy number, chromosomal position, or orientation between individuals (Escaramis et al., 2015). A major class of genomic structural variation is copy number variation (CNV), which includes deletion and duplication of sequences (Ionita-Laza et al., 2009). In addition to single nucleotide polymorphisms (SNP) (Sachidanandam et al., 2001), CNVs are a major source of variation in the human genome, with significant effects on evolution and disease susceptibility (Conrad and Hurles, 2007). For example, at least 15% of neurodevelopmental diseases are caused by local dose imbalances in dozens of genes due to CNVs (Girirajan et al., 2011). Furthermore, CNVs make significant contributions to development of bone disorders and CNV analysis increases the diagnostic yield for these diseases. CNVs are significantly associated with osteoporosis (Yang et al., 2008; Costantini et al., 2018) and CNV is an important genetic factor for the etiology of fetal skeletal dysplasia (Wit et al., 2014; Bai et al., 2022). In addition, studies have shown that many CNVs confer greater disease risk than SNPs (Ionita-Laza et al., 2009). However, the contribution of CNVs to the pathogenesis and development of HV has not been investigated.

Aiming to explore the contribution of potential pathogenic CNVs to the development of HV deformity, we conducted the first study on the structural variants underlying familial HV using whole exome sequencing (WES) approach, which may be helpful for future risk prediction of HV.

Materials and methods

Samples and ethics statement

In this study, we recruited a total of 22 Chinese participants from 5 families (including 17 cases and 3 controls without any foot deformity) and two sporadic cases (Figure 1). The recruitment and the WES study were approved by Tianjin Hospital, and all participants provided written informed consent. The diagnosis of HV was made by clinical experts of the Foot and Ankle Surgery Group of Orthopaedic Branch of Chinese Medical Association and Foot and Ankle Surgery Professional Committee of Orthopaedic Physician Branch of Chinese Medical Association, according to the expert consensus (Foot and Ankle Working Committee et al., 2015). HV is diagnosed by combining the evaluation of clinical presentations, physical examination, auxiliary imaging examination and medical history. The severity of HV was determined by hallux valgus angle (HVA) and intermetatarsal angle (IMA) (normal: HVA < 16°, IMA < 10°; mild: HVA < 20°, IMA < 13°; moderate: 20° < HVA ≤ 40°, 13° < IMA ≤ 16°; severe: HVA > 40°, IMA > 16°). The single nucleotide variant analysis of the WES data of three families has been reported in our previous publication (Jia et al., 2021). With the addition of data from HV families and sporadic cases, we performed the current CNV study. The genetic genealogy of the 5 families shows that each proband has at least one first-degree relative as a HV patient.

Genomic DNA extraction and whole-exome sequencing

Genomic DNA was extracted from peripheral blood sample of each subject following standard procedures. The TargetSeqTM Enrichment Kit (iGeneTechTM) Human Exome Capture Kit was used for library construction. The Illumina sequencing platform was used for paired next-generation sequencing.

Quality control of sequencing result files

The Trim Galore software (Martin, 2011) (https://github.com/FelixKrueger/TrimGalore) was used to remove low-quality base (−q 25), limit maximum allowable error rate (default −e 0.1), remove reads < 36 nt (-length 36), remove double-ended overlap>3, and remove reads as a unit (-paired) from the raw fastQ files. Burrows-Wheeler Aligner (BWA)-MEM (version 0.7.17) (Li, 2013) was used to align the reads with the reference genome to obtain the SAM file.

After reordering SAM files were converted to BAM (Raw BAM) files and polymerase chain reaction (PCR) duplicates were marked using Picard (v1.91). SamTools (version 1.58) (Li et al., 2009) was used for quality control of the binary alignment graph files generated between them. Then, the Genome Analysis ToolKit (GATK) (version 3.8) (McKenna et al., 2010) was used to re-align indel regions and correct the base mass fraction. And after Base quality
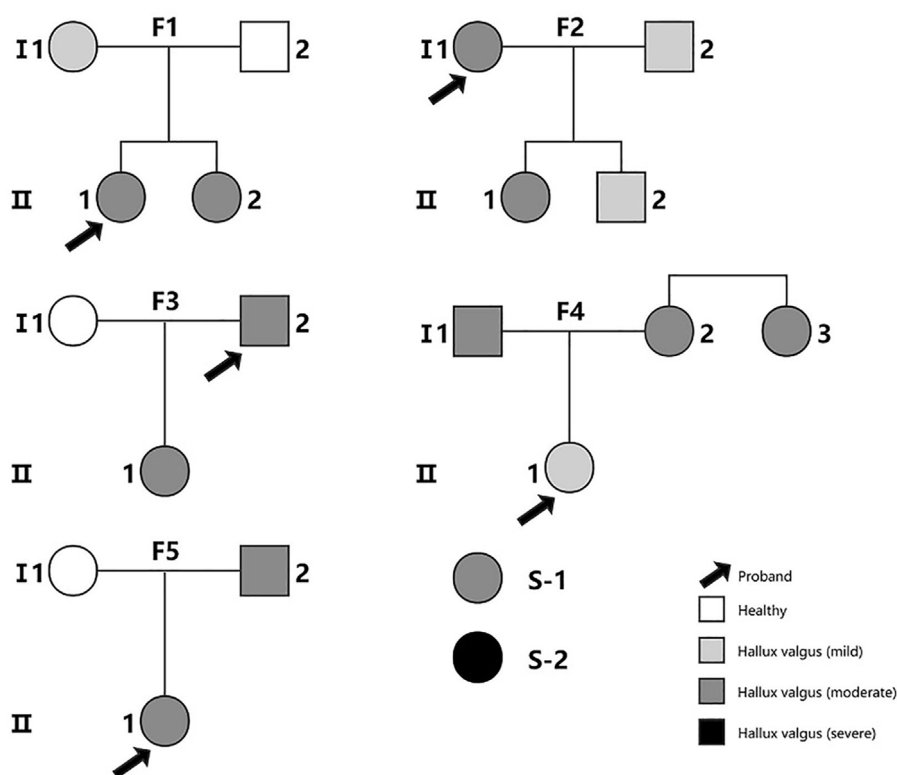
FIGURE 1
The family pedigree diagrams for all the subjects in the study. S-1 and S-2 are two sporadic cases. Families F1, F3 and F5 were included in our previous study of single nucleotide variants (Jia et al., 2021).The color indicates the severity of the patient's illness from mild to severe.

score recalibrations (BQSRs), we used verifyBamID (version 1.1.3) (Jun et al., 2012) to confirm that there was no contamination of the cross samples. The resulting BAM files were used for CNV calling.

CNVs calling by GATK and XHMM

The eXome-Hidden Markov Model (XHMM) (Fromer and Purcell, 2014) was used to call CNVs from filtered normalized targeted exome sequence data. The XHMM divided the chromosomal regions into three types: diploid, deletion, and duplication. The entire pipeline is plotted as a flowchart and shown in Figure 2.

CNVs annotation

We used "scan_regoin.pl" program in GenGen(Wang et al., 2007) to scan genomic features and to find the CNVs that overlapped with those in the database of genomic variants (DGV) (MacDonald et al., 2014) and filtered them out. The DGV database contains genomic structural variants over 50 bp in healthy individuals.

We used "scan_regoin.pl" program in GenGen (Wang et al., 2007) with the hg19_refGene and hg19_refLink files to annotate the CNV regions against the RefGene annotation to find the genes overlapped the CNV regions.

Then, we filtered out CNVs present among controls, and retained case-only CNVs that were repeated among the HV cases but not carried by any of the controls.

Quality control filtering of CNVs

We performed quality control filtering based on "Q_EXACT" and "Q_SOME" scores in the XCNV file according to the instruction of XHMM software. "Q_EXACT" and "Q_SOME" represent the phred-scaled quality of a CNV event along the entire interval and the same CNV event in the interval respectively.

Pathway enrichment of CNVs

The 67 genes overlapped with the 43 CNVs were used as input to the web-portal of OBAS (http://kobas.cbi.pku.edu.cn/) (Bu et al., 2021) and STRING (version 11.0) (https://string-db.org/) (Szklarczyk et al., 2021) for pathway enrichment and protein-protein interaction (PPI) analysis. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database and Search Tool for the Retrieval of Interacting Genes/Proteins database were used as the reference databases.

FIGURE 2
The pipeline of copy number variation analysis based on whole exome sequencing data.
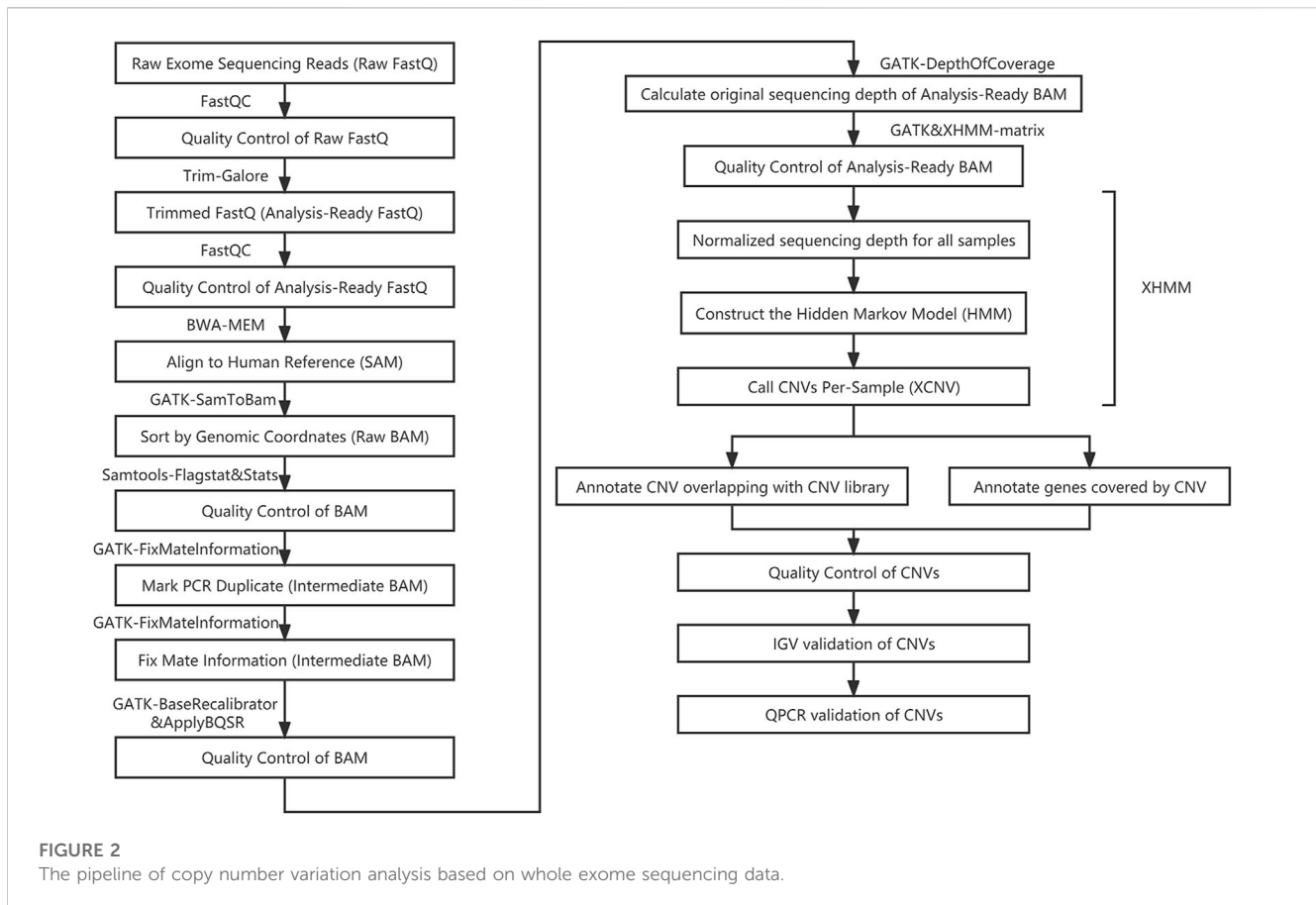
Examination of CNVs using integrative genomics viewer

For candidate CNVs of interest, we further conducted visual inspection using the Integrative Genomics Viewer, based on the aligned bam files of the WES data of each CNV carrier and non-carriers within the same family.

qPCR validation of CNVs

For the potential pathogenic CNVs, we chose a ~100 bp fragment in each CNV for qPCR validation. The qPCR experiment was carried out with the Sybr Green I system. Each reaction in a 10 μL system contains 4 ng of genomic DNA, 10 μm per-primer, 2× ChamQ universal SYBR qPCR Master Mix, and DD-H_2O. Each sample was repeated for three times. The geometric mean values of the CT values of the control sequence GAPDH and the sample sequence were calculated to obtain △CT values for each sample. We calculated the $2^{-\triangle\triangle CT}$ of each sample by taking the $2^{-\triangle CT}$ of members without CNV in the family as the relative reference value. Finally, the existence of CNV was examined by the relative value of $2^{-\triangle\triangle CT}$ of each sample in the pedigree.

Results

The identification of 43 CNVs related to HV

To assess the potential contribution of structural variants to HV, we carried out a CNV study based on the WES data of five HV families including 17 cases and 3 family members without HV and two sporadic cases (Supplementary Table S1). A total of 965 CNVs were detected based on the WES data and 372 passed quality control filtering (Supplementary Table S2).

Among these 372 CNVs, 325 CNVs were present only in cases, including 142 deletions and 183 duplications. They were not detected in the any of the three control samples. Furthermore, 51 CNVs (26 deletions and 25 duplications) were carried by more than one samples. In order to screen for potential pathogenic CNVs, we further filtered out any CNVs with 70% overlap with those carried by healthy individuals in the DGV. Then, 43 CNVs were retained for further analysis, covering the exons of the 67 genes, with 1 CNV co-occurring in four cases, 5 CNVs co-occurring in three cases, and 37 CNVs co-occurring in two cases (Supplementary Table S3).
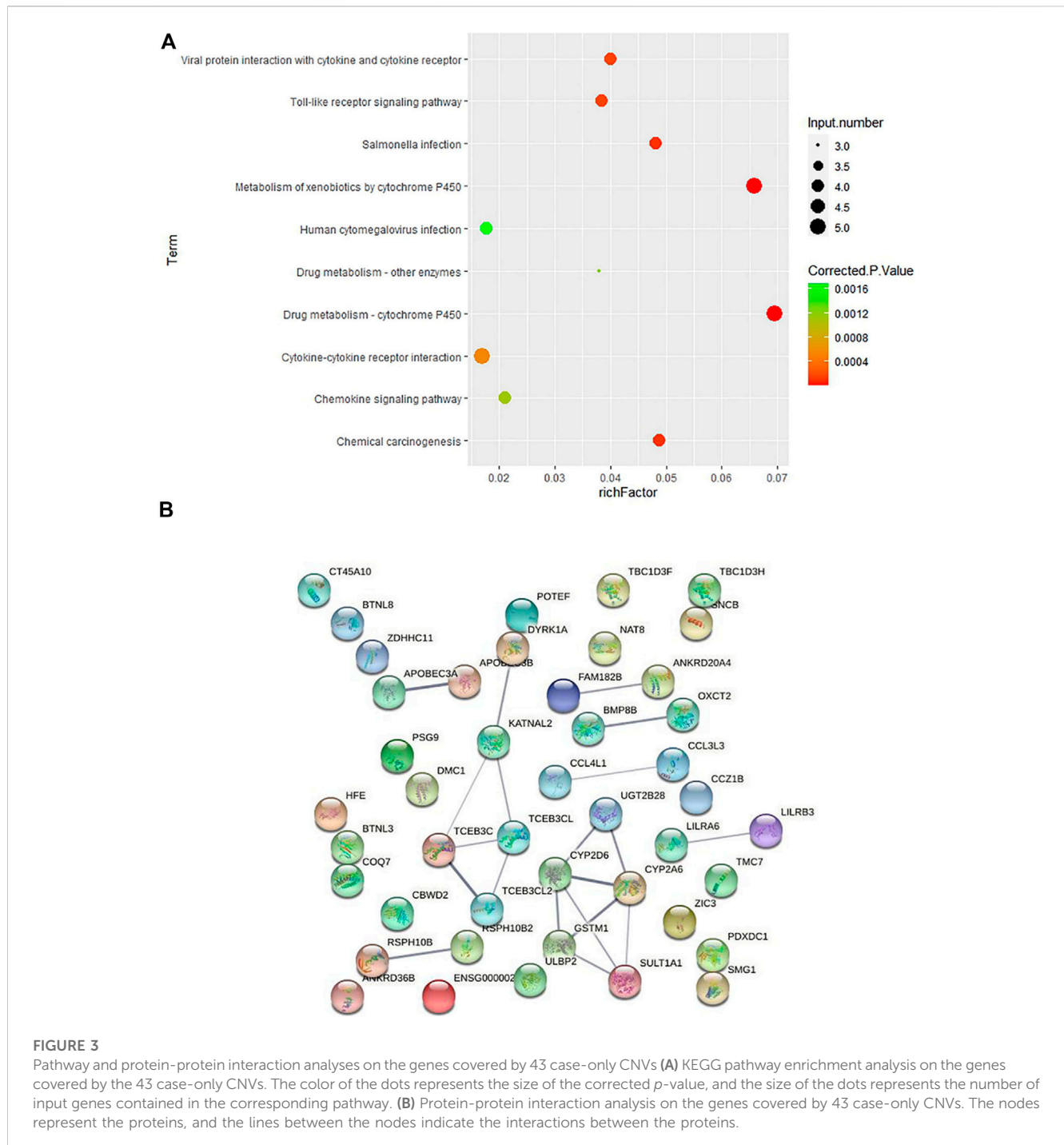
FIGURE 3
Pathway and protein-protein interaction analyses on the genes covered by 43 case-only CNVs **(A)** KEGG pathway enrichment analysis on the genes covered by the 43 case-only CNVs. The color of the dots represents the size of the corrected *p*-value, and the size of the dots represents the number of input genes contained in the corresponding pathway. **(B)** Protein-protein interaction analysis on the genes covered by 43 case-only CNVs. The nodes represent the proteins, and the lines between the nodes indicate the interactions between the proteins.

The involvement of immune and metabolism genes

In order to further explore how these CNVs may contribute to the pathogenesis of HV, we carried out pathway enrichment analysis among the 67 genes covered by case-only CNVs through KEGG PATHWAY database. Ten pathways were statistically significantly enriched (*p* < 0.001) for cytokine genes and genes involved in cytochrome P450 related metabolism and immunity/inflammation (Figure 3A; Supplementary Table S4), such as KEGG pathways "Drug metabolism—cytochrome P450", "Toll-

like receptor signaling pathway" and "Cytokine-cytokine receptor interaction", suggesting their contribution to the pathogenesis of HV.

We then conducted the PPI analysis on these genes (Figure 3B). The PPI network also supports the consistent involvement of a group of immune genes and genes functioning in drug metabolism. In the PPI network, we observed the cluster formed by proteins involved in drug metabolism (UGT2B28, CYP2D6, CYP2A6, GSTM1, SULT1A1). In addition to the well-known involvement of cytokine genes CCL4L1 and CCL3L3 in immunity, several other genes also function in regulation of the immune system. HLA-H

TABLE 1 Candidate pathogenic CNVs.

SAMPLE	CNV	INTERVAL	Q_NON_DIPLOID	Q_SOME	GENE
F1-I-1,F1-II-1,F1-II-2	DEL	chr22: 42522498–42536739	99	99	CYP2D6, CYP2D7
F2-I-1,F2-II-1	DEL	chr6: 29855550–29895036	99	93	HCG4B, HLA-H

SAMPLE, sample ID; CNV, type of copy number variation (DEL or DUP); INTERVAL, genomic range of the called CNV; Q_NON_DIPLOID, Phred-scaled quality of not being diploid, i.e., DEL or DUP event in the interval; Q_SOME, Phred-scaled quality of some CNV event in the interval; GENE, gene name.

FIGURE 4
qPCR validation of two CNVs. **(A)** Copy number detected by qPCR at chr22: 42522498–42536739 for F1-I-1, F1-II-1, F1-II-2 and F1-I-2. **(B)** Copy number detected by qPCR at chr6: 29855550–29895036 for F2-I-1, F2-II-1 and F2-I-2. **(C)** Copy number detected by qPCR at chr6: 29855550–29895036 for F2-I-1, F2-II-1 and F2-II-2.
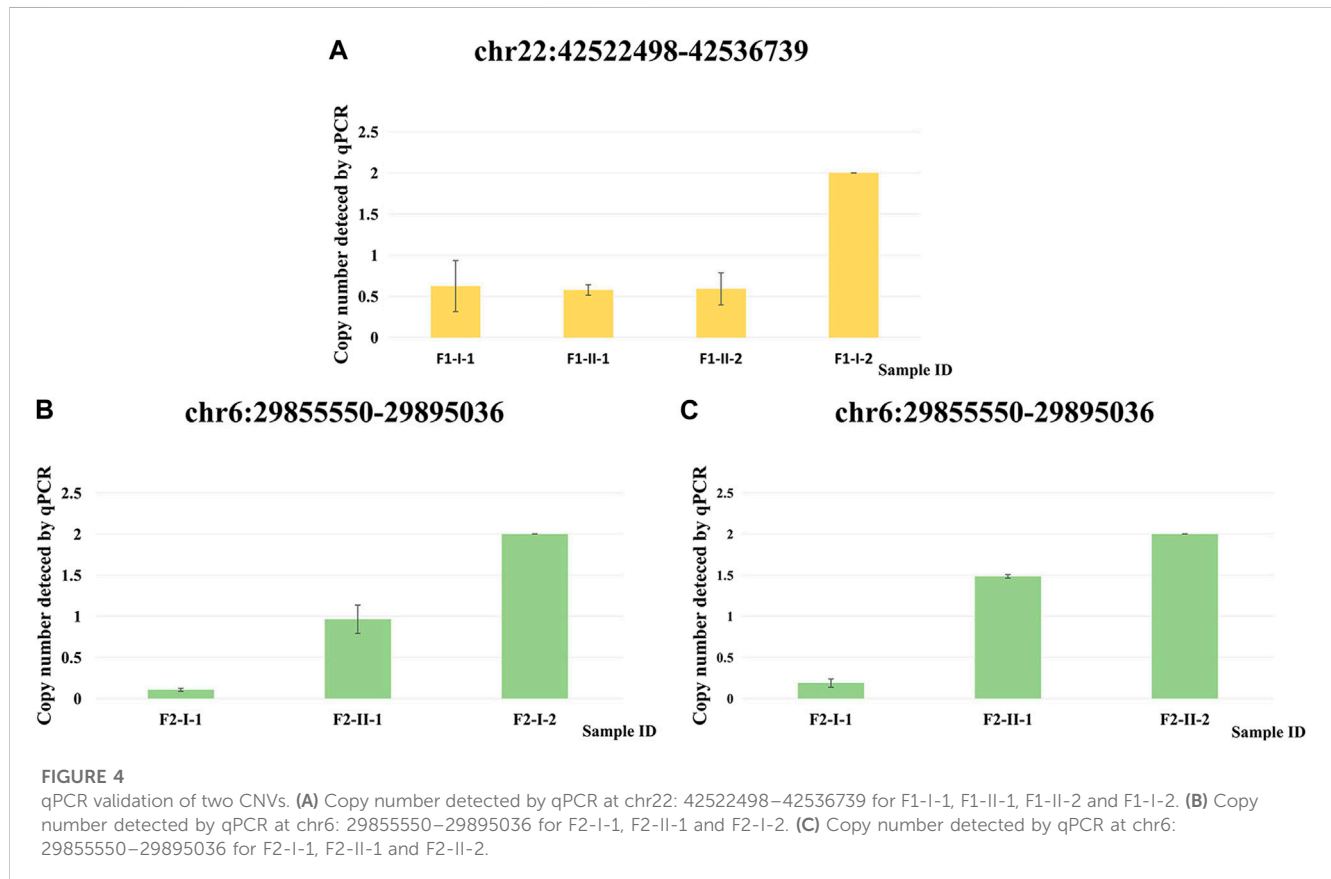
(HFE) is an important link between iron homeostasis and immune regulation (Bahram et al., 1999; Lio et al., 2002); DYRK1A gene plays a key role in regulating the differentiation of Th17 and regulatory T cells (Khor et al., 2015); and evidence suggests that genetic polymorphisms of CYP2D6 is associated with autoimmune bullous diseases induction (Rychlik-Sych et al., 2013).

Inheritance pattern of HV-related CNVs

To identify the potential disease-contributing CNVs, we examined the inheritance pattern of the CNVs which passed quality control filtering. We did not find any CNV that fits the recessive pattern. Multiple case-only CNVs consistent with dominant inheritance pattern were found in each family. For family 1, 3 and 5, dominant CNVs were referred to those present in cases (child and parent with HV) but absent in controls (parent without HV). For family 2 and family 4 in which all family members

were affected individuals with HV, the CNVs consistent with dominant inheritance were those carried by the child and one of the parents (Supplementary Table S5). Our results are consistent with the hypothesis that HV has a dominant inheritance (Pique-Vidal et al., 2007), but certainly we can not exclude the possibility of recessive inheritance model as a large proportion of the genome were not assayed in our analysis by WES.

Validation of disease-contributing CNVs

Two inherited CNVs were of particular interest and they have the highest quality (Phred-scaled quality of Non-Diploidy = 99) in our analyses (Table 1). Chr22: 42522498–42536739 deletion occurs to all patients in family 1 and absent in the healthy member of the family. This segment of CNV covered genes CYP2D7 and CYP2D6 which are involved in bone excitation effects (Jayaraman et al., 2021) and metabolic pathways related to

human immunity (Effner et al., 2017). Chr6: 29855550–29895036 deletion appears to F2-I-1 and F2-II-1 the two individuals with moderate HV but not in family members F2-I-2 and F2-II-2 who have only mild HV condition. HCG4B and HLA-H genes are in this segment of CNV. HCG4B is a proinflammatory gene whose expression is positively correlated with the expression of HLA-A and may regulate the expression of HLA-A (Chen et al., 2017). HLA-A and HLA-H function in immune homeostasis (Jordier et al., 2019). Therefore evidence from literature suggests that these two CNVs may contribute to the pathogenesis of HV.

We performed qPCR (Supplementary Table S6) as an independent experimental approach to validate these two CNVs. The presence of these two CNVs in family 1 and family 2 respectively was consistent with the results from the CNV calling of WES data (Figure 4).

Discussion

This is the first study that identified the structural variants underlying familial HV using WES approach, The identification provided us new knowledge about the genetics basis of HV, which highlighted the potential contribution of immunity/inflammation, and cytochrome P450 metabolism to the etiology of HV.

Previous studies indicated that HV malformations are likely to be caused by a variety of contributing factors, including genetics, dorsiflexion of the first metatarsal, gastrocnemius equinus, abnormal foot mechanics, and joint hypermobility (Coughlin and Jones, 2007). Interestingly, patients with autoimmune arthritis conditions such as rheumatoid arthritis and psoriatic arthritis are more likely to develop HV malformations (Rojas-Villarraga et al., 2009; Hyslop et al., 2010).

The CNV chr6: 29855550–29895036 observed in both F2-I-1 and F2-II-1 covers the HCG4B and HLA-H genes. It has been found that HCG4B is a pro-inflammatory factor and may regulate the expression of HLA-A by acting on competing endogenous RNAs spongingmiR-122 and miR-1352 (Chen et al., 2017). HLA-A and HLA-H play essential roles in the immune system, including regulation of the innate immune system, class I MHC-mediated antigen processing and presentation, interferon gamma signaling, and antigen processing-cross-presentation (Allan et al., 2002; Nitschke et al., 2016; Jordier et al., 2019). In psoriatic arthritis and rheumatoid arthritis, HLA alleles have been shown to influence susceptibility and severity of these autoimmune joint disease (Mc et al., 2015).

The CNV chr 22: 42529569–42529669 seen in F1-I-1, F1-II-1 and F1-II-2 covers CYP2D6 and CYP2D7.These genes have roles in the cytochrome P450 related xenobiotic metabolic process and oxidation-reduction process (Lessard et al., 1997). Cytochrome P450 changes the structure of cell antigens by metabolizing foreign organisms to produce reactive oxygen species, and then initiates and/or amplifies the autoimmune phenomenon through molecular simulation of the autoimmune response to the original antigen (Namazi, 2009). In addition, studies have shown that secondary metabolites produced by CYP2D6-dependent biotransformation have bone excitation effects (Jayaraman et al., 2021).

These two CNVs may have pathological effects and lead to bone structure deformities through the dysregulation of the immune system and the process of reactive oxygen species production.

Previous genetic studies on HV, including our WES study (Jia et al., 2021), revealed the association of a few genetic variants with HV. Genome-wide association study in population of European ancestry found the sex-specific association of SNVs close to genes AXIN2, ESD, ANXA1 and MRGPRX3 at the marginal significance level (Hsu et al., 2015). A later GWAS with increased sample size identified the genome-wide significant association with SNP in gene CLCA2, being an expression quantitative trait locus for COL24A1 (Arbeeva et al., 2020). Candidate gene study in Chinese population suggested variants in VDR gene and SNP rs1800629 at the 5' of TNF gene are related to HV (Tao et al., 2018; Arbeeva et al., 2020). In our study, we did not find outstanding CNVs in these genes, which could be due to differences in study population, variant types. We also focused on different regions of the human genome from the previous GWAS (the coding exons *versus* the non-coding regions). However, similar underlying genetic mechanisms were revealed. As discussed by Hsu and colleagues, AXIN2 and ANXA1 have important functions in regulation of innate and adaptive immunity, suggesting the potential role of inflammation in HV (Hsu et al., 2015). Hsu and colleagues also discussed the involvement of the serine hydrolase gene ESD gene in the recycling of sialic acids which is related to decreased antioxidant levels. Interestingly, CYP2D6 and CYP2D7 also have roles in oxidation-reduction process. The SNP rs1800629 associated with HV in Chinese population is an eQTL SNP for genes including complement gene C4A and CYP21A1P a member of the Cytochrome P450 Family. Therefore, these identifications and results from our study converge on the similar immune and drug metabolism signaling.

Because of the importance of CNV in the etiology of complex human diseases, various CNV detection methods and computational algorithms have been developed. The traditional array-based approach including array-based comparative hybridization and SNP-array approaches can efficiently detect large CNVs with a relatively accurate rate (Gabrielaite et al., 2021). CNV calling from high-throughput sequencing data has become more and more commonly utilized in research and clinical diagnosis of complex human disorders. Though still being limited by short read length, high-throughput sequencing offers the potential to identify novel CNVs in the genomic regions with sparse coverage or lack of coverage by array probes. As the most frequently used sequencing approach, WES focuses on the exomes, the copy number alterations of which are more likely to have a pathogenic effect than the other regions of the human genome. Calling CNVs from WES data also has the advantages of detecting small CNVs ranging from 1–100 kb (Gordeeva et al., 2021). This approach has the drawback due to non-uniform read depth distribution between exons which makes it prone to false positive detection (Ozden et al., 2022). It is also subjected to missing CNVs in regions with few exons. As another sequencing-based CNV detection approach, CNV-seq technology can potentially provide a unbiased coverage of the human genome and more precise estimation of CNV breakpoints (Liang et al., 2014), however it also has the limitation of relative low read depth and resolution of 0.1 Mb in size which render it more suitable to detect large CNVs. Owing to the low cost, popularity in research use, abundant existing data, simultaneous usage for single nucleotide variant detection, and a number of well-tested CNV-calling algorithms, WES data are still a common choice of CNV analysis. With the gradually reduced sequencing cost and further improved CNV-calling algorithms, WES and CNV-seq are complimentary to each other and be combined in research

and clinical diagnosis to improve clinical efficiency and diagnostic yield of multisystem anomalies (Chen et al., 2022).

There are a few limitations in our study. There are certain technical drawbacks for CNV calling based on WES as aforementioned. Small sample size is another major limitation of our study. To optimize study power and increase the possibility of identifying genetic components of HV, we adopted a family-based approach (Borecki and Province, 2008) focusing on familial Hallux Valgus. To find the potentially pathogenic CNVs related to HV, we also filtered the case-only CNVs against all the CNVs carried by healthy individuals in the DGV database. Larger sample size will be needed to further investigate the substantial contribution of various CNV regions in the human genome (Zarrei et al., 2015) to the etiology of HV, and particularly will gain statistical evidence for CNVs of uncertain clinical significance.

In summary, from an unbiased genomic approach, this first study on the structural variants in familial HV gained us new insights into the HV pathogenesis, mediated by genetic variation of immunity and inflammation, and abnormal cytochrome P450 related metabolism. It provides evidence supporting the dominant inheritance mode of HV. Our results suggest that HV is a degenerative joint disease involving the dysregulation of immune system and metabolism system.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: https://db.cngb.org/ [Accession number CNP0004036].

Ethical statement

The study involving human participants was approved by the institutional review board of Tianjin Hospital. All the participants provided the informed consents.

Author contributions

Conceptualization: XZ, QX, and JL; methodology: QX, HH, and JL; data analysis: WZ, FM, JYL, XM, XQ, ZD, and GZ; resources, JJ and XZ; interpretation: WZ, JJ, and HQ; writing—original draft preparation, WZ, JJ, and FM; writing—review and editing, HQ, XM, JYL, HH, XZ, JL, and QX; supervision, HH, JL, and QX; All authors have read and agreed to the published version of the manuscript.

Funding

Acknowledgments

Conflict of interest

ZD is an employee of Fynn Biotechnologies Ltd.

The remaining authors declare no conflicts of interests and no competing interests of this work. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1116284/full#supplementary-material

References

Allan, D. S., Lepin, E. J., Braud, V. M., O'Callaghan, C. A., and Mcmichael, A. J. (2002). Tetrameric complexes of HLA-E, HLA-F, and HLA-G. *J. Immunol. Methods* 268, 43–50. doi:10.1016/s0022-1759(02)00199-0

Arbeeva, L., Yau, M., Mitchell, B. D., Jackson, R. D., Ryan, K., Golightly, Y. M., et al. (2020). Genome-wide meta-analysis identified novel variant associated with hallux valgus in Caucasians. *J. Foot Ankle Res.* 13 (1), 11. doi:10.1186/s13047-020-0379-1

Bahram, S., Gilfillan, S., Kuhn, L. C., Moret, R., Schulze, J. B., Lebeau, A., et al. (1999). Experimental hemochromatosis due to MHC class I HFE deficiency: Immune status and iron metabolism. *Proc. Natl. Acad. Sci. U. S. A.* 96 (23), 13312–13317. doi:10.1073/pnas.96.23.13312

Bai, Y., Sun, Y., Liu, N., Wang, L., Jiao, Z., Hou, Y., et al. (2022). Genetic analysis of 55 cases with fetal skeletal dysplasia. *Orphanet J. Rare Dis.* 17 (1), 410. doi:10.1186/s13023-022-02559-4

Bayar, B., Erel, S., Şimşek, İ. E., SüMER, E., and Bayar, K. (2011). The effects of taping and foot exercises on patients with hallux valgus: A preliminary study. *Turkish J. Med. Sci.* 41, 403–409. doi:10.3906/sag-0912-499

Borecki, I. B., and Province, M. A. (2008). Genetic and genomic discovery using family studies. *Circulation* 118 (10), 1057–1063. doi:10.1161/CIRCULATIONAHA.107.714592

Bu, D., Luo, H., Huo, P., Wang, Z., Zhang, S., He, Z., et al. (2021). KOBAS-I: Intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* 49, W317–W325. doi:10.1093/nar/gkab447

Chen, X., Jiang, Y., Chen, R., Qi, Q., Zhang, X., Zhao, S., et al. (2022). Clinical efficiency of simultaneous CNV-seq and whole-exome sequencing for testing fetal structural anomalies. *J. Transl. Med.* 20 (1), 10. doi:10.1186/s12967-021-03202-9

Chen, X., Lu, X., Chen, J., Wu, D., Qiu, F., Xiong, H., et al. (2017). Association of nsv823469 copy number loss with decreased risk of chronic obstructive pulmonary disease and pulmonary function in Chinese. *Sci. Rep.* 7, 40060. doi:10.1038/srep40060

Clinical, G. G. O. M., Professional, C. F. P. D., and Group, O. G. D. P. (2019). Expert consensus on the application of low-depth whole genome sequencing in prenatal diagnosis.

Zhonghua Yi Xue Yi Chuan Xue Za Zhi 36 (4), 293–296. doi:10.3760/cma.j.issn.1003-9406.2019.04.001

Foot and Ankle Working Committee, Chinese Association of Orthopaedic Surgeons Orthopaedic Branch, Chinese Association of Orthopaedic Surgeons, Gao, P., Gui, J. C., Hong, J-J., Hong, J-S., Hu, Y., et al. (2015). Consensus on surgical management of hallux valgus from China. *Orthop. Surg.* 7, 291–296. doi:10.1111/os.12207

Conrad, D. F., and Hurles, M. E. (2007). The population genetics of structural variation. *Nat. Genet.* 39, S30–S36. doi:10.1038/ng2042

Costantini, A., Skarp, S., Kampe, A., Makitie, R. E., Pettersson, M., Mannikko, M., et al. (2018). Rare copy number variants in array-based comparative genomic hybridization in early-onset skeletal fragility. *Front. Endocrinol. (Lausanne)* 9, 380. doi:10.3389/fendo.2018.00380

Coughlin, M. J., and Jones, C. P. (2007). Hallux valgus: Demographics, etiology, and radiographic assessment. *Foot Ankle Int.* 28, 759–777. doi:10.3113/FAI.2007.0759

Effner, R., Hiller, J., Eyerich, S., Traidl-Hoffmann, C., Brockow, K., Triggiani, M., et al. (2017). Cytochrome P450s in human immune cells regulate IL-22 and c-Kit via an AHR feedback loop. *Sci. Rep.* 7, 44005. doi:10.1038/srep44005

Escaramis, G., Docampo, E., and Rabionet, R. (2015). A decade of structural variants: Description, history and methods to detect structural variation. *Brief. Funct. Genomics* 14, 305–314. doi:10.1093/bfgp/elv014

Fromer, M., and Purcell, S. M. (2014). Using XHMM software to detect copy number variation in whole-exome sequencing data. *Curr. Protoc. Hum. Genet.* 81, 7 23 1–21. doi:10.1002/0471142905.hg0723s81

Gabrielaite, M., Torp, M. H., Rasmussen, M. S., Andreu-Sanchez, S., Vieira, F. G., Pedersen, C. B., et al. (2021). A comparison of tools for copy-number variation detection in germline whole exome and whole genome sequencing data. *Cancers (Basel)* 13 (24), 6283. doi:10.3390/cancers13246283

Girirajan, S., Campbell, C. D., and Eichler, E. E. (2011). Human copy number variation and complex genetic disease. *Annu. Rev. Genet.* 45, 203–226. doi:10.1146/annurev-genet-102209-163544

Golightly, Y. M., Hannan, M. T., Dufour, A. B., and Jordan, J. M. (2012). Racial differences in foot disorders and foot type. *Arthritis Care Res. Hob.* 64 (11), 1756–1759. doi:10.1002/acr.21752

Gordeeva, V., Sharova, E., Babalyan, K., Sultanov, R., Govorun, V. M., and Arapidi, G. (2021). Benchmarking germline CNV calling tools from exome sequencing data. *Sci. Rep.* 11 (1), 14416. doi:10.1038/s41598-021-93878-2

Hannan, M. T., Menz, H. B., Jordan, J. M., Cupples, L. A., Cheng, C. H., and Hsu, Y. H. (2013). High heritability of hallux valgus and lesser toe deformities in adult men and women. *Arthritis care & Res.* 65, 1515–1521. doi:10.1002/acr.22040

Hecht, P. J., and Lin, T. J. (2014). Hallux valgus. *Hallux valgus* 98, 227–232. doi:10.1016/j.mcna.2013.10.007

Hsu, Y. H., Liu, Y., Hannan, M. T., Maixner, W., Smith, S. B., Diatchenko, L., et al. (2015). Genome-wide association meta-analyses to identify common genetic variants associated with hallux valgus in Caucasian and African Americans. *J. Med. Genet.* 52 (11), 762–769. doi:10.1136/jmedgenet-2015-103142

Hyslop, E., Mcinnes, I. B., Woodburn, J., and Turner, D. E. (2010). Foot problems in psoriatic arthritis: High burden and low care provision. *Ann. Rheum. Dis.* 69, 928. doi:10.1136/ard.2009.111971

Ionita-Laza, I., Rogers, A. J., Lange, C., Raby, B. A., and Lee, C. (2009). Genetic association analysis of copy-number variation (CNV) in human disease pathogenesis. *Genomics* 93, 22–26. doi:10.1016/j.ygeno.2008.08.012

Jayaraman, S., Reid, J. M., Hawse, J. R., and Goetz, M. P. (2021). Endoxifen, an estrogen receptor targeted therapy: From bench to bedside. *Endocrinology* 162 (12), bqab191. doi:10.1210/endocr/bqab191

Jia, J., Li, J., Qu, H., Li, M., Zhang, S., Hao, J., et al. (2021). New insights into hallux valgus by whole exome sequencing study. *Exp. Biol. Med. (Maywood)* 246 (14), 1607–1616. doi:10.1177/15353702211008641

Jordier, F., Gras, D., De Grandis, M., D'Journo, X. B., Thomas, P. A., Chanez, P., et al. (2019). HLA-H: Transcriptional activity and HLA-E mobilization. *Front. Immunol.* 10, 2986. doi:10.3389/fimmu.2019.02986

Jun, G., Flickinger, M., Hetrick, K. N., Romm, J. M., Doheny, K. F., Abecasis, G. R., et al. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* 91, 839–848. doi:10.1016/j.ajhg.2012.09.004

Khor, B., Gagnon, J. D., Goel, G., Roche, M. I., Conway, K. L., Tran, K., et al. (2015). The kinase DYRK1A reciprocally regulates the differentiation of Th17 and regulatory T cells. *Elife* 4, e05920. doi:10.7554/eLife.05920

Lee, C. H., Lee, S., Kang, H., Jung, D. E., Song, Y. M., Lee, K., et al. (2014). Genetic influences on hallux valgus in Koreans: The healthy twin study. *Twin Res. Hum. Genet.* 17, 121–126. doi:10.1017/thg.2014.10

Lessard, E., Fortin, A., Belanger, P. M., Beaune, P., Hamelin, B. A., and Turgeon, J. (1997). Role of CYP2D6 in the N-hydroxylation of procainamide. *Pharmacogenetics* 7, 381–390. doi:10.1097/00008571-199710000-00007

Li, H. (2013). "Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (preprint),". arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Sequence Alignment/Map format SAMtools* 25, 2078–2079. doi:10.1093/bioinformatics/btp352

Liang, D., Peng, Y., Lv, W., Deng, L., Zhang, Y., Li, H., et al. (2014). Copy number variation sequencing for comprehensive diagnosis of chromosome disease syndromes. *J. Mol. Diagn* 16 (5), 519–526. doi:10.1016/j.jmoldx.2014.05.002

Lieberson, S., and Mendes, D. G. (1991). *Congenital hallux valgus*. NJ: SLACK Incorporated Thorofare.

Lio, D., Balistreri, C. R., Colonna-Romano, G., Motta, M., Franceschi, C., Malaguarnera, M., et al. (2002). Association between the MHC class I gene HFE polymorphisms and longevity: A study in Sicilian population. *Genes Immun.* 3 (1), 20–24. doi:10.1038/sj.gene.6363823

Macdonald, J. R., Ziman, R., Yuen, R. K., Feuk, L., and Scherer, S. W. (2014). The database of genomic variants: A curated collection of structural variation in the human genome. *Nucleic acids Res.* 42, D986–D992. doi:10.1093/nar/gkt958

Mann, R. A., and Coughlin, M. J. (1981). Hallux valgus-etiology, anatomy, treatment and surgical considerations. *Clin. Orthop. Relat. Res.* 157, 31–41. doi:10.1097/00003086-198106000-00008

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* 17, 10. doi:10.14806/ej.17.1.200

Mc, A. A., Flatley, B., Pennington, S. R., and Fitzgerald, O. (2015). Early biomarkers of joint damage in rheumatoid and psoriatic arthritis. *Arthritis Res. Ther.* 17, 141. doi:10.1186/s13075-015-0652-z

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

Namazi, M. R. (2009). Cytochrome-P450 enzymes and autoimmunity: Expansion of the relationship and introduction of free radicals as the link. *J. Autoimmune Dis.* 6, 4. doi:10.1186/1740-2557-6-4

Nitschke, K., Luxenburger, H., Kiraithe, M. M., Thimme, R., and Neumann-Haefelin, C. (2016). CD8+ T-cell responses in hepatitis B and C: The (HLA-) A, B, and C of hepatitis B and C. *Dig. Dis.* 34, 396–409. doi:10.1159/000444555

Nix, S., Smith, M., and Vicenzino, B. (2010). Prevalence of hallux valgus in the general population: A systematic review and meta-analysis. *J. Foot Ankle Res.* 321, 21. doi:10.1186/1757-1146-3-21

Ozden, F., Alkan, C., and Cicek, A. E. (2022). Polishing copy number variant calls on exome sequencing data via deep learning. *Genome Res.* 32 (6), 1170–1182. doi:10.1101/gr.274845.120

Perera, A. M., Mason, L., and Stephens, M. M. (2011). The pathogenesis of hallux valgus. *J. Bone Jt. Surg. Am.* 93, 1650–1661. doi:10.2106/JBJS.H.01630

Pique-Vidal, C., Sole, M. T., and Antich, J. (2007). Hallux valgus inheritance: Pedigree research in 350 patients with bunion deformity. *J. Foot Ankle Surg.* 46, 149–154. doi:10.1053/j.jfas.2006.10.011

Rojas-Villarraga, A., Bayona, J., Zuluaga, N., Mejia, S., Hincapie, M. E., and Anaya, J. M. (2009). The impact of rheumatoid foot on disability in Colombian patients with rheumatoid arthritis. *BMC Musculoskelet. Disord.* 10, 67. doi:10.1186/1471-2474-10-67

Rychlik-Sych, M., Baranska, M., Waszczykowska, E., Torzecka, J. D., Zebrowska, A., and Skretkowicz, J. (2013). Genetic polymorphisms of CYP2D6 oxidation in patients with autoimmune bullous diseases. *Postepy Dermatol Alergol.* 30 (4), 211–217. doi:10.5114/pdia.2013.37030

Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., et al. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409, 928–933. doi:10.1038/35057149

Sammarco, G. J., and Idusuyi, O. B. (2001). Complications after surgery of the hallux. *Clin. Orthop. Relat. Research*® 391, 59–71. doi:10.1097/00003086-200110000-00008

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612. doi:10.1093/nar/gkaa1074

Tao, T., Jiang, Y., Li, W., Li, Y., Du, J., and Gui, J. (2018). Association of vitamin D receptor gene TaqI, BsmI, FokI, and ApaI polymorphisms and susceptibility to hallux valgus in the Chinese population. *J. Foot Ankle Surg.* 57 (4), 753–758. doi:10.1053/j.jfas.2018.01.007

Wang, K., Li, M., and Bucan, M. (2007). Pathway-based approaches for analysis of genomewide association studies. *Am. J. Hum. Genet.* 81, 1278–1283. doi:10.1086/522374

Wit, J. M., van Duyvenvoorde, H. A., van Klinken, J. B., Caliebe, J., Bosch, C. A., Lui, J. C., et al. (2014). Copy number variants in short children born small for gestational age. *Horm. Res. Paediatr.* 82 (5), 310–318. doi:10.1159/000367712

Yang, J., Wang, J., Liang, X., Zhao, H., Lu, J., Ma, Q., et al. (2019). Relationship between genetic polymorphisms of the TNF gene and hallux valgus susceptibility. *Genet. Test. Mol. Biomarkers* 23 (6), 380–386. doi:10.1089/gtmb.2018.0269

Yang, T. L., Chen, X. D., Guo, Y., Lei, S. F., Wang, J. T., Zhou, Q., et al. (2008). Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. *Am. J. Hum. Genet.* 83 (6), 663–674. doi:10.1016/j.ajhg.2008.10.006

Zarrei, M., Macdonald, J. R., Merico, D., and Scherer, S. W. (2015). A copy number variation map of the human genome. *Nat. Rev. Genet.* 16 (3), 172–183. doi:10.1038/nrg3871