



## OPEN ACCESS

## EDITED BY

Bo Xiong,  
Huazhong University of Science  
and Technology, China

## REVIEWED BY

Yi Zhang,  
Central South University, China  
Yan Wang,  
Beijing Institutes of Life Science (CAS),  
China

## \*CORRESPONDENCE

Hong Yao  
yaohong6319@sina.com  
Bo Tan  
tanbo@hospital.cqmu.edu.cn

†These authors have contributed  
equally to this work and share first  
authorship

## SPECIALTY SECTION

This article was submitted to  
Neurogenomics,  
a section of the journal  
Frontiers in Neuroscience

RECEIVED 28 June 2022

ACCEPTED 11 August 2022

PUBLISHED 07 September 2022

## CITATION

Liu L, Feng X, Liu S, Zhou Y, Dong X,  
Yao H and Tan B (2022)  
Whole-genome sequencing combined  
RNA-sequencing analysis of patients  
with mutations in SET binding  
protein 1.  
*Front. Neurosci.* 16:980000.  
doi: 10.3389/fnins.2022.980000

## COPYRIGHT

© 2022 Liu, Feng, Liu, Zhou, Dong, Yao  
and Tan. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Whole-genome sequencing combined RNA-sequencing analysis of patients with mutations in SET binding protein 1

Li Liu<sup>1†</sup>, Xiaoshu Feng<sup>2†</sup>, Sihan Liu<sup>2†</sup>, Yanqiu Zhou<sup>1</sup>,  
Xiaojing Dong<sup>1</sup>, Hong Yao<sup>1\*</sup> and Bo Tan<sup>1\*</sup>

<sup>1</sup>Department of Gynecology and Obstetrics, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, <sup>2</sup>Institute of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China

SET binding protein 1 (SETBP1) is essential for human development, and pathogenic germline variants in *SETBP1* lead to a recognizable developmental syndrome and variable clinical features. In this study, we assessed a patient with facial dysmorphism, intellectual disability and delayed motor development. Whole genome sequencing identified a novel *de novo* variation of the *SETBP1* (c.2631C > A; p. S877R) gene, which is located in the SKI domain, as a likely pathogenic variant for the proband's phenotype. RNA sequencing was performed to investigate the potential molecular mechanism of the novel variation in *SETBP1*. In total, 77 and 38 genes were identified with aberrant expression and splicing, respectively. Moreover, the biological functions of these genes were involved in DNA/protein binding, expression regulation, and the cell cycle, which may advance our understanding of the pathogenesis of *SETBP1 in vivo*.

## KEYWORDS

RNA-seq, *de novo*, missense variant, *SETBP1*, clinical diagnosis

## Introduction

The SET binding protein 1 (*SETBP1*) gene is an oncogene located on the long (q) arm of chromosome 18 at position 12.3. The protein encoded by the *SETBP1* gene contains several motifs and has been shown to bind the SET nuclear oncogene, which is associated with DNA replication and gene expression regulation (Piazza et al., 2018). Mutations in *SETBP1* are involved in multiple diseases, leading to extremely complex genotype-phenotype correlations for the *SETBP1* gene (Acuna-Hidalgo et al., 2017). Somatic mutations of *SETBP1* appear to be gain-of-function mutations and are associated with several hematological malignancies, such as myeloid leukemia (Makishima et al., 2013; Meggendorfer et al., 2013; Piazza et al., 2013; Sakaguchi et al., 2013; Thol et al., 2013;

Fabiani et al., 2014; Patnaik et al., 2014; Inoue et al., 2015). In addition, germline loss-of-function mutations in the *SETBP1* gene are correlated with developmental delay, which has a spectrum of symptoms, including absent speech/expressive language delays and mild-severe intellectual disability (Filges et al., 2011). In contrast, germline gain-of-function mutations in the *SETBP1* gene are linked with Schinzel–Giedion syndrome (SGS; OMIM 269150) (Acuna-Hidalgo et al., 2017; Leonardi et al., 2020).

SGS is a rare genetic disorder characterized by characteristic facial features, multiple malformations, and neurological problems (Schinzel and Giedion, 1978; Minn et al., 2002; Al-Mudaffer et al., 2008). Germline *de novo* mutations in the *SETBP1* gene cluster to a hotspot of 12 base pairs in exon 4 of the *SETBP1* protein cause SGS (Hoischen et al., 2010). This mutational hotspot is highly conserved and is part of a degron motif targeted by the SCF- $\beta$ TrCP1 E3 ligase (Piazza et al., 2013). Previous studies have demonstrated that somatic mutations occurring in the hotspot region in the *SETBP1* gene may cause a functional loss of the degron motif, resulting in accumulation of *SETBP1* protein in cells and inhibition of the PP2A phosphatase through the *SETBP1*–*SET*–PP2A axis (Cristobal et al., 2010; Oakley et al., 2012; Makishima et al., 2013). However, the understanding of the pathogenic mechanism of germline mutations in the *SETBP1* gene is inadequate. Additional clinical and functional investigation is warranted to promote our understanding of the molecular mechanisms of SGS.

Recently, several studies have implemented total RNA sequencing integrated with whole-genome sequencing (WGS) to facilitate interpretation of the pathogenicity of variants by revealing expression and splicing outliers (Kremer et al., 2017; Hollein et al., 2020; Peymani et al., 2022; Yopez et al., 2022). This approach provides an opportunity to explore the molecular mechanisms of germline mutations in the *SETBP1* gene. In this study, we present the clinical characterization of a patient diagnosis as SGS and conducted WGS of parent-offspring trio. The results revealed a novel *de novo* mutation in *SETBP1* that was predicted to be deleterious based on the concordance of generic damage prediction tools. Furthermore, RNA sequencing was performed in this family, and numerous aberrant expression/splicing genes provided supporting evidence for the role of *SETBP1* and insight into the molecular mechanisms of germline mutations in the *SETBP1* gene.

## Materials and methods

### Ethical compliance

Informed consent was obtained from the patient's parents. This study was approved by the ethics committee of the Second Affiliated Hospital of Chongqing Medical University.

## DNA isolation and whole genome sequencing

We sequenced the patient and her parents following the MGI-2000 protocol outsourced to BGI. Genomic DNA was isolated from peripheral blood using a blood genomic DNA extraction kit (Tiangen Biotech, Beijing, China) in accordance with the manufacturer's protocol. One microgram of genomic DNA was randomly fragmented by Covaris, and the fragmented DNA was selected by an Agencourt AMPure XP-Medium kit to an average size of 200–400 bp, followed by adapter ligation and PCR amplification. The products were recovered by the AxyPrep Mag PCR clean up Kit. The double-stranded PCR products were heat-denatured and circularized by the splint oligo sequence. The single-strand circle DNA (ssCir DNA) was formatted as the final library and qualified by QC. WGS was performed on the MGI-2000 platform with an average depth of 30x, meaning that the entire genome was sequenced an average of 30 times.

## RNA isolation and sequencing

Total RNA was extracted from peripheral blood and enriched by oligo-dT bead capture, and cDNA was synthesized according to the manufacturer's protocol. cDNA libraries were constructed using the Illumina trueSeq stranded mRNA sample prep kit protocol (Illumina). Pooled samples were sequenced using a NovaSeq 6000 sequencing system.

## Single-nucleotide variant/INDEL identification, annotation and interpretation

The raw data produced on the MGI-2000 platform were filtered and aligned against the human reference genome (hg19) using the Burrows–Wheeler Alignment tool (Li and Durbin, 2009) after evaluation according to Illumina Sequence Control Software (SCS). The single-nucleotide polymorphisms (SNPs) were called by using Genome Analysis ToolKit software (Van der Auwera et al., 2013).

Variants were annotated using ANNOVAR (Wang et al., 2010). The effects of single-nucleotide variants (SNVs) were predicted by the SIFT, Polyphen-2, and MutationTaster programs. Variants were filtered by a minor allele frequency (MAF) of < 0.1% in the gnomAD (Karczewski et al., 2020), 1000 Genome (Genomes Project et al., 2015), ExAC (Lek et al., 2016) databases and the Exome Variant Server (EVS; NHLBI Exome Sequencing Project).

All variants were interpreted according to ACMG/AMP standards and categorized as pathogenic, likely pathogenic, variants of unknown clinical significance (VUS), likely benign and benign (Richards et al., 2015). Variant validation

was performed using Sanger sequencing (ABI 3730xl Genetic Analyzer).

## Copy number variation identification and annotation

Copy number variations (CNVs) were detected by CNVnator, 100-bp bins and standard parameters were used to calculate the read-depth (RD) signal (Abyzov et al., 2011). The CNVs identified were compared with CNVs from the Database of Genomic Variants<sup>1</sup> to exclude previously reported polymorphisms. The non-polymorphic CNVs were compared with the entries in the DECIPHER,<sup>2</sup> ISCA,<sup>3</sup> ClinGen<sup>4</sup>, or

1 <http://projects.tcag.ca/variation/>

2 <http://decipher.sanger.ac.uk/>

3 <http://www.iccg.org/>

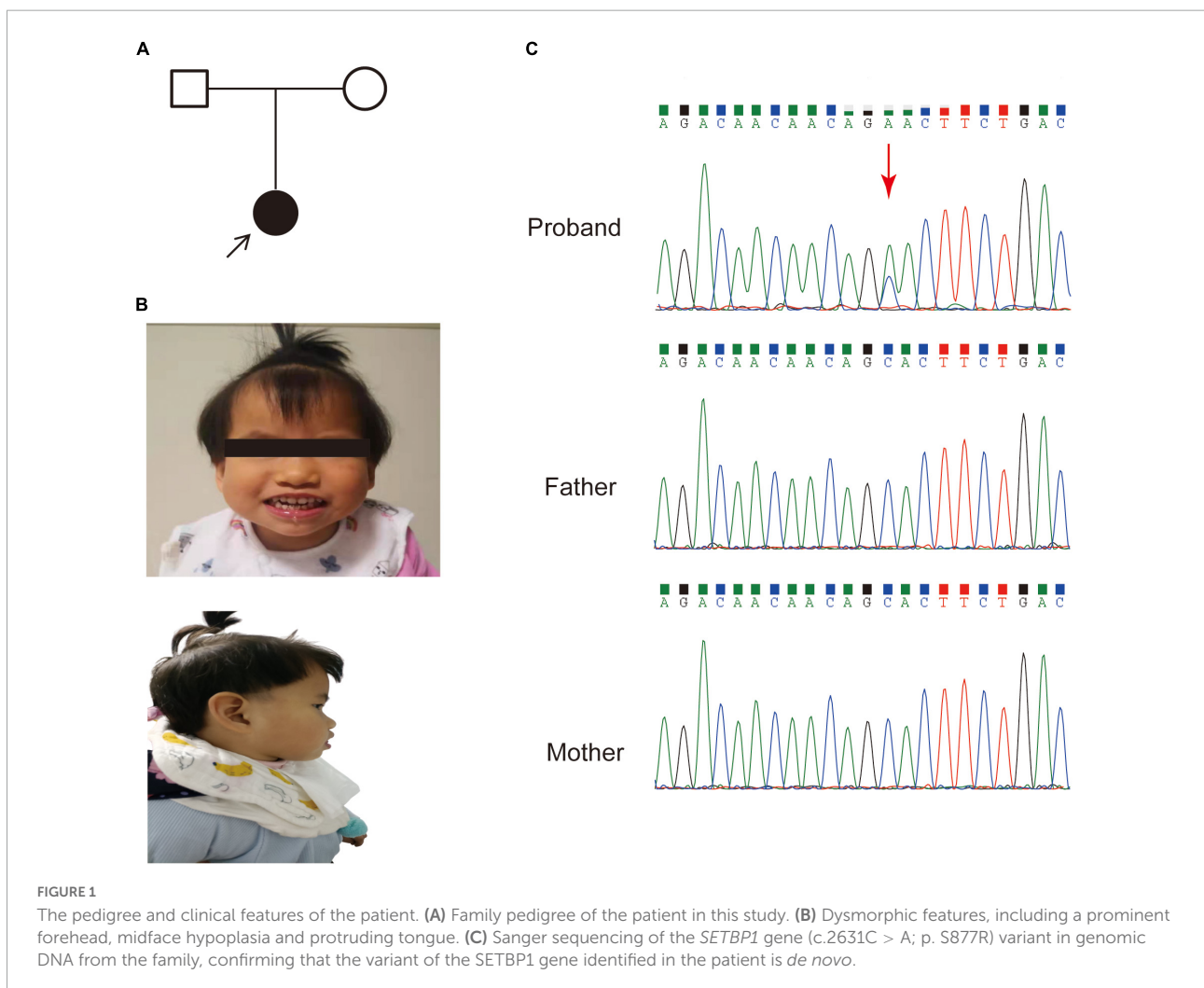
4 <https://search.clinicalgenome.org/kb/gene-dosage>

ClinVar<sup>5</sup> databases, evaluated against the literature for known syndromes and overlapping causal aberrations and further analyzed according to the type and size of aberration, function, and expression profile of genes.

## Quality control for RNA-seq data

Fastp was used to filter low-quality reads from raw sequencing reads to obtain clean reads (Chen et al., 2018b). Then, FastQC and multiQC were used to evaluate the quality of sequencing data, and the average quality score for overall RNA sequences was > 30, indicating that a large percentage of the sequences were high quality (Ewels et al., 2016). DROP v1.2.1 was used to compute the evaluation metrics of mapping with sequencing depth, percentage of mapped reads, and the number of expressed genes (Yepez et al., 2021). The match between the RNA-seq sample and its annotated DNA sample was also determined by DROP with a cutoff of 0.8.

5 <http://www.ncbi.nlm.nih.gov/clinvar/>



### Detection of aberrant expression

Aberrant expression was fully detected based on DROP v1.2.1 (Yepez et al., 2021). The clean RNA-sequencing reads were mapped to the human reference genome (hg19) using STAR (2.7.8a) with the Gencode v29 annotation (Dobin et al., 2013). The summarize Overlaps function from the Genomic Alignments R package was used to count reads. To increase statistical power, we performed aberrant expression and splicing analysis by combining our data with 367 blood samples from GTEx data.<sup>6</sup> Genes with a 95th percentile FPKM (Fragments Per Kilobase of transcript per Million mapped reads) < 1 were considered as lowly expressed in samples and were removed in downstream analysis. In total, nearly 10,000 genes were included. OUTRIDER was applied to identify expression outliers (Brechtmann et al., 2018). Technical and biological covariates, such as sex, age and sequencing batch, were automatically controlled by OUTRIDER, which used an autoencoder implementation. Genes were defined as having aberrant expression with a  $p < 0.01$ . Reverse transcription-quantitative

PCR (RT-qPCR) was performed to validate candidate gene expression.

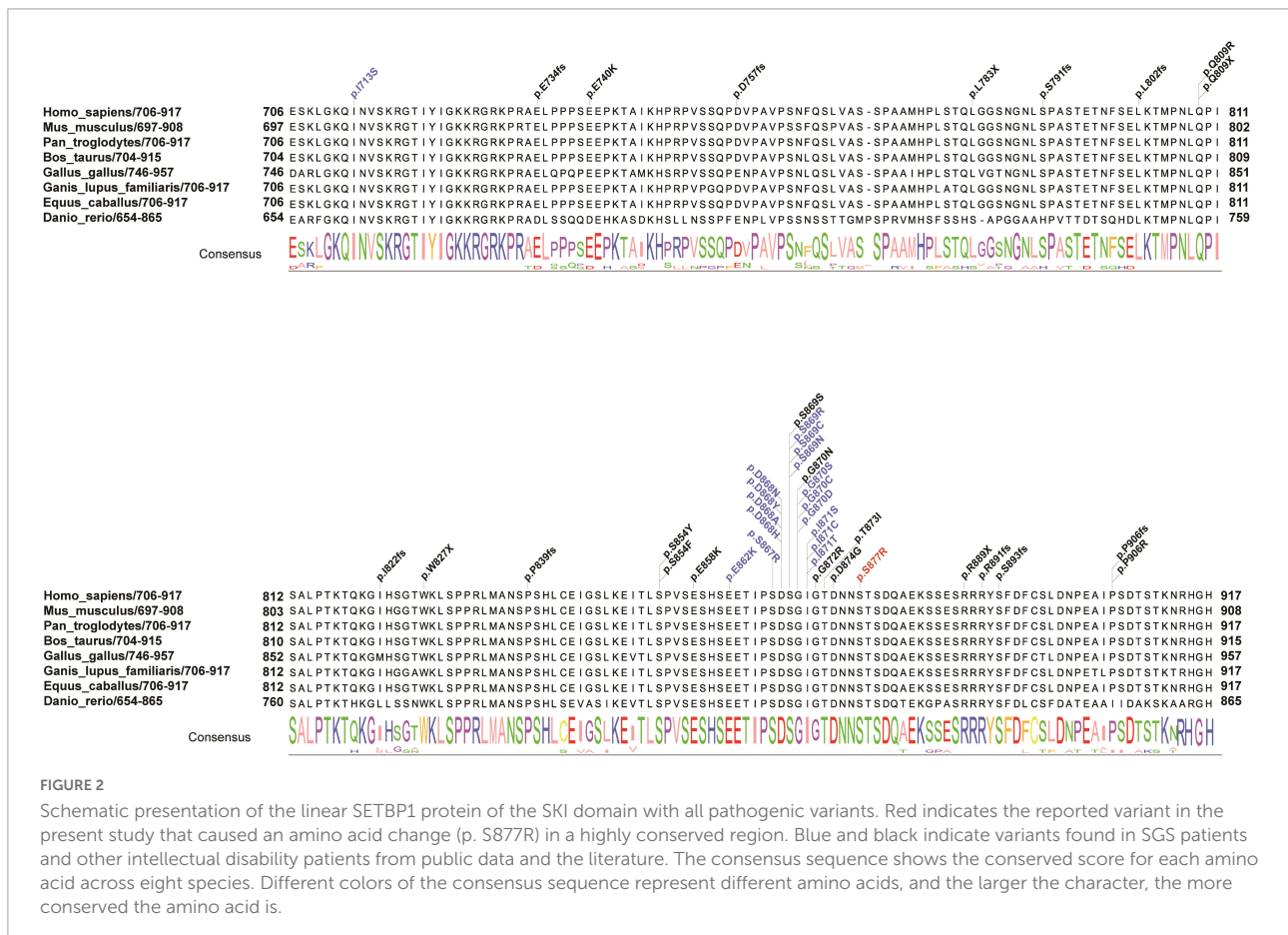
### Detection of aberrant splicing

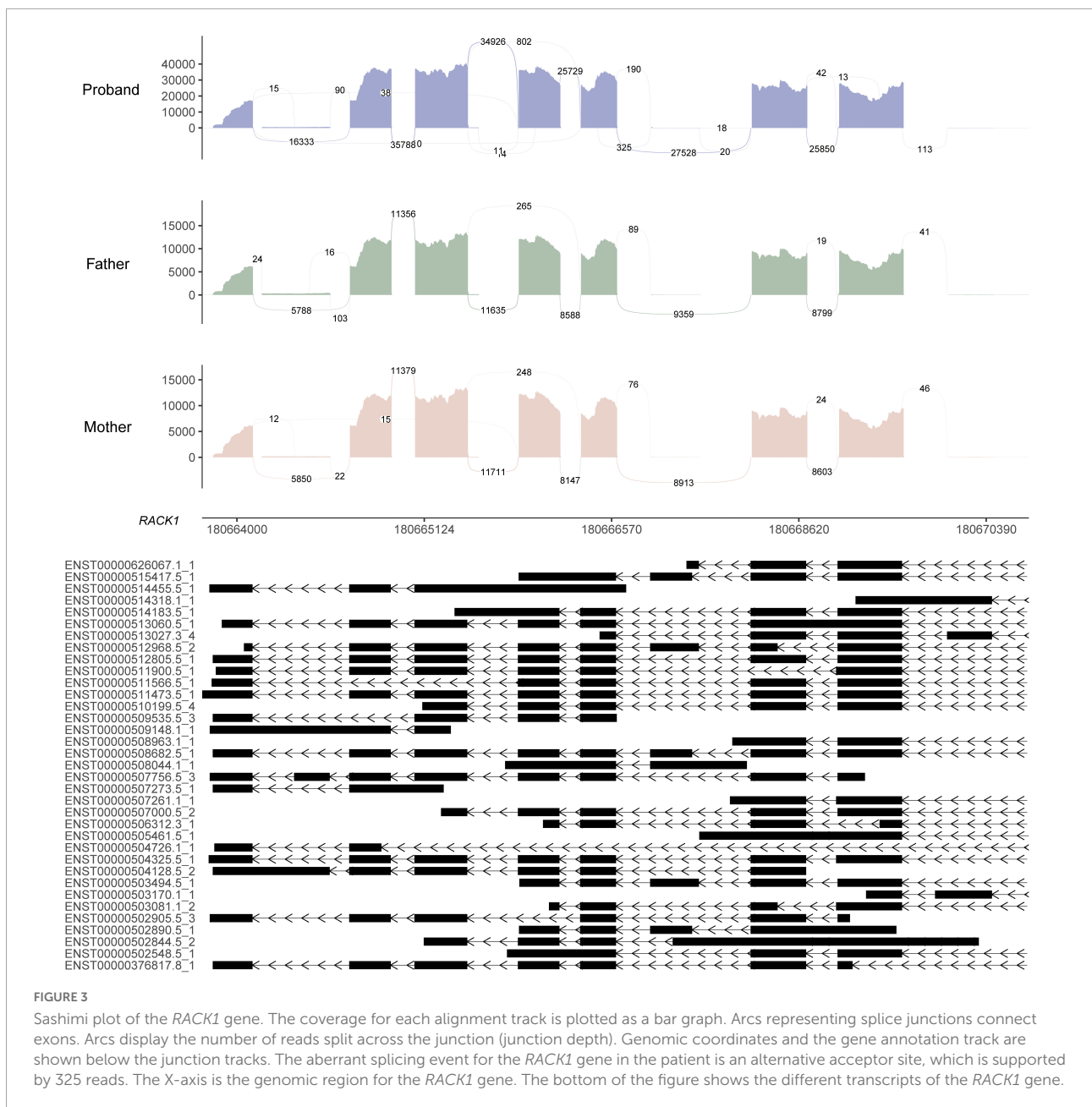
FRASER, which has been included in DROP, was used to obtain splicing outliers (Mertes et al., 2021). Exon-exon and exon-intron junctions with less than 20 reads in all samples were filtered out. In addition, junctions in which the total number of reads at the donor/acceptor splice site was 0 in more than 90% of the samples were also filtered out. Similar to OUTRIDER, FRASER also applies an autoencoder implementation to automatically control the technical and biological covariates. Splicing outlier genes were defined as genes with an adjusted  $p < 0.05$ . Outlier junctions were defined as those in splicing outlier genes, with an adjusted  $p < 0.05$ .

### Pathway enrichment analysis

Functional enrichment of the aberrantly expressed and spliced genes was performed with KOBAS-i, a service that provides comprehensive pathway enrichment analysis using several databases, including GO, KEGG, Reactome, and GWAS catalogs (Bu et al., 2021). An adjusted  $p < 0.1$  was selected as the threshold for significant pathways.

<sup>6</sup> [https://zenodo.org/record/5638707#\\_Yt4B63ZByUk](https://zenodo.org/record/5638707#_Yt4B63ZByUk)





**FIGURE 3**  
Sashimi plot of the *RACK1* gene. The coverage for each alignment track is plotted as a bar graph. Arcs representing splice junctions connect exons. Arcs display the number of reads split across the junction (junction depth). Genomic coordinates and the gene annotation track are shown below the junction tracks. The aberrant splicing event for the *RACK1* gene in the patient is an alternative acceptor site, which is supported by 325 reads. The X-axis is the genomic region for the *RACK1* gene. The bottom of the figure shows the different transcripts of the *RACK1* gene.

## Results

### Clinical features of the patient

A 3-year-old female was referred to our hospital with global developmental delay, hypertonia and facial dysmorphism. The patient was born after 39 weeks with a normal gestation history. Her parents had no medical history (Figure 1A). She was found to have a motor and language development delay at 2 years old. She had characteristic facial features, including microcephaly, a prominent forehead, midface hypoplasia, a high palatal arch and a protruding tongue (Figure 1B).

Brain MRI at 8 months of age showed delayed myelination of brain white matter and enlargement of the lateral ventricle, and the bilateral frontotemporal extracerebral space was significantly widened. Her karyotype analysis revealed normal results.

### Whole genome sequencing analysis

An average of 102G sequencing data were acquired after WGS for the family member, and no pathogenic CNVs were detected in the proband's WGS data (Supplementary Figure 1).

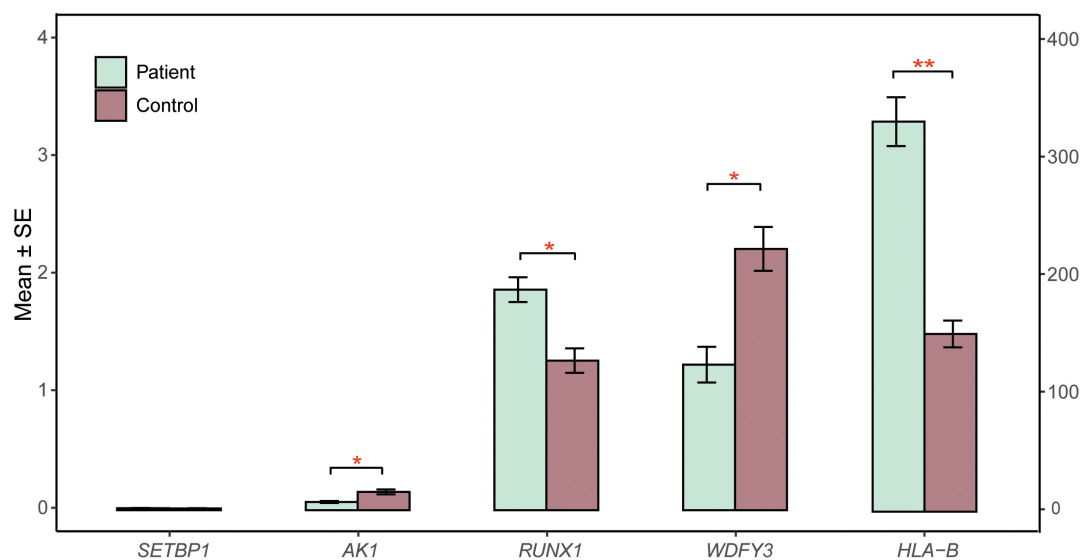


FIGURE 4

RT-qPCR validation. The relative expression of candidate genes (*SETBP1*, *WDFY3*, *AK1*, *RUNX1*, *HLA-B*) in patient and control blood cells was quantified by RT-qPCR.  $\beta$ -Actin mRNA was used as an internal control. The experiments were performed in triplicate. Data are shown as the mean  $\pm$  SE. Two sample *t*-test was performed to test whether the expression of these genes between case and control are significant difference or not. Statistical significance is presented relative to control as \* $P < 0.05$  and \*\* $P < 0.01$ . The X-axis on the left presents the mean expression of the *SETBP1*, *WDFY3*, *AK1*, and *RUNX1* genes. The X-axis on the right presents the mean expression of the *HLA-B* gene. SE, standard error.

After variant pathogenicity classification according to ACMG guidelines, one *de novo* missense variant located in the *SETBP1* gene (PS2 + PM2\_Supporting + PP3), NM\_015559.2: g.42531936C > A, c.2631C > A (NM\_000052.7), affected highly conserved residues in close proximity to the canonical region in the SKI domain (p. S877R shown in red in Figure 2), indicating a likely pathogenic variant for the proband's phenotype. Sanger sequencing confirmed that the c.2631C > A variant was a novel *de novo* variant (Figure 1C and Supplementary Figure 2). In addition, this variant has never been reported in the ClinVar database, HGMD database, or gnomAD database before.

## Transcriptome analysis

RNA sequencing was performed to investigate the potential molecular mechanism of the novel mutation in *SETBP1* (c.2631C > A; p.S877R). Aberrant analysis results and RT-qPCR showed that the RNA expression of the *SETBP1* gene in the patient and control was similar. In total, 77 and 38 genes were identified with aberrant expression and splicing in the patient, respectively (Supplementary Tables 1, 2 and Supplementary Figures 3–8). Several genes directly targeted by *SETBP1* or associated with neurodevelopmental disorders (NDDs) have been identified as aberrant genes in patient. For example, the receptor for activated C kinase 1 (*RACK1*) gene and RUNX Family Transcription Factor 1 (*RUNX1*) gene. *RACK1* was a part of the IRE1-RACK1-PP2A complex and was

aberrantly spliced in the patient (Figure 3). The *RACK1* gene can modulate neurodegeneration by promoting ERK degradation in Machado-Joseph disease (MJD) and Huntington's disease (HD) models and participates in the process of neuronal differentiation by regulating *SCN1A* expression (Adams et al., 2011; Dong et al., 2014; Xie et al., 2021). *RUNX1* is a direct transcriptional target of *SETBP1* and encodes a transcription factor involved in the generation of hematopoietic stem cells and their differentiation into myeloid and lymphoid lines (Vishwakarma et al., 2016). Relative quantification of a subset of genes (*SETBP1*, *WDFY3*, *AK1*, *RUNX1*, and *HLA-B*) by means of RT-qPCR confirmed the accuracy of aberrant analysis with RNA-seq data (Figure 4; two sample *t*-tests;  $P_{SETBP1} = 0.485$ ;  $P_{WDFY3} = 0.015$ ;  $P_{AK1} = 0.037$ ;  $P_{RUNX1} = 0.015$  and  $P_{HLA-B} = 0.004$ ). We next performed enrichment analyses of the aberrantly expressed and spliced genes to delineate the most relevant biological pathways. Functional annotation demonstrated that the biological functions of these genes were involved in DNA/protein binding, expression regulation, and the cell cycle (Figure 5 and Supplementary Tables 3, 4).

## Mutation pattern and genotype-phenotype correlations in the SKI domain of *SETBP1*

To evaluate the correlation between associated variants in the SKI domain and the phenotype of *SETBP1*, a systematic

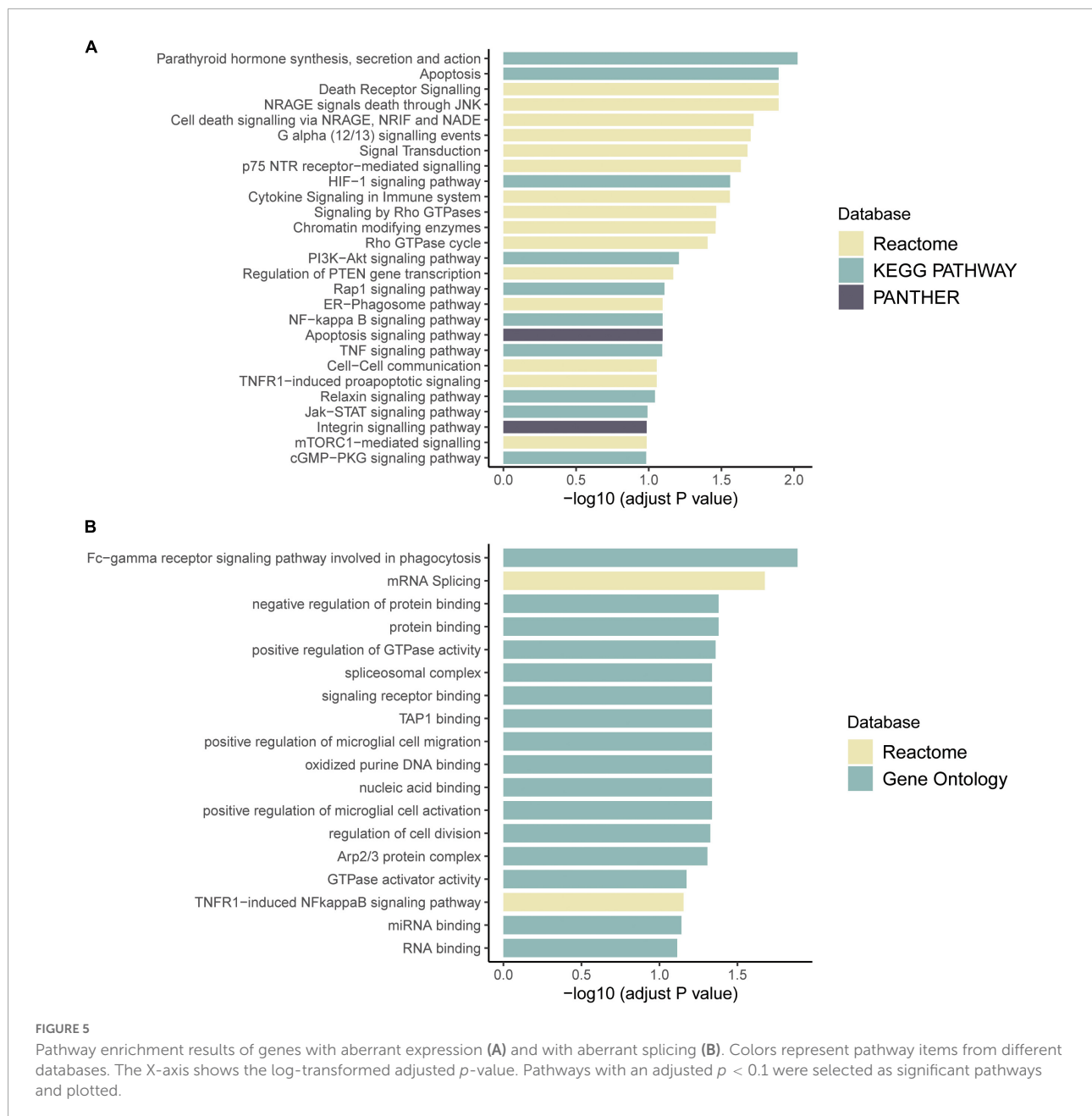


FIGURE 5

Pathway enrichment results of genes with aberrant expression (A) and with aberrant splicing (B). Colors represent pathway items from different databases. The X-axis shows the log-transformed adjusted  $p$ -value. Pathways with an adjusted  $p < 0.1$  were selected as significant pathways and plotted.

search of original papers was performed (Hoischen et al., 2010; Filges et al., 2011; De Rubeis et al., 2014; Carvalho et al., 2015; Miyake et al., 2015; Takeuchi et al., 2015; Volk et al., 2015; Li et al., 2016; Acuna-Hidalgo et al., 2017; Stessman et al., 2017; Chen et al., 2018a; Daum et al., 2019; Hildebrand et al., 2020; Kurtz-Nelson et al., 2020; Leonardi et al., 2020). A total of 41 variants with *SETBP1*-related NDDs located in the SKI domain were included, including 28 with missense mutations and 13 with loss-of-function mutations (Supplementary Table 5). Overall, variants clustering to a hotspot of 12 base pairs coding for residues 868 to 871 of the *SETBP1* protein are known to be associated with severe forms of SGS, possibly through a

dominant negative effect (Hoischen et al., 2010). In our study, a novel variant near this region was associated with a similar form of this disease (Figure 2).

## Discussion

In this study, we reported a patient of SGS with severe intellectual disability, developmental delay, epilepsy, hypertonia and distinctive facial dysmorphism. WGS and Sanger validation revealed that these phenotypes may be caused by a novel *de novo* germline missense mutation of the *SETBP1* gene

(NM\_015559.2: c.2631C > A), which can cause the amino acid change p. S877R (serine - arginine). Due to the extremely low prevalence and great phenotypic heterogeneity of SGS, it is difficult to recognize by clinicians and is usually diagnosed based on the reminiscent clinical features (Acuna-Hidalgo et al., 2017). In this case, an accurate genetic diagnosis significantly improved the management of the patient and reproduction of this family.

Our results revealed that germline *de novo* heterozygous missense variants adjacent to the mutation hotspot of the *SETBP1* gene tend to cause atypical SGS. Previous studies have shown that germline *de novo* mutations in the *SETBP1* gene cluster to residues 868–871 of the SETBP1 protein, which are associated with severe forms of SGS (Hoischen et al., 2010). In this study, the *de novo* variant identified in this patient is located in residue 877, which is close to the mutation hotspot of the *SETBP1* gene. Some phenotypes in this patient were mild relative to typical SGS patients, including ventriculomegaly, skeletal abnormalities, and hydronephrosis. In addition, several individuals with atypical SGS carrying heterozygous missense variants outside the mutation hotspot have been reported. Acuna-Hidalgo et al. (2017) identified four individuals carrying *SETBP1* variants in close proximity to the canonical mutation hotspot, including p.(Glu862Lys), p.(Ser867Arg), and p.(Thr873Ile), who showed a milder developmental phenotype with clinical characteristics that partially overlapped with classical SGS. Moreover, individuals with variants located further from the mutation hotspot showed a variable clinical phenotype ranging from mild to severe intellectual disability (Leonardi et al., 2020; Wong et al., 2022). These findings highlight that the variable severity of broad clinical features depends on the proximity of variants to the mutation hotspot.

Integrative analyses identified that the missense variant reported in this study likely disrupts SETBP1 protein functions via mechanisms including DNA/protein binding, transcription and the cell cycle. While the transcription of the *SETBP1* gene was not affected, it is possible that this missense *SETBP1* mutation has a subtle but distinct effect on the regulation, since 77 aberrantly expressed and 38 spliced genes have been identified in the patient. Consistent findings have been reported by other researchers. In a recent paper, Wong et al. revealed through cellular experiments that classical SGS variants located in the mutation hotspot showed increased protein stability and higher SETBP1 protein levels, while *SETBP1* variants outside the mutation hotspot disrupt DNA binding and transcription independent of protein abundance (Wong et al., 2022). Future studies that delineate the structural impact of *SETBP1* variants and how they affect interactions with other genes will contribute to the understanding of their impacts on protein functions and thus etiology.

Taken together, our findings expand the current understanding of the genetics and clinical spectrum of

*SETBP1* variants. In addition, by integrating WGS and RNA-seq analyses, we provide insight into the pathogenicity of a germline *de novo* *SETBP1* variant in a patient diagnosed with atypical SGS.

## Data availability statement

The original contributions presented in this study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee at The Second Affiliated Hospital of Chongqing Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

## Author contributions

LL performed the experiments and wrote the manuscript. XF and SL performed RNA-seq analysis and wrote the manuscript. YZ and XD collected the clinical information of the patient and her parents. HY and BT designed and supervised the study and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Joint International Research Lab for Reproduction and Development, Ministry of Education, China, Program for Youth Innovation in Future Medicine, Chongqing Medical University (W0122), "Kuanren talents" project of the Second Affiliated Hospital of Chongqing Medical University (13-003-003), Nan'an District Science and Health Joint Medical Scientific Research Project (2020-01), and the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. 81).

## Acknowledgments

We thank the patient and the family members for their participation in this study.



## 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.

## 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/fnins.2022.980000/full#supplementary-material>

## References

- Abyzov, A., Urban, A. E., Snyder, M., and Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 21, 974–984. doi: 10.1101/gr.114876.110
- Acuna-Hidalgo, R., Deriziotis, P., Steehouwer, M., Gilissen, C., Graham, S. A., van Dam, S., et al. (2017). Overlapping SETBP1 gain-of-function mutations in Schinzel-Giedion syndrome and hematologic malignancies. *PLoS Genet.* 13:e1006683. doi: 10.1371/journal.pgen.1006683
- Adams, D. R., Ron, D., and Kiely, P. A. (2011). RACK1, A multifaceted scaffolding protein: Structure and function. *Cell Commun. Signal.* 9:22. doi: 10.1186/1478-811X-9-22
- Al-Mudaffer, M., Oley, C., Price, S., Hayes, I., Stewart, A., Hall, C. M., et al. (2008). Clinical and radiological findings in Schinzel-Giedion syndrome. *Eur. J. Pediatr.* 167, 1399–1407. doi: 10.1007/s00431-008-0683-4
- Brechtmann, F., Mertes, C., Matusciuciu, A., Yopez, V. A., Avsec, Z., Herzog, M., et al. (2018). OUTRIDER: A Statistical Method for Detecting Aberrantly Expressed Genes in RNA Sequencing Data. *Am. J. Hum. Genet.* 103, 907–917. doi: 10.1016/j.ajhg.2018.10.025
- 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
- Carvalho, E., Honjo, R., Magalhaes, M., Yamamoto, G., Rocha, K., Naslavsky, M., et al. (2015). Schinzel-Giedion syndrome in two Brazilian patients: Report of a novel mutation in SETBP1 and literature review of the clinical features. *Am. J. Med. Genet. A* 167A, 1039–1046. doi: 10.1002/ajmg.a.36789
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018b). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, S., Fragoza, R., Klei, L., Liu, Y., Wang, J., Roeder, K., et al. (2018a). An interactome perturbation framework prioritizes damaging missense mutations for developmental disorders. *Nat. Genet.* 50, 1032–1040. doi: 10.1038/s41588-018-0130-z
- Cristobal, I., Blanco, F. J., Garcia-Orti, L., Marcotegui, N., Vicente, C., Rifon, J., et al. (2010). SETBP1 overexpression is a novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute myeloid leukemia. *Blood* 115, 615–625. doi: 10.1182/blood-2009-06-227363
- Daum, H., Meiner, V., Elpeleg, O., Harel, T., and Collaborating, A. (2019). Fetal exome sequencing: Yield and limitations in a tertiary referral center. *Ultrasound Obstet. Gynecol.* 53, 80–86. doi: 10.1002/uog.19168
- De Rubeis, S., He, X., Goldberg, A. P., Poultney, C. S., Samocha, K., Cicek, A. E., et al. (2014). Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 515, 209–215. doi: 10.1038/nature13772
- 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
- Dong, Z. F., Tang, L. J., Deng, G. F., Zeng, T., Liu, S. J., Wan, R. P., et al. (2014). Transcription of the human sodium channel SCN1A gene is repressed by a scaffolding protein RACK1. *Mol. Neurobiol.* 50, 438–448. doi: 10.1007/s12035-014-8633-9
- Ewels, P., Magnusson, M., Lundin, S., and Kaller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048. doi: 10.1093/bioinformatics/btw354
- Fabiani, E., Falconi, G., Fianchi, L., Criscuolo, M., Leone, G., and Voso, M. T. (2014). SETBP1 mutations in 106 patients with therapy-related myeloid neoplasms. *Haematologica* 99:e152–e153. doi: 10.3324/haematol.2014.108159
- Filges, I., Shimojima, K., Okamoto, N., Rothlisberger, B., Weber, P., Huber, A. R., et al. (2011). Reduced expression by SETBP1 haploinsufficiency causes developmental and expressive language delay indicating a phenotype distinct from Schinzel-Giedion syndrome. *J. Med. Genet.* 48, 117–122. doi: 10.1136/jmg.2010.084582
- Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393
- Hildebrand, M. S., Jackson, V. E., Scerri, T. S., Van Reyk, O., Coleman, M., Braden, R. O., et al. (2020). Severe childhood speech disorder: Gene discovery highlights transcriptional dysregulation. *Neurology* 94:e2148–e2167. doi: 10.1212/WNL.00000000000009441
- Hoischen, A., van Bon, B. W., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., et al. (2010). De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* 42, 483–485. doi: 10.1038/ng.581
- Hollein, A., Twardziok, S. O., Walter, W., Hutter, S., Baer, C., Hernandez-Sanchez, J. M., et al. (2020). The combination of WGS and RNA-Seq is superior to conventional diagnostic tests in multiple myeloma: Ready for prime time? *Cancer Genet.* 242, 15–24. doi: 10.1016/j.cancergen.2020.01.001
- Inoue, D., Kitaura, J., Matsui, H., Hou, H. A., Chou, W. C., Nagamachi, A., et al. (2015). SETBP1 mutations drive leukemic transformation in ASXL1-mutated MDS. *Leukemia* 29, 847–857. doi: 10.1038/leu.2014.301
- Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. doi: 10.1038/s41586-020-2308-7
- Kremer, L. S., Bader, D. M., Mertes, C., Kopajtic, R., Pichler, G., Iuso, A., et al. (2017). Genetic diagnosis of Mendelian disorders via RNA sequencing. *Nat. Commun.* 8:15824. doi: 10.1038/ncomms15824
- Kurtz-Nelson, E. C., Beighley, J. S., Hudac, C. M., Gerdt, J., Wallace, A. S., Hoekzema, K., et al. (2020). Co-occurring medical conditions among individuals with ASD-associated disruptive mutations. *Child. Health Care* 49, 361–384. doi: 10.1080/02739615.2020.1741361
- 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
- Leonardi, E., Bettella, E., Pelizza, M. F., Aspromonte, M. C., Polli, R., Boniver, C., et al. (2020). Identification of SETBP1 Mutations by Gene Panel Sequencing in Individuals With Intellectual Disability or With “Developmental and Epileptic Encephalopathy”. *Front. Neurol.* 11:593446. doi: 10.3389/fneur.2020.593446
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

- Li, J., Cai, T., Jiang, Y., Chen, H., He, X., Chen, C., et al. (2016). Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. *Mol. Psychiatry* 21, 290–297. doi: 10.1038/mp.2015.40
- Makishima, H., Yoshida, K., Nguyen, N., Przychodzen, B., Sanada, M., Okuno, Y., et al. (2013). Somatic SETBP1 mutations in myeloid malignancies. *Nat. Genet.* 45, 942–946. doi: 10.1038/ng.2696
- Meggendorfer, M., Bacher, U., Alpermann, T., Haferlach, C., Kern, W., Gambacorti-Passerini, C., et al. (2013). SETBP1 mutations occur in 9% of MDS/MPN and in 4% of MPN cases and are strongly associated with atypical CML, monosomy 7, isochromosome i(17)(q10), ASXL1 and CBL mutations. *Leukemia* 27, 1852–1860. doi: 10.1038/leu.2013.133
- Mertes, C., Scheller, I. F., Yezpez, V. A., Celik, M. H., Liang, Y., Kremer, L. S., et al. (2021). Detection of aberrant splicing events in RNA-seq data using FRASER. *Nat. Commun.* 12:529. doi: 10.1038/s41467-020-20573-7
- Minn, D., Christmann, D., De Saint-Martin, A., Alembik, Y., Eliot, M., Mack, G., et al. (2002). Further clinical and sensorial delineation of Schinzel-Giedion syndrome: Report of two cases. *Am. J. Med. Genet.* 109, 211–217. doi: 10.1002/ajmg.10348
- Miyake, F., Kuroda, Y., Naruto, T., Ohashi, I., Takano, K., and Kurosawa, K. (2015). West syndrome in a patient with Schinzel-Giedion syndrome. *J. Child Neurol.* 30, 932–936. doi: 10.1177/0883073814541468
- Oakley, K., Han, Y., Vishwakarma, B. A., Chu, S., Bhatia, R., Gudmundsson, K. O., et al. (2012). Setbp1 promotes the self-renewal of murine myeloid progenitors via activation of Hoxa9 and Hoxa10. *Blood* 119, 6099–6108. doi: 10.1182/blood-2011-10-388710
- Patnaik, M. M., Itzykson, R., Lasho, T. L., Kosmider, O., Finke, C. M., Hanson, C. A., et al. (2014). ASXL1 and SETBP1 mutations and their prognostic contribution in chronic myelomonocytic leukemia: A two-center study of 466 patients. *Leukemia* 28, 2206–2212. doi: 10.1038/leu.2014.125
- Peymani, F., Farzeen, A., and Prokisch, H. (2022). RNA sequencing role and application in clinical diagnostic. *Pediatr. Investig.* 6, 29–35. doi: 10.1002/ped4.12314
- Piazza, R., Magistroni, V., Redaelli, S., Mauri, M., Massimino, L., Sessa, A., et al. (2018). SETBP1 induces transcription of a network of development genes by acting as an epigenetic hub. *Nat. Commun.* 9:2192. doi: 10.1038/s41467-018-04462-8
- Piazza, R., Valletta, S., Winkelmann, N., Redaelli, S., Spinelli, R., Pirola, A., et al. (2013). Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat. Genet.* 45, 18–24. doi: 10.1038/ng.2495
- 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
- Sakaguchi, H., Okuno, Y., Muramatsu, H., Yoshida, K., Shiraishi, Y., Takahashi, M., et al. (2013). Exome sequencing identifies secondary mutations of SETBP1 and JAK3 in juvenile myelomonocytic leukemia. *Nat. Genet.* 45, 937–941. doi: 10.1038/ng.2698
- Schinzel, A., and Giedion, A. (1978). A syndrome of severe midface retraction, multiple skull anomalies, clubfeet, and cardiac and renal malformations in sibs. *Am. J. Med. Genet.* 1, 361–375. doi: 10.1002/ajmg.1320010402
- Stessman, H. A., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., et al. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat. Genet.* 49, 515–526. doi: 10.1038/ng.3792
- Takeuchi, A., Okamoto, N., Fujinaga, S., Morita, H., Shimizu, J., Akiyama, T., et al. (2015). Progressive brain atrophy in Schinzel-Giedion syndrome with a SETBP1 mutation. *Eur. J. Med. Genet.* 58, 369–371. doi: 10.1016/j.ejmg.2015.05.006
- Thol, F., Suchanek, K. J., Koenecke, C., Stadler, M., Platzbecker, U., Thiede, C., et al. (2013). SETBP1 mutation analysis in 944 patients with MDS and AML. *Leukemia* 27, 2072–2075. doi: 10.1038/leu.2013.145
- 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, 111011–111033. doi: 10.1002/0471250953.bi111033
- Vishwakarma, B. A., Nguyen, N., Makishima, H., Hosono, N., Gudmundsson, K. O., Negi, V., et al. (2016). Runx1 repression by histone deacetylation is critical for Setbp1-induced mouse myeloid leukemia development. *Leukemia* 30, 200–208. doi: 10.1038/leu.2015.200
- Volk, A., Conboy, E., Wical, B., Patterson, M., and Kirmani, S. (2015). Whole-Exome Sequencing in the Clinic: Lessons from Six Consecutive Cases from the Clinician's Perspective. *Mol. Syndromol.* 6, 23–31. doi: 10.1159/000371598
- 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
- Wong, M. M., Kampen, R. A., Braden, R. O., Alagöz, G., Hildebrand, M. S., Barnett, C., et al. (2022). SETBP1 variants outside the degron disrupt DNA-binding and transcription independent of protein abundance to cause a heterogeneous neurodevelopmental disorder. *medRxiv* [Preprint]. doi: 10.1101/2022.03.04.22271462
- Xie, J., Han, Y., and Wang, T. (2021). RACK1 modulates polyglutamine-induced neurodegeneration by promoting ERK degradation in Drosophila. *PLoS Genet.* 17:e1009558. doi: 10.1371/journal.pgen.1009558
- Yezpez, V. A., Gusic, M., Kopajtich, R., Mertes, C., Smith, N. H., Alston, C. L., et al. (2022). Clinical implementation of RNA sequencing for Mendelian disease diagnostics. *Genome Med.* 14:38. doi: 10.1186/s13073-022-01019-9
- Yezpez, V. A., Mertes, C., Muller, M. F., Klaproth-Andrade, D., Wachutka, L., Fresard, L., et al. (2021). Detection of aberrant gene expression events in RNA sequencing data. *Nat. Protoc.* 16, 1276–1296. doi: 10.1038/s41596-020-00462-5