frontiers | Research Topics

The genetics of neurodevelopmental disorders

Edited by
Tianyun Wang, Bo Xiong, Xiaoli Chen and Davide Vecchio

Published in
Frontiers in Genetics
Frontiers in Neuroscience

ISSN 1664-8714
ISBN 978-2-83251-871-7
DOI 10.3389/978-2-83251-871-7

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: frontiersin.org/about/contact

The genetics of neurodevelopmental disorders

Topic editors

Tianyun Wang — Peking University, China

Bo Xiong — Huazhong University of Science and Technology, China

Xiaoli Chen — Capital Institute of Pediatrics, China

Davide Vecchio — University Hospital Pediatric Department, Bambino Gesù Children's Hospital (IRCCS), Italy

Citation

Wang, T., Xiong, B., Chen, X., Vecchio, D., eds. (2023). *The genetics of neurodevelopmental disorders*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-83251-871-7

Table of contents

frontiers in Genetics

CASE REPORT
published: 29 March 2022
doi: 10.3389/fgene.2022.853028

Case Report: Exome and RNA Sequencing Identify a Novel *de novo* Missense Variant in HNRNPK in a Chinese Patient With Au-Kline Syndrome

Xin Pan[1†], *Sihan Liu*[2†], *Li Liu*[1], *Xu Zhang*[1], *Hong Yao*[1] *and Bo Tan*[1*]

[1]*Department of Gynecology and Obstetrics, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China,*
[2]*Institute of Rare Diseases, West China Hospital of Sichuan University, Chengdu, China*

OPEN ACCESS

Edited by:
Tianyun Wang,
University of Washington,
United States

Reviewed by:
Madelyn Gillentine,
Seattle Children's Hospital,
United States
Maurizio Romano,
University of Trieste, Italy
Jian Wang,
Shanghai Children's Medical Center,
China

***Correspondence:**
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 Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *12 January 2022*
Accepted: *14 March 2022*
Published: *29 March 2022*

Citation:
Pan X, Liu S, Liu L, Zhang X, Yao H and Tan B (2022) Case Report: Exome and RNA Sequencing Identify a Novel de novo Missense Variant in HNRNPK in a Chinese Patient With Au-Kline Syndrome. Front. Genet. 13:853028. doi: 10.3389/fgene.2022.853028

Au-Kline syndrome is a severe multisystemic syndrome characterized by several congenital defects, including intellectual disability. Loss-of-function and missense variants in the *HNRNPK* gene are associated with a range of dysmorphic features. This report describes an eleven-year-old Chinese boy with intellectual disability and developmental delays. Family-based whole-exome and Sanger sequencing identified a *de novo* missense variant in *HNRNPK* (NM_002140.3: c.143T > A, p. Leu48Val). In silico analysis predicted that this variant would be damaged in a highly conserved residue in the K homology 1 (KH1) domain. Bioinformatic analysis showed that the affinity change (ΔΔG) caused by this variant was -0.033 kcal/mol, indicating that it would have reduced affinity for RNA binding. Transcript analysis of the peripheral blood from this case found 42 aberrantly expressed and 86 aberrantly spliced genes (p-value <0.01). Functional enrichment analysis confirmed that the biological functions of these genes, including protein binding and transcriptional regulation, are associated with *HNRNPK*. In summary, this study identifies the first Chinese patient with a novel *de novo* heterozygous *HNRNPK* gene variant that contributes to Au-Kline syndrome and expands current knowledge of the clinical spectrum of *HNRNPK* variants.

Keywords: missense variant, Au-Kline syndrome, RNA-seq, hnRNPK, clinical diagnosis

INTRODUCTION

Au-Kline syndrome (AKS) was first described in 2015 in two unrelated boys who presented a wide spectrum of abnormalities, including atypical facial features, developmental delays, and hypotonia with intellectual disability. AKS-associated facial features include long faces, ptosis, cleft palate, and oligodontia. Genetic alterations of the heterogeneous nuclear ribonucleoprotein K (*HNRNPK*) gene are responsible for the development of AKS.

HNRNPK is a member of the RNA-binding protein family and is involved in both physiological and pathological processes, including spermatogenesis, nervous system and ovary development, erythroid differentiation, organogenesis, and carcinogenesis (Barboro et al., 2014; Gallardo et al., 2016; Geuens et al., 2016). *HNRNPK* contains three repeat K homology domains, KH1, KH2, and KH3, which recognize target RNAs and play a central role in regulating gene expression, chromatin structure, and other genetic functions. To date, the genotypes and detailed clinical features of over 30

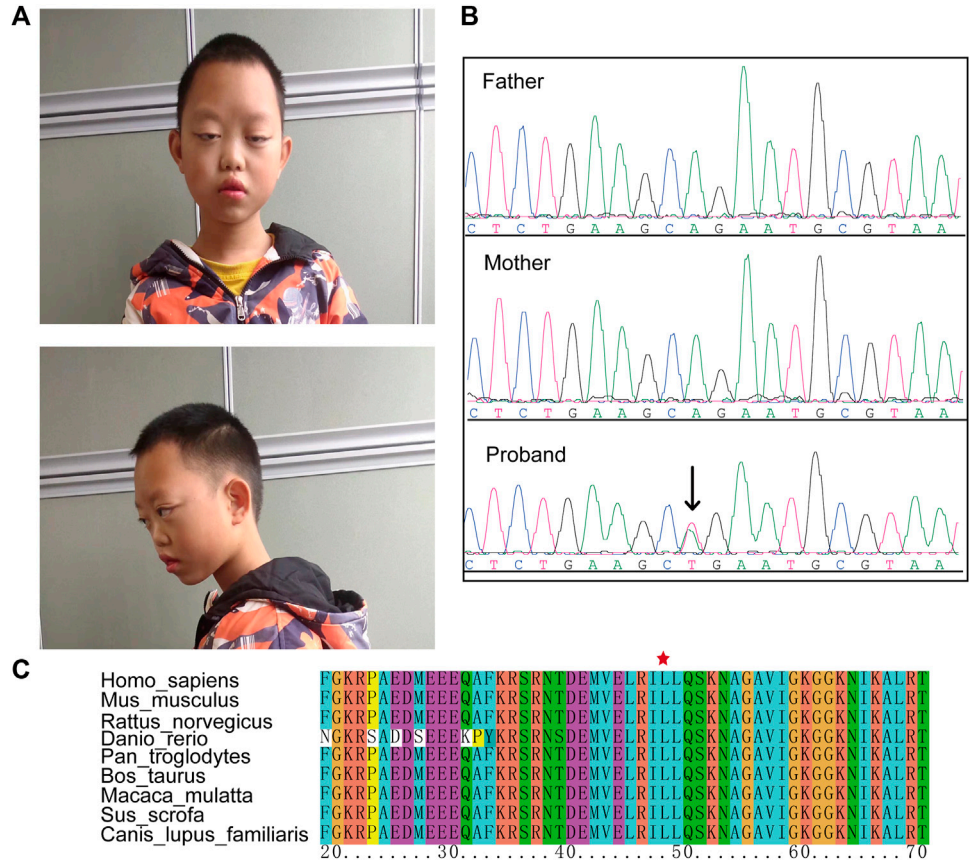


FIGURE 1 | The clinical phenotype of the case. **(A)** Dysmorphic features including a long face, long palpebral fissures, ptosis, and hypoplastic alar nasi. **(B)** Sanger sequencing confirmed a *de novo HNRNPK* variant (NM_002140.3: c.143T > A) in the proband. **(C)** This variant caused an amino acid change (p. L48G) that in a highly conserved region.

AKS patients have been well-characterized (Dentici et al., 2018; Gillentine et al., 2021). However, there are no reports of Chinese cases of AKS. In addition, more cases are needed to better understand the relationship between AKS and its associated pathogenic variants.

This study describes the clinical and molecular characteristics of the first Chinese AKS patient who had a novel *de novo* missense variant of *HNRNPK* (NM_002140.3: c.143T > A) and expands the current understanding of the genotypic spectrum of AKS.

Case Presentation

The proband, a boy 11 years and 10 months of age, was the first child of nonconsanguineous Chinese parents. Prenatal ultrasounds were normal. The proband failed to raise his head and exhibited hypotonia at 6 months of age and a language delay with his first words spoken at 3 years of age. He had a moderate degree of intellectual disability. Dysmorphic features included a long face, long palpebral fissures, ptosis, and hypoplastic alar nasi (**Figure 1A**).

An ultrasound confirmed mild hydronephrosis and cryptorchidism at 11 years of age (**Supplementary Figure S1**), and oligodontia was determined by panoramic radiographs. The proband's hearing was normal, and no significant abnormalities were found by brain magnetic resonance imaging (MRI). Chromosomal karyotype and microarray analysis results were normal.

METHODS

Ethical Compliance

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

DNA Extraction, Whole-Exome Sequencing, and Variant Analysis

Genomic DNA was isolated from each participant's peripheral blood using a blood genomic DNA extraction kit (Tiangen Biotech, Beijing, China) according to the manufacturer's protocol. The gDNA was fragmented, captured, and sequenced using the MGI-2000 sequencing system (BGI China).

Raw sequencing reads were filtered to obtain clean reads using Fastp,(Chen et al., 2018) and FastQC (Trivedi et al., 2014) was used to evaluate the quality of sequencing data in each sample.

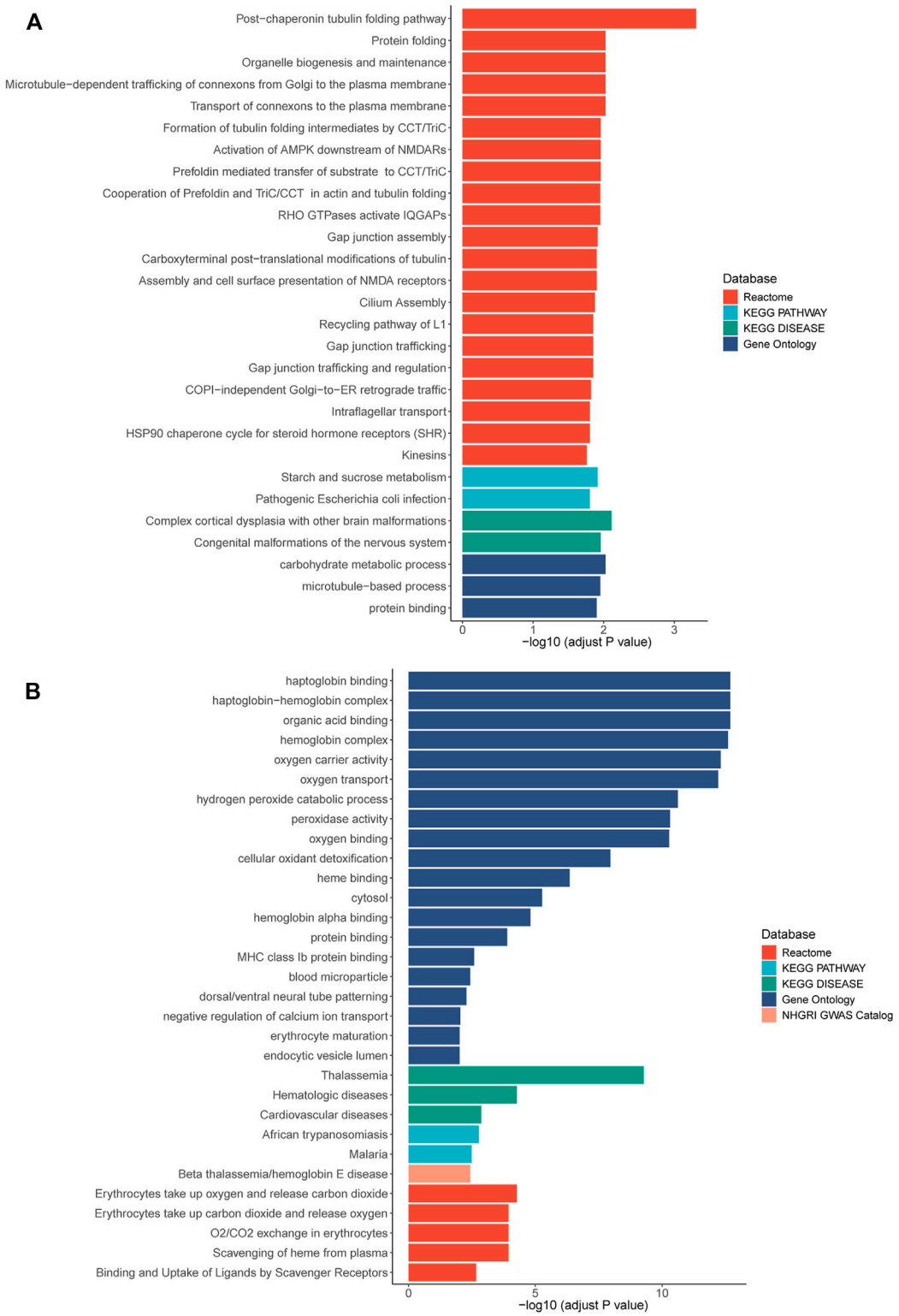


FIGURE 2 | Pathway enrichment results of genes with aberrant expression **(A)** and with aberrant splicing **(B)**. Colors stand for pathways items from different database. The *X* axis shows adjusted *p*-value with log transformed. Pathways with an adjusted *p*-value <0.05 was selected as significant pathways and plotted.

TABLE 1 | Pathogenic variants identified in *HNRNPK* gene (NM_002140.3).

Patient index	Genomic (hg38)	cDNAchange	AAchange	Function	References
1	chr9:g.86592674_86592675insC	c.85_86insG	p.Glu29Glyfs[a]	LOF	Wang et al. (2020)
2	chr9:g.86592661delA	c.99delT	p.Phe33Leufs[a]25	LOF	Gillentine et al. (2021)
3	chr9:g.83973901C > T	c.402+1G > A	N/A	LOF	Gillentine et al. (2021)
4	chr9:g.83972056dupC	c.779dupG	p.Asp262[a]	LOF	Au et al. (2018)
5	chr9:g.83971976G > A	c.859C > T	p.Arg287[a]	LOF	Au et al. (2018)
6	chr9:g.83971903_83971904insAA	c.931_932insTT	p.Pro311Leufs[a]40	LOF	Gillentine et al. (2021)
7	chr9:g.83971903_83971904insAA	c.931_932insTT	p.Pro311Leufs[a]40	LOF	Gillentine et al. (2021)
8	chr9:g.83971881dupC	c.953+1dupG	p.Gly319Alafs[a]6	LOF	Au et al. (2018)
9	chr9:g.83971682dupT	c.998dupA	p.Tyr333[a]	LOF	Dentici et al. (2018)
10	chr9:g.83971671C > T	c.1008+1G > A	N/A	LOF	Au et al. (2018)
11	chr9:g.83971356del	c.1009del	p.Val337Leufs[a]13	LOF	Au et al. (2018)
12	chr9: g.83970911delC	c.1094delG	p.Gly365Valfs[a]28	LOF	Au et al. (2018)
13	chr9:g.83970896C > A	c.1108+1G > T	N/A	LOF	Gillentine et al. (2021)
14	chr9:g.83970832A > G	c.1109-13T > C	N/A	LOF	Gillentine et al. (2021)
15	chr9:g.83970161C > T	c.1361+1G > A	N/A	LOF	Okamoto, (2019)
16	chr9:g.83969356A > ATTCT	c.1385_1386insAGAA	p.Phe462LfsThr[a]10	LOF	Gillentine et al. (2021)
17	chr9:g.83977061A > C	c.157-10T > G	p.52Lys_56AsninsLeuLeuGln	LOF	Yamada et al. (2020)
18	chr9:g.83975540T > C	c.214-35A > G	N/A	LOF	Murdock et al. (2021)
19	chr9:g.83975462C > T	c.257G > A	p.Arg86His? Splicing changes?	LOF	Au et al. (2018)
20	chr9:g.83975457C > T	c.257+5G > A	p.Ile87Tyrfs[a]12	LOF	Maystadt et al. (2020)
21	chr9:g.83974592G > A	c.258-3C > T	N/A	LOF	Gillentine et al. (2021)
22	chr9:g.83970334G > T	c.1192-3C > A	N/A	LOF	Gillentine et al. (2021)
23	chr9:g.83970229delC	c.1294delG	p.Asp432Ilefs[a]24	LOF	Jarvela et al. (2021)
24	chr9:g.83977780C > T	c.65G > A	p.Arg22His	missense	Farwell Hagman et al. (2017)
25	chr9:83971694G > A	c.986C > T	p.Pro329Leu	missense	Gillentine et al. (2021)
26	chr9:g.83971691C > T	c.989G > A	p.Gly330Glu	missense	Gillentine et al. (2021)
27	chr9:g.83977035A > G	c.173T > C	p.Ile58Thr	missense	Shashi et al. (2019)
28	chr9:g.83977032C > T	c.176G > A	p.Gly59Glu	missense	Wang et al. (2020)
29	chr9:g.83977009C > G	c.199G > C	p.Ala67Pro	missense	Gillentine et al. (2021)
30	chr9:g.83975466C > T	c.253G > A	p.Gly85Lys	missense	Gillentine et al. (2021)
31	chr9:g.83975466C > T	c.253G > A	p.Gly85Lys	missense	Gillentine et al. (2021)
32	chr9:g.83973359C > A	c.443G > T	p.Arg148Met	missense	Gillentine et al. (2021)
33	chr9:g.83973338A > G	c.464T > C	p.Leu155Pro	missense	Miyake et al. (2017)
34	chr9:g.83970744G > A	c.1184C > T	p.Pro395Leu	missense	Gillentine et al. (2021)
35	chr9:g.83977702A > T	c.143T > A	p. Leu48Gln	missense	This study

A total of 33 variants from 35 patients with AKS were curated from the published literature or online databases and this Study, study, including 21 loss-of-function variants and 12 missense variants. LOF, loss-of-function variants.
[a]*Nucleotide numbering and to indicate a translation termination (stop) codon.*

Clean DNA sequencing reads were mapped to the human reference genome hg19 (GRCh37) using the BWA-MEM algorithm (Li and Durbin, 2009), and ambiguously mapped reads (MAPQ <10) and duplicated reads were removed using SAMtools (Li et al., 2009) and PicardTools [http://broadinstitute.github.io/picard/], respectively. SNPs and small insertions and deletions (INDEL) were identified according to the Genome Analysis Toolkit software's best practices and variants were annotated using the Ensembl Variant Effect Predictor (VEP) (McLaren et al., 2016). According to guidelines from the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG) (Richards et al., 2015), variants were classified as pathogenic (P), likely pathogenic (LP), benign (B), likely benign (LB), or variants of uncertain significance (VUS). Variant validation was performed using Sanger sequencing (ABI 3730xl Genetic Analyzer).

RNA Sequencing and Data Preprocessing

Total RNA was isolated 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.

Raw sequencing reads were filtered to obtain clean reads using Fastp, and FastQC was used to evaluate the quality of sequencing data based on several measures, including sequence quality per base, sequence duplication level, and quality score distribution for each sample. The average quality score for overall RNA sequences was >30, indicating that a large percentage of the sequences were high quality. The clean RNA-sequencing reads were mapped to the human reference genome (hg19) using STAR (2.4.2a) with the Gencode v19 annotation (Dobin et al., 2013).

Identification of Aberrant Gene Expression and Pathway Enrichment Analysis

Aberrant gene expression, splicing, and monoallelic expression were detected using DROP (Yepez et al., 2021) with the default filter parameters. To increase the power to detect aberrantly

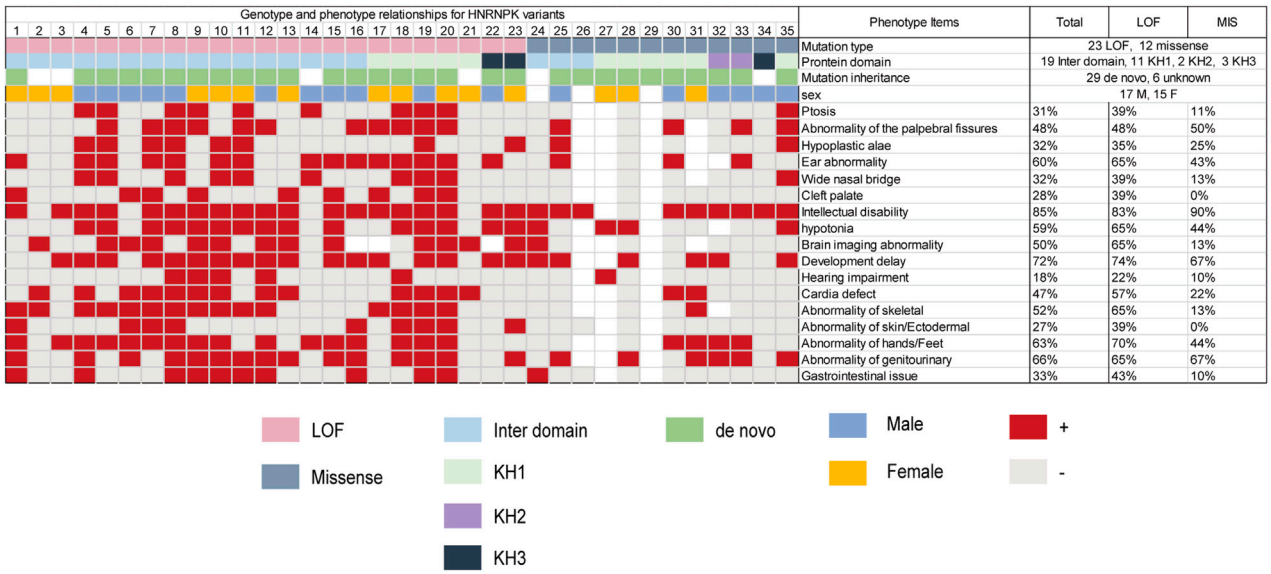


FIGURE 3 | Genotype and phenotype relationships for *HNRNPK* variants. Pink boxes represent loss-of-function variants; Navy blue boxes represent missense variants; Green boxes represent *de novo*; blue boxes represent male; yellow boxes represent female; red boxes represent present. Light gray boxes represent absent; white boxes represent unreported. Total (%) represents the number of patients who were reported to have the specific phenotype (numerator) divided by the number of patients. LOF (%) represents the number of patients with LOF variants who were reported to have the specific phenotype (numerator) divided by the number of patients with LOF variants. MIS (%) represents the number of patients with missense variants who were reported to have the specific phenotype (numerator) divided by the number of patients with missense variants.

expressed genes, in-house data with the same sequencing and analysis pipeline was included. Genes were defined as having aberrant expression, splicing, or monoallelic expression with a p-value <0.01. Functional enrichment of the aberrant 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-value <0.05 was selected as the threshold for significant pathways.

RESULTS

After trio whole-exome sequencing (Trio-WES) was performed on each family member, the causal variants were evaluated using ACMG guidelines. Results identified a novel missense *HNRNPK* variant (NM_002140.3: c.143T > A) in the patient that was absent in the parents (**Figure 1B**). This variant was classified as LP with the following evidence (PS2_Moderate + PM2 +PP3). This variant was not reported in the dbSNP and gnomAD databases, and predicted as pathogenic using in silico prediction tools (SIFT = 0.001, Polyphen2 = 0.99, MutationTaster = 1, and EVE = 0.945) (Frazer et al., 2021). The *HNRNPK* variant caused a missense substitution (p. Leu48Gln) localized in the K homology domains of a highly conserved region, suggesting that the variant may disrupt the binding ability of the HNRNPK protein (**Figure 1C**).

To prove this hypothesis, we predicted the effect of this missense variant on protein affinity with mCSM-NA (Pires and Ascher, 2017) using the PDB (Protein Data Bank) file provided by AlphaFold. The predicted affinity change (ΔΔG) was -0.033 kcal/mol, indicating that HNRNPK had reduced affinity for RNA binding. In addition, the predicted stability effect of HNRNPK was -3.541 kcal/mol, indicating that the protein was destabilized. These results suggest that the *de novo* variants (NM_002140.3:c.143T > A) may lead to loss of function of HNRNPK.

RNA sequencing was also performed and 42 and 86 genes with aberrant expression and splicing were identified, respectively (**Supplementary Figures S2–S5**, **Supplementary Tables S1, S2**). While the *HNRNPK* gene was not differentially expressed, two target genes regulated by *HNRNPK*, *TUBB2A* (p = 0.0093) and *TUBB2B* (p = 0.0092), were differentially expressed. Pathway enrichment analysis showed that the biological functions of these genes correlated with protein binding, transcriptional regulation, and nervous system regulation (**Figure 2**; **Supplementary Tables S3, S4**).

DISCUSSION AND CONCLUSION

In this study, we reported a patient with AKS who had multiple system anomalies, including developmental delay, facial dysmorphism, and kidney and genital abnormalities. Exome sequencing and Sanger validation showed that these phenotypes may be explained by a novel *de novo* missense variant of the *HNRNPK* gene (NM_002140.3: c.143T > A). In addition, there is a missense variant (NM_002140.3:c.142C > G; p. Leu48Val) interpreted as uncertain significance in ClinVar database which affect the same site with different amino acid.

With the addition of our patient, a total of 33 pathogenic *HNRNPK* variants (21 LOF and 12 missense) that caused AKS

have been reported in 35 patients (**Table 1**) (Miyake et al., 2017; Au et al., 2018; Gillentine et al., 2021). Three-quarters (75%) of the missense variants occurred within the KH domain (42% in the KH1 domain). In contrast, most LOF variants was in the interdomain. The most common phenotype in patients with AKS includes intellectual disability (85%), developmental delay (72%), genitourinary abnormalities (66%), and hand and foot abnormalities (63%). Phenotypic differences between individuals with loss-of-function and missense variants were observed in ptosis, wide nasal bridge, brain imaging abnormalities, cardiac defects, and skeletal and gastrointestinal abnormalities. The patient reported here had a unique phenotype of craniofacial features and oligodontia (**Figure 3**; **Supplementary Table S5**).

We hypothesized that missense variants affect RNA binding and thus cause dysfunction of related biological pathways that lead to the development of AKS. The *de novo* missense variant identified in this study was in the KH1 domain, and the predicted affinity change (ΔΔG) showed a reduced affinity for RNA binding. Transcriptome results identifying 42 aberrantly expressed and 86 spliced genes in the patient provided additional evidence to support the hypothesis. These genes were significantly associated with protein binding, transcriptional regulation, and nervous system function. Two of the downregulated genes, *TUBB2A* ($p = 0.0093$) and *TUBB2B* ($p = 0.0092$), are known to interact with HNRNPK at the protein level (Cerami et al., 2011). Both genes encode for the tubulin protein, which plays a critical role in neuronal function, neuronal migration, and postmigration development. Prior studies have shown that reduced expression of tubulin can lead to intellectual disability, matching the phenotype of the case in this study (Breuss et al., 2017; Jimenez et al., 2019; Schmidt et al., 2021). Further experiments are needed to assess the mechanism by which HNRNPK variants impact the development of AKS.

In summary, by integrating Trio-WES and RNA-seq analyses, we were able to better understand the role of an *HNRNPK* variant in a patient with AKS. Our findings expand the current understand of the clinical spectrum of *HNRNPK* variants.

DATA AVAILABILITY STATEMENT

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of the Second Affiliated Hospital of Chongqing Medical University in China. 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

XP performed the experiments and prepared the figures. SL performed RNA-seq analysis and wrote the manuscript. HY, XZ and LL provided intellectual and material support. 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 Chongqing Natural Science Foundation (cstc2019jcyj-msxmx0318); "Kuanren talents" project of the Second Affiliated Hospital of Chongqing Medical University (13-003-003) and 2020 Nan'an District Science and Health Joint Medical Scientific Research Project (2020-01).

ACKNOWLEDGMENTS

We thank the patient and the family members for their participation in this study. We thank professor Fengxiao Bu for his helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.853028/full#supplementary-material

REFERENCES

Au, P. Y. B., Goedhart, C., Goedhart, C., Ferguson, M., Breckpot, J., Devriendt, K., et al. (2018). Phenotypic Spectrum of Au-Kline Syndrome: a Report of Six New Cases and Review of the Literature. *Eur. J. Hum. Genet.* 26 (9), 1272–1281. doi:10.1038/s41431-018-0187-2

Barboro, P., Ferrari, N., and Balbi, C. (2014). Emerging Roles of Heterogeneous Nuclear Ribonucleoprotein K (hnRNP K) in Cancer Progression. *Cancer Lett.* 352 (2), 152–159. doi:10.1016/j.canlet.2014.06.019

Breuss, M. W., Nguyen, T., Srivatsan, A., Leca, I., Tian, G., Fritz, T., et al. (2017). Uner Tan Syndrome Caused by a homozygousTUBB2Bmutation Affecting Microtubule Stability. *Hum. Mol. Genet.* 26 (2), ddw383. doi:10.1093/hmg/ddw383

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 (W1), W317–W325. doi:10.1093/nar/gkab447

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 (Database issue), D685–D690. doi:10.1093/nar/gkq1039

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: an Ultra-fast All-In-One FASTQ Preprocessor. *Bioinformatics* 34 (17), i884–i890. doi:10.1093/bioinformatics/bty560

Dentici, M. L., Barresi, S., Niceta, M., Pantaleoni, F., Pizzi, S., Dallapiccola, B., et al. (2018). Clinical Spectrum of Kabuki-like Syndrome Caused by HNRNPK Haploinsufficiency. *Clin. Genet.* 93 (2), 401–407. doi:10.1111/cge.13029

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). STAR: Ultrafast Universal RNA-Seq Aligner. *Bioinformatics* 29 (1), 15–21. doi:10.1093/bioinformatics/bts635

Farwell Hagman, K. D., Shinde, D. N., Mroske, C., Smith, E., Radtke, K., Shahmirzadi, L., et al. (2017). Candidate-gene Criteria for Clinical Reporting: Diagnostic Exome Sequencing Identifies Altered Candidate Genes Among 8% of Patients with Undiagnosed Diseases. *Genet. Med.* 19 (2), 224–235. doi:10.1038/gim.2016.95

Frazer, J., Notin, P., Dias, M., Gomez, A., Min, J. K., Brock, K., et al. (2021). Disease Variant Prediction with Deep Generative Models of Evolutionary Data. *Nature* 599 (7883), 91–95. doi:10.1038/s41586-021-04043-8

Gallardo, M., Hornbaker, M. J., Zhang, X., Hu, P., Bueso-Ramos, C., and Post, S. M. (2016). Aberrant hnRNP K Expression: All Roads lead to Cancer. *Cell Cycle* 15 (12), 1552–1557. doi:10.1080/15384101.2016.1164372

Geuens, T., Bouhy, D., and Timmerman, V. (2016). The hnRNP Family: Insights into Their Role in Health and Disease. *Hum. Genet.* 135 (8), 851–867. doi:10.1007/s00439-016-1683-5

Gillentine, M. A., Wang, T., Wang, T., Hoekzema, K., Rosenfeld, J., Liu, P., et al. (2021). Rare Deleterious Mutations of HNRNP Genes Result in Shared Neurodevelopmental Disorders. *Genome Med.* 13 (1), 63. doi:10.1186/s13073-021-00870-6

Järvelä, I., Määttä, T., Acharya, A., Leppälä, J., Jhangiani, S. N., Arvio, M., et al. (2021). Exome Sequencing Reveals Predominantly De Novo Variants in Disorders with Intellectual Disability (ID) in the Founder Population of Finland. *Hum. Genet.* 140 (7), 1011–1029. doi:10.1007/s00439-021-02268-1

Jimenez, J., Herrera, D. A., Vargas, S. A., Montoya, J., and Castillo, M. (2019). β-Tubulinopathy Caused by a Mutation of the TUBB2B Gene: Magnetic Resonance Imaging Findings of the Brain. *Neuroradiol J.* 32 (2), 148–150. doi:10.1177/1971400919828142

Li, H., and Durbin, R. (2009). Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* 25 (14), 1754–1760. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* 25 (16), 2078–2079. doi:10.1093/bioinformatics/btp352

Maystadt, I., Deprez, M., Moortgat, S., Benoît, V., and Karadurmus, D. (2020). A Second Case of Okamoto Syndrome Caused by HNRNPK Mutation. *Am. J. Med. Genet.* 182 (6), 1537–1539. doi:10.1002/ajmg.a.61568

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., et al. (2016). The Ensembl Variant Effect Predictor. *Genome Biol.* 17 (1), 122. doi:10.1186/s13059-016-0974-4

Miyake, N., Inaba, M., Mizuno, S., Shiina, M., Imagawa, E., Miyatake, S., et al. (2017). A Case of Atypical Kabuki Syndrome Arising from a Novel Missense Variant in HNRNPK. *Clin. Genet.* 92 (5), 554–555. doi:10.1111/cge.13023

Murdock, D. R., Dai, H., Burrage, L. C., Rosenfeld, J. A., Ketkar, S., Müller, M. F., et al. (2021). Transcriptome-directed Analysis for Mendelian Disease Diagnosis Overcomes Limitations of Conventional Genomic Testing. *J. Clin. Invest.* 131 (1), e141500. doi:10.1172/JCI141500

Okamoto, N. (2019). Okamoto Syndrome Has Features Overlapping with Au-Kline Syndrome and Is Caused by HNRNPK Mutation. *Am. J. Med. Genet.* 179 (5), 822–826. doi:10.1002/ajmg.a.61079

Pires, D. E. V., and Ascher, D. B. (2017). mCSM-NA: Predicting the Effects of Mutations on Protein-Nucleic Acids Interactions. *Nucleic Acids Res.* 45 (W1), W241–W246. doi:10.1093/nar/gkx236

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 (5), 405–424. doi:10.1038/gim.2015.30

Schmidt, L., Wain, K. E., Hajek, C., Estrada-Veras, J. I., Guillen Sacoto, M. J., Wentzensen, I. M., et al. (2021). Expanding the Phenotype of TUBB2A-Related Tubulinopathy: Three Cases of a Novel, Heterozygous TUBB2A Pathogenic Variant p.Gly98Arg. *Mol. Syndromol* 12 (1), 1–8. doi:10.1159/000512160

Shashi, V., Schoch, K., Spillmann, R., Cope, H., Tan, Q. K., Walley, N., et al. (2019). A Comprehensive Iterative Approach Is Highly Effective in Diagnosing Individuals Who Are Exome Negative. *Genet. Med.* 21 (1), 161–172. doi:10.1038/s41436-018-0044-2

Trivedi, U. H., Cã©zard, T. e., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., et al. (2014). Quality Control of Next-Generation Sequencing Data without a Reference. *Front. Genet.* 5, 111. doi:10.3389/fgene.2014.00111

Wang, T., Hoekzema, K., Vecchio, D., Wu, H., Sulovari, A., Coe, B. P., et al. (2020). Large-scale Targeted Sequencing Identifies Risk Genes for Neurodevelopmental Disorders. *Nat. Commun.* 11 (1), 4932. doi:10.1038/s41467-020-18723-y

Yamada, M., Shiraishi, Y., Uehara, T., Suzuki, H., Takenouchi, T., Abe-Hatano, C., et al. (2020). Diagnostic Utility of Integrated Analysis of Exome and Transcriptome: Successful Diagnosis of Au-Kline Syndrome in a Patient with Submucous Cleft Palate, Scaphocephaly, and Intellectual Disabilities. *Mol. Genet. Genomic Med.* 8 (9), e1364. doi:10.1002/mgg3.1364

Yépez, V. A., Mertes, C., Müller, M. F., Klaproth-Andrade, D., Wachutka, L., Frésard, L., et al. (2021). Detection of Aberrant Gene Expression Events in RNA Sequencing Data. *Nat. Protoc.* 16 (2), 1276–1296. doi:10.1038/s41596-020-00462-5

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.

frontiers in Genetics

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fgene.2022.852764

Identification of a Novel Deep Intronic Variant by Whole Genome Sequencing Combined With RNA Sequencing in a Chinese Patient With Menkes Disease

Xiufang Zhi[1,2†], *Qi Ai*[3,4†], *Wenchao Sheng*[1,2†], *Yuping Yu*[1,2†], *Jianbo Shu*[2,5,6], *Changshun Yu*[7], *Xiaoli Yu*[2,8*], *Dong Li*[2,8*] *and Chunquan Cai*[2,5,6*]

[1]Graduate College of Tianjin Medical University, Tianjin, China, [2]Tianjin Children's Hospital (Children's Hospital of Tianjin University), Tianjin, China, [3]Key Laboratory of Cancer Prevention and Therapy, Department of Pediatric Oncology, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital, Tianjin Medical University, Tianjin, China, [4]Department of Hematology and Oncology, Tianjin Children's Hospital, Tianjin, China, [5]Tianjin Pediatric Research Institute, Tianjin, China, [6]Tianjin Key Laboratory of Birth Defects for Prevention and Treatment, Tianjin, China, [7]Tianjin Kingmed Center for Clinical Laboratory, Tianjin, China, [8]Department of Neurology, Tianjin Children's Hospital, Tianjin, China

OPEN ACCESS

Edited by:
Xiaoli Chen, Capital Institute of Pediatrics, China

Reviewed by:
Yu An,
Fudan University, China
Ana Topf,
Newcastle University, United Kingdom
Jing Peng,
Central South University, China

***Correspondence:**
Xiaoli Yu
yxlniu@163.com
Dong Li
lidongtjetyy@163.com
Chunquan Cai
cqcns6@126.com

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

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: 11 January 2022
Accepted: 09 March 2022
Published: 31 March 2022

Citation:
Zhi X, Ai Q, Sheng W, Yu Y, Shu J, Yu C, Yu X, Li D and Cai C (2022) Identification of a Novel Deep Intronic Variant by Whole Genome Sequencing Combined With RNA Sequencing in a Chinese Patient With Menkes Disease. Front. Genet. 13:852764. doi: 10.3389/fgene.2022.852764

Background: Menkes disease (MD) is a rare X-linked connective tissue disorder of copper metabolism caused by pathogenic variant(s) in *ATP7A* gene. The aim of the present study is to determine the clinical characteristics and molecular basis of one patient with MD.

Methods: One 10-month-old Chinese boy who met the clinical manifestations of MD was enrolled in this study. Whole genome sequencing (WGS) was performed in the patient in order to identify the variant(s), followed by Sanger sequencing. RNA sequencing (RNA-seq) from whole blood was subsequently applied to assess the effect of variant on transcription levels, and reverse transcriptase-polymerase chain reaction (RT-PCR) was performed for further validation. In addition, X chromosome inactivation (XCI) status of the patient's mother at the DNA level was measured by capillary electrophoresis.

Results: The patient suffered from intermittent convulsions for more than 6 months, with psychomoto retardation and neurodegenerations. The patient also had curly hair, hypopigmented skin, cutis laxa, decreased muscle strength and hypotonia. MRI showed the intracranial arteries were tortuous with some "spiral" changes. The patient's serum ceruloplasmin level was low. WGS revealed one novel hemizygous variant, c.2627-501C > T (NM_000,052.7), located in the deep intronic sequence of *ATP7A* gene. Sanger sequencing confirmed that the variant was inherited from his mother. RNA-seq confirmed the variant itself, and identified a pseudo-exon inserted between exons 12 and 13 in mRNA of *ATP7A*. The sequencing results of RT-PCR from the patient confirmed this finding, while neither of his parents detected aberrant splicing. The Capillary electrophoresis results showed that the patient's mother had a skewed XCI.

Conclusion: Our finding of the variant enlarges the variant spectrum in the *ATP7A* gene. This is a novel deep intronic variant which leads to the activation of a pseudo-exons in the *ATP7A* gene, and it demonstrates the usefulness of WGS combined with RNA-seq, in terms of revealing disease-causing variants in non-coding regions. Furthermore, the fact

that the deep intronic variants cause disease by the activation of pseudo-exon inclusion indicates that in MD this might be an important mechanism.

Keywords: ATP7A gene, Menkes disease, whole genome sequencing, RNA sequencing, deep intronic variants

INTRODUCTION

MD, also known as Menkes kinky hair syndrome, is a rare X-linked recessive genetic disease of impaired copper metabolism caused by variants in *ATP7A* gene (Caicedo-Herrera et al., 2018; Patel et al., 2017). Clinical manifestations include neurological symptoms and systemic multi-system dysfunction (Altarelli et al., 2019). According to clinical manifestations and progression, it can be divided into three subtypes: 1) classical type accounts for 90–95%, manifested as progressive neurodegeneration, seizures, connective tissue abnormalities and "kinky" with hair, and ultimately death in early childhood; 2) atypical MD has various clinical phenotypes of the classical type, but the degree is milder and the survival time is longer; and 3) occipital horn syndrome only affects connective tissue and bones, with the mildest form (Bonati et al., 2017; Kaler and Distasio, 1993; Kodama et al., 2012; Smpokou et al., 2015).

ATP7A gene, located on the Xq13.3 chromosomal region, contains 23 exons and encodes the transmembrane P-type ATP enzyme which can transport copper to the extracellular or Golgi apparatus by phosphorylation/dephosphorylation (Hartwig et al., 2019; Horn and Wittung-Stafshede, 2021; Skjorringe et al., 2017). Variants in *ATP7A* gene cause the inactivation of copper transport ATP enzyme, which block the absorption of copper in intestinal mucosa and the transport of copper across blood-brain barrier, ultimately leading to nervous system abnormalities and multi-system dysfunction (Chen et al., 2020; Maung et al., 2021; Petris et al., 2000; Prithvi et al., 2019). MD is mostly caused by nonsense variants, frame shift variants and splice site variants (Cao et al., 2017; de Gemmis et al., 2017; Kaler et al., 2020; Tumer, 2013). About 22% of *ATP7A* variants are splice site variants (Moller et al., 2009). Splice site variants may cause one or more exons complete skipping or retention of introns by activation of cryptic splice sites within exons or introns (Anna and Monika, 2018; Parada et al., 2014), and eventually affect the structure and function of transmembrane P-type ATP enzyme or reduce the expression of normal products (Moller et al., 2000). In the clinically discovered *ATP7A* variant-related MD patients, the vast majority of splice site variants occur in the intrinsic splicing zone (Skjorringe et al., 2011). The pathogenesis of deep intronic variants is rarely reported.

One possible important reason why deep intronic variants are always overlooked is that the detection methods are inappropriate. As an important means of identify disease-causing genetic variants of individuals affected with a genetically undiagnosed rare disorder, the main disadvantage of whole exome sequencing (WES) is that it cannot indicate disease-causing variants in non-coding regions. WGS in principle reveals all genetic variants, however, the main challenges are the sizeable variants number and poor interpretation of the non-coding variants. RNA-seq can directly detect gene expression, which help to reveal splice defects, the mono-allelic expression of heterozygous loss-of-function variants, and expression outliers, thereby providing functional data to support the clinical interpretation of variants (Cummings et al., 2017; Kremer et al., 2017). Therefore, the combination of WGS and RNA-seq displays strong superiority in the detection and analysis of non-coding variants.

In the present study, we identified a novel hemizygous variant of *ATP7A* located in deep intronic sequences by WGS combined with RNA-seq in a Chinese family. In addition, we described the clinical profile, laboratory and radiological findings of the patient and performed a variant analysis of *ATP7A* in the patient with MD.

Patients, Materials and Methods

Ethics Statement

The study was approved by the Ethics Committee of Tianjin Children's Hospital (Tianjin University Children's Hospital), and the study approval reference number was 2016021. Written informed consent of the patient was obtained from his parents, and the parents gave their own written informed consent to take part in this study. All study procedures adhered to the tenets of the Declaration of Helsinki.

Patients

The patient, a 10-month-old Chinese boy with non-consanguineous parents, was admitted to Tianjin Children's Hospital because of convulsions. All of the patient's clinical information was collected from medical records. The peripheral blood samples were collected from the patient and his parents. All the experiments performed with human materials were in accordance with the relevant guidelines and regulations.

DNA and RNA Isolation

Genomic DNA was isolated from peripheral venous blood of the patient and both parents using Blood Genomic DNA Mini Kit (CWBIO, Jiangsu, China) according to the manufacturer's protocol. Total RNA was extracted from peripheral venous blood using Lymphocyte Separation Medium (SOLARBIO SCIENCE &TECHNOLOGY CO., LTD., Beijing, China) and Trizol reagent (TIANGEN BIOTECH CO., LTD., Beijing, China). DNA and RNA was quantified by Nanodrop Spectrophotometer (THERMO FISHER SCIENTIC, Shanghai, China). DNA and blood samples were stored at −20°C. RNA was stored at −80°C.

WGS and in Silico Analysis of Potential Splice Variant

WGS was performed by KingMed company (Guangzhou, China) on the patient. The DNA was performed with 150-bp paired-end sequencing and average coverage of 50-fold in more than 95% of

the target regions, including the coding region and intronic region. WGS data were processed following the standard procedure. Low-quality reads, adapters, and sequences with more than 5% unknown bases were removed from the raw reads. Sequenced reads were aligned to the human reference genome hg19 using the Burrows-Wheeler Aligner software. Duplications were marked with Picard and BAM was sorted with Samtools2. GATK Best Practices Pipeline was used in the process of SNP and INDEL calling. SNPs and INDELs were annotated using Annovar and local software. The HGMD, 1,000 genomes, OMIM, dbSNP, Clinvar and local database were used for variants annotation. Specifically, all the variants were firstly filtered using the dbSNP, 1,000 genomes, and local databases, deleting all variants with frequency >1% in the population. All the remaining filtered variants were searched in the OMIM and HGMD databases for identifying the variants of genes associated with the disease phenotype. Then, the variant pathogenicity classification followed the variant classification criteria formulated by American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) (Richards et al., 2015; Zhang et al., 2020). Clinical significance grading was done as described previously (Cai et al., 2021). Sanger sequencing was performed for the candidate variant(s) in the patient and his parents for confirmation.

To analyze the effect of the candidate variant(s) on pre-mRNA splicing, we performed *in silico* analysis using several different freely available software including NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/), Splice Site Prediction by Neural Network (www.fruitfly.org/seq_tools/splice.html), and ASSP model (wangcomputing.com/assp). The variant(s) designations were based on *ATP7A* gene (NM_000,052.7) from NCBI, and the gene sequence is from NC_000,023.11. We analyzed the new splice site and its effect on splicing by comparing the difference of the splice site between the wild type and the variant(s).

RNA-Seq

RNA-seq assay was applied to confirm the splicing effect of the candidate variant(s). Whole-blood samples were collected in PAXgene whole blood RNA tubes (BD), and intracellular RNA was extracted using PAXgene blood RNA kit (QIAGEN) according to the manufacturer's recommendation. RNA was quantified and then processed to sequencing library preparation using Ribo-off rRNA depletion kit (human/mouse/rat) and VAHTS® universal V8 RNA-seq library prep kit for Illumina (Vazyme). The qualified library was sequenced on an Illumina Novaseq 6,000 plateform with PE150 mode at the Tianjin KingMed Center for Clinical Laboratory for Translational Genomics, and approximately 30–50 million raw reads were expected to be generated.

Raw reads were processed by trimming poly-A/T stretches and Illumina adapters using Cutadapt software and resulting reads shorter than 30bp were discarded. Sequencing data was mapped to the GRCh37 reference genome using STAR and HISAT2 software with default parameters. Transcripts were assembled by Stringtie software from the reads alignment data for abnormal transcripts detection. The iGV v2.9.4 software was applied for manual analysis for the genomic regions of interest.

RT-PCR

Complementary DNAs (cDNA) were synthesized with the FastKing RT Kit (With gDNase) (TIANGEN BIOTECH CO.,LTD., Beijing, China) according to the manufacturer's instruction. To determine the consequence of the variant for the *ATP7A* gene transcript, we performed PCR experiments on cDNA using primers located in exon 9 (Forward, 5′-CTTCTA TGTTCCTGGAGCG-3′) and in exon 15 (Reverse, 5′-ATCGTT TCTGTTCGGGAG-3′). Due to the low expression of *ATP7A* in whole blood, we performed two sequential amplification reactions, and both of which used the same primers. Alternatively, the product of the first amplification reaction was used as the template for the second PCR, which was amplified by the same primers as the first primer pair. The final PCR products from cDNA was resolved by 1.5% agarose gel electrophoresis and sequenced (GENEWIZ, Hangzhou, China).

XCI Status Measurement by the Human Androgen Receptor (HUMARA) Assay

The XCI status at the DNA level was measured using HUMARA assay based on the highly polymorphic CAG trinucleotide repeat in the promoter area of the androgen receptor (AR). Specifically, genomic DNA was digested with methylation-sensitive restriction endonuclease HhaI (TAKARA BIO, Japan). The CAG repeat regions from digested and undigested DNA samples were amplified by PCR with forward primer (GCT GTGAAGGTTGCTGTTCCTCAT) and reverse primer (TCCAGAATCTGTTC CAGAGCGTGC) (Pan et al., 2021), then the PCR products were resolved by 1.5% agarose gel electrophoresis to confirm that DNA had been fully digested. Additionally, the same DNA samples were amplified by PCR using same primers but the forward primer was end-labeled with FAM fluorescence. The products were analysed by capillary electrophoresis (BECKMAN COULTER, United States). The proportion of variant allele on the inactive X chromosome was calculated according to the standard formula (Thouin et al., 2003). XCI status was classified as random (XCI<70%), or skewed (XCI≥70%) (Pu et al., 2010; Sato et al., 2004). If the ratio was ≥90%, the XCI pattern was considered to be extremely skewed (Yoon et al., 2008).

RESULTS

Clinical Characteristics of the Patient

The male Chinese patient (proband) was born at term after an uncomplicated pregnancy, with third degree contamination of amniotic fluid, and neonatal jaundice after birth. His birth weight was 2.25 kg. Convulsions were seen as early as 4 months after birth, the patient presented with right-sided eyelid myoclonus for several seconds over 10 times a day. The second seizure type appeared at 6 months of age. The patient showed flexor spasms in either a isolation seizure or a cluster of seizures. When a cluster of seizures occurred, seizure occurred 2–8 times per day and the maximum duration was about 20 s. At 8 months of age, the third

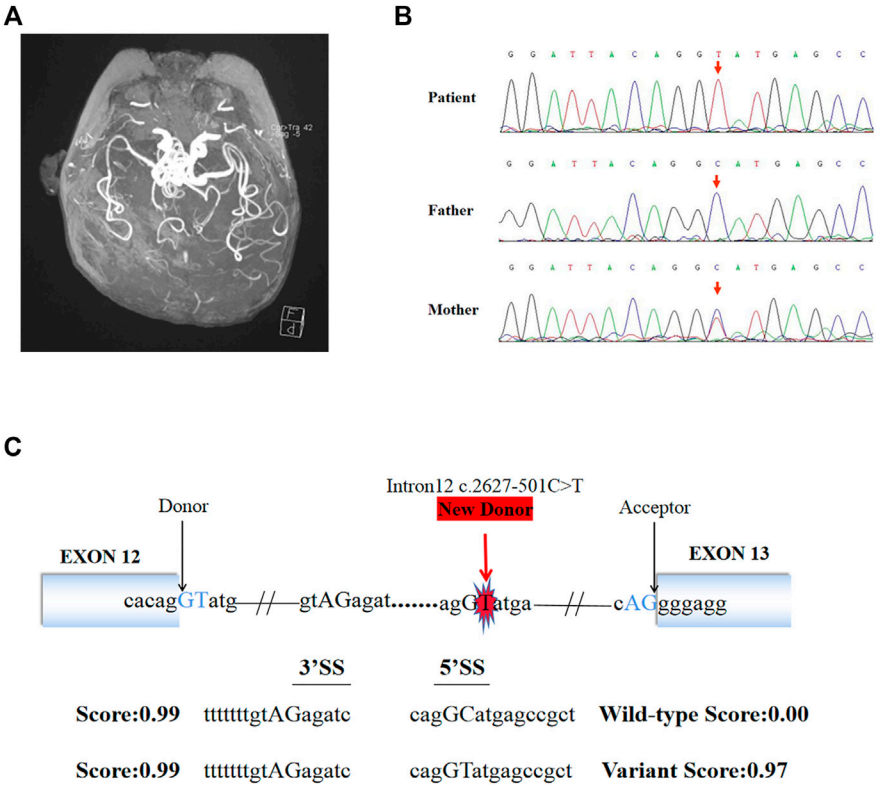


FIGURE 1 | Brain MRI, Sanger sequencing results and splice site prediction schematic of patient. **(A)** Magnetic resonance angiogram showed the intracranial arteries were tortuous with some "spiral" changes. **(B)** Sanger sequencing confirmed that the variant (c.2627-501C > T) of *ATP7A* gene was inherited from the patient's mother. **(C)** The prediction results *in silico* showed the existence of a new 5′ donor splice site and a cryptic 3′ receptor splice sites. The boxes represented exons, the black arrows pointed to the normal classic splice site, and the red arrow pointed to the new donor splice site caused by the variant.

type, generalized tonic-clonic seizures, occurred. The patient received 7–8 times of seizure per day, each lasting for 30 s to 1 min. There were no loss of consciousness in the first two types of seizure. Only spasms were described in the last 1 month with an exacerbation of symptoms. The patient received antiepileptic therapy with topiramate and levetiracetam, but the effect was poor. At present, the patient weighed 7.1 kg. Developmental milestones of the patient was normal before 4 months of age. However, developmental regression was beginning to appear at 4 months of age. The age of the patient was10.6 months old at this visit. He could not raise his head, turn over, or sit alone. His deciduous teeth had not erupted yet. He could not make a laugh and sought to track parallax. He could only eat with a dropper at presentation, showing global developmental delay.

On physical examination, his head was flat with irregular shape. The hair was sparse, short, and curly while the skin was fair, rough, and loose. One 3 mm × 3 mm café au lait spot could be seen on the inside of the right upper arm. He had pectus excavatum, decreased muscle strength, hypotonia and decreased muscle volume.

Video EEG showed hypsarrhythmia. Brain MRI showed patchy flaky slightly long T2 signal in the white matter of bilateral frontal, temporal, and parietal cortices. DWI showed a small patch hyperintensity in the bilateral centrum semiovale. Moreover, MRI showed widened ventricles, sulci and cisterns. Brain MRA showed the intracranial arteries were tortuous with some "spiral" changes (**Figure 1A**). The serum ceruloplasmin was 0.05 g/L (normal range 0.15–0.30 g/L). MD was suspected clinically. Given the severe neurological symptoms, it was considered to be the classical type. An extensive previous genetic workup, including karyotype, WES, trio WES, were negative.

Identification and Analysis of the Candidate Variant

140G sequencing data was acquired after WGS for the proband and 40.5 million SNVs, 931 thousands indels, and no pathogenic CNVs were detected in the proband's WGS data. After variant pathogenicity classification and clinical significance grading screening process, only one candidate variant, NC_000,023.10: g.77275240 C > T (hg37, chrX), c.2627-501C > T (NM_000,052.7), located in deep intron 12 of *ATP7A* gene was identified in the sequencing data as a potential causative variant for the proband's phenotype. Sanger sequencing confirmed that the c.2627-501C > T variant was inherited from the mother carrying heterozygous variant (**Figure 1B**). This variant has never been reported in ClinVar database, HGMD database, or gnomAD database before. We submitted the sequence of the interest region to splice site predictor programs (NetGene2, Splice Site

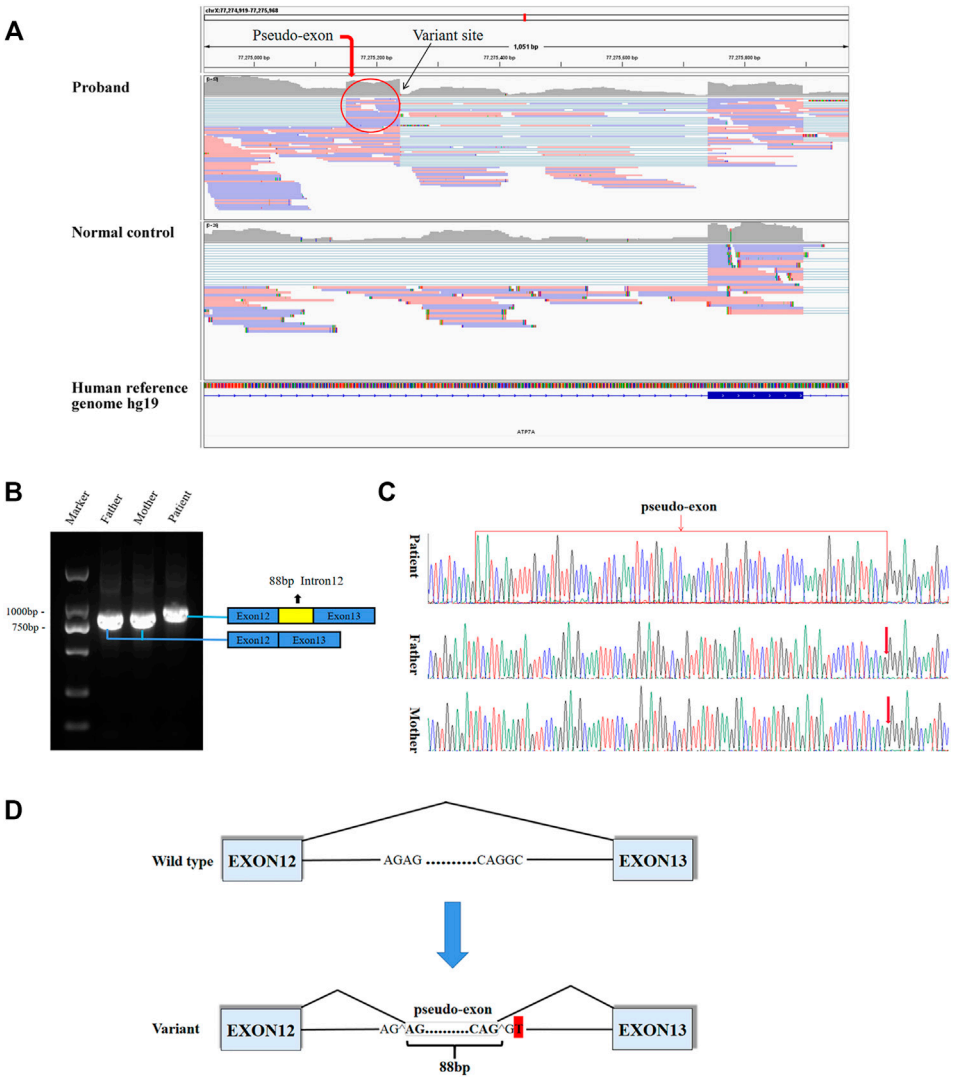


FIGURE 2 | Aberrant splicing caused by the varant. **(A)** RNA-seq demonstrated an abnormal splicing pattern that resulted in an out-of-frame pseudo-exon between exons 12 and 13. The black arrow pointed to the variant site, and the red circle was the fragment from intron 12 inserted between exons 12 and 13. **(B)** In agarose gel electrophoresis of RT-PCR, a lager product size in patient compared with the normal size of his parents. The blue boxes represented exons, and the red box represented intron fragment inserted between exons 12 and 13. **(C)** The sequencing results of RT-PCR showed there was a 88bp pseudo-exon fragment between exons 12 and 13 in mature ATP7A mRNA of the patient compared to his normal parents. **(D)** Schematic diagram of WT and variant splicing patterns showed that the variant site created a new 5′ donor splice site and activated the upstream 3′ acceptor splice site concomitantly, thereby creating the aberrant splicing. Boxes represented exons. Straight lines represented base sequences. Broken lines represented splicing modes.

Prediction by Neural Network, ASSP model) and all of them confirmed that the variant activated a new splice donor site (**Figure 1C**).

Aberrant Splicing Caused by the Varant in Transcription Level

8.8G sequencing data was acquired after RNA-seq for the proband. The *ATP7A* gene region was carefully looked through manually using the iGV software and we successfully detected the c.2627-501C > T variant. The data also demonstrated clearly an abnormal splicing pattern that resulted in an out-of-frame pseudo-exon between exons 12 and 13 (**Figure 2A**) which predicted to cause frame shift and protein truncation (p. Gly876Cysfs*26). By the way, the sequencing depth for the c.2627-501C site was only 5x and variants with such low depth were rendered as too much low quality data (although the C > T mutation frequency was 5/5) and were filtered after variant calling with default parameters. For further verification of the impact of the novel variant on splice site, we performed RT-PCR. In agarose gel electrophoresis of RT-PCR, cDNA analysis revealed a lager product size in patient compared with the normal size of his parents (**Figure 2B**). Direct sequencing of the patient's amplified fragment revealed aberrant RNA splicing effect

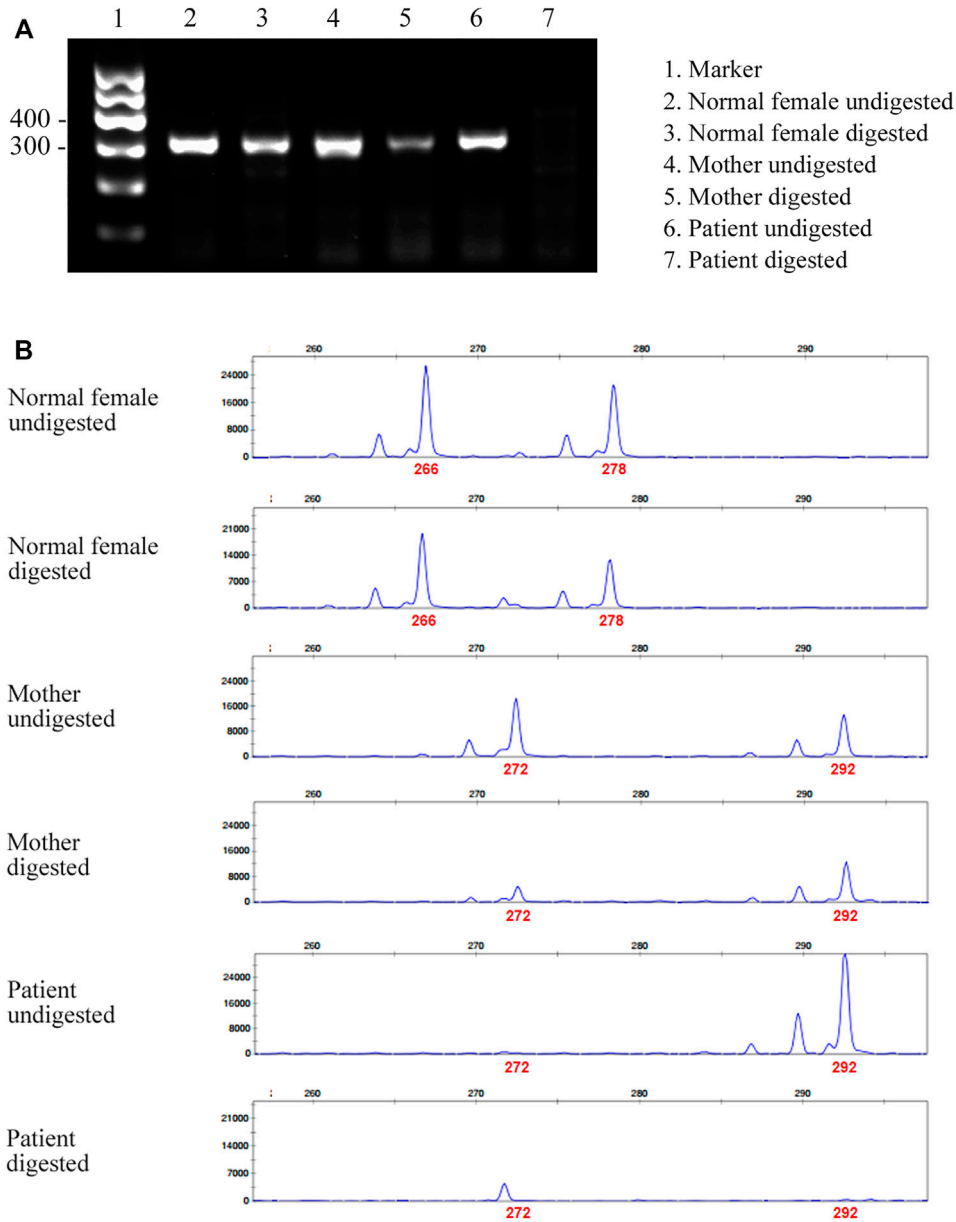


FIGURE 3 | Agarose gel electrophoresis and capillary electrophoresis of XCI. **(A)** The result of agarose gel electrophoresis showed that there was no amplification band in the patient's and his father's DNA after digestion, indicating that the DNA was completely digested. **(B)** The result of capillary electrophoresis showed that there was one patient's product with a size of 292 bp before DNA digestion, suggesting that this was the X chromosome carrying the variant. There were two mother's products, 272 bp (WT) and 292 bp (variant), both in digested DNA and undigested DNA. The result showed skewed pattern of XCI.

between exons 12 and 13, which resulted in the appearance of a pseudo-exon in the splicing process (**Figure 2C**), which was consistent with the RNA-seq result in terms of variant analysis. Of note, only wild-type (WT) *ATP7A* mRNA expression was detected in both electrophoresis and direct sequencing in the patient's mother.

Pathogenic Classification of the Variant

According to the ACMG/AMP variant classification criteria, the c.2627-501C > T variant could be classified as 'likely pathogenic' (PP4-Moderate, PS3-Moderate, PM2-Supporting, PP3-Supporting) (Richards et al., 2015; Zhang et al., 2020).

Evaluation of XCI

Given that only WT *ATP7A* mRNA expression was detected in the patient's mother, we performed the HUMARA assay for XCI on samples from the patient and his parents to evaluate the XCI state and to better understand the mechanism. The results of agarose gel electrophoresis showed that the DNA was completely digested (**Figure 3A**). The results of capillary electrophoresis

showed that there was one patient's product with a size of 292 bp before DNA digestion, suggesting that this was the X chromosome carrying the variant (**Figure 3B**). There were two mother's products, 272 bp (WT) and 292 bp (variant), both in completely digested DNA and undigested DNA (**Figure 3B**). The proportion of variant allele on the inactive X chromosome was calculated as 78.0% indicating a skewed XCI model compare with 45.3% in normal female control.

DISCUSSION

In this study, the patient presented with epilepsy, rough and loose skin, sparse and curly hair, and psychomoto retardation. The level of serum ceruloplasmin was low. Otherwise, the patient's MRI showed that the intracranial arteries were tortuous with some "spiral" changes. All the above manifestations were highly matched with MD. The complete diagnosis of MD still required genetic variant evidence. As a result, the patient and his parents underwent WES, trio WES in other medical institution, but no abnormality consistent with the patient's clinical manifestations was found. Considering the limitations of WES including the inability to detect non-coding variants in intronic regions (Burdick et al., 2020; Jelin and Vora, 2018) and balanced translocation of chromosomes, we performed a karyotype analysis for the patient and the chromosomal analysis revealed a normal karyotype with 46, XY.

Due to these negative results and strong demands from the patient's parents to find the etiology of the diease, we performed WGS for its capability to detect a full range of common and rare genetic variants of different types across almost the entire genome (Lappalainen et al., 2019). WGS found a deep intronic variant in the patient's *ATP7A* gene (c.2627-501C > T). Then, the Sanger sequencing found that the variant was inherited from his mother.

It has been reported that deep intronic variant results in intron fragments (also known as cryptic exons or pseudo-exons) being inserted into mature transcripts, and leads to frame shift or premature termination of translation (Anna and Monika, 2018). Functionally, these variants create new acceptor or donor sites, which are recognized by the splicing complex and interact with existing intron cryptic splice sites. It is also possible that deep intronic variants lead to the generation of new regulatory elements (such as splicing enhancers) and the recognition of specific introns. To explore the effect of the variant on transcription level, we performed RNA-seq. RNA-seq can directly detect gene expression and provide functional data that supports clinical interpretation of variants. Although RNA-seq is a supplementary tool, its advantage lies in detecting the functional consequences of DNA variants that affect splicing and gene expression levels. Therefore, it supports the interpretation of variants of uncertain significance, and reorders variants that are ignored by DNA sequencing. The combination of RNA-seq with WES or WGS analysis may provide more information than WES or WGS analysis alone (Fresard et al., 2019; Gonorazky et al., 2019; Lee et al., 2020). In our case, based on the guidance of WGS to find the causative deep intronic variants, RNA-seq analysis found abnormal gene expression and splicing. RNA-seq result of the patient showed that a pseudo-exon (88bp) from intron 12 located between exons 12 and 13 was contained in the mature mRNA, and confirmed by RT-PCR. Due to the variant, a new donor splice site was generated, and the upstream cryptic splice acceptor site was activated, leading to 88 bp insertion from the new donor splice site upstream in intron 12 (**Figure 2D**). Finally, a pseudo-exon fragment from intron 12 was included in the mature mRNA, which resulted in a frame shift and a premature termination codon (p. Gly876Cysfs*26). A deep intron c.3718-2477C > T variant located in intron 19 of *CFTR* gene has been reported in the literature. This variant leads to the generation of a new donor site, resulting in an 84 bp pseudo-exon containing into the mature mRNA. This pseudo-exon contains an in-frame termination codon, so the translated protein is shorter and non-functional (Anczukow et al., 2012; Anna and Monika, 2018). However, there is a limitation in our study. At present, we cannot explain the mechanisms of how the newly formed donor splice site and cryptic acceptor splice site caused by the deep intronic variant lead to aberrant splicing. We will further verify these issues in follow-up research.

In this study, it was learned through WGS and Sanger sequencing that the patient's variant was inherited from his mother, but the subsequent RT-PCR results showed that only WT *ATP7A* mRNA expression was detected in the patient's mother, which was same with that of the father's. We performed the HUMARA assay for XCI and the result suggested that the patient's mother had a skewed XCI with a preferential inactivation of the X chromosome that carried the variant allele. This result may explain why the patient's mother could only detect WT *ATP7A* mRNA expression.

Due to the high mortality rate of MD, patients with classic MD generally die before the age of three. Early diagnosis of MD and identification of pathogenic variants are the basis of genetic counseling. The results of this study are of great significance for this family's genetic counseling, preventing the birth of children with MD (Elliott, 2020; Perge and Igaz, 2019; Yang and Kim, 2018). Early diagnosis may also be meaningful for MD gene therapy. It has been reported that Antisense oligonucleotides can improve the abnormal splicing effect of the aforementioned *BRCA2* gene intron 12, c.6937 + 594T > G to a certain extent (Anczukow et al., 2012). Gene therapy is still in the exploratory stage, however, it is expected that MD caused by such splice site variants can be treated by gene therapy in the future.

CONCLUSION

Our finding of the variant enlarges the variant spectrum in the *ATP7A* gene. This is a novel deep intronic variant which leads to the activation of a pseudo-exons in the *ATP7A* gene, and it demonstrates the usefulness of WGS combined with RNA-seq, in terms of revealing disease-causing variants in non-coding regions. Furthermore, the fact that the deep intronic variants

cause disease by the activation of pseudo-exon inclusion indicates that this might be an important mechanism in MD.

DATA AVAILABILITY STATEMENT

All data relevant to the study are included in the article. The whole-genome sequencing data are not publicly available as these could compromise research participant privacy. Specific variant requests or other data are available from the corresponding author CC upon reasonable request.

ETHICS STATEMENT

The study was approved by the Ethics Committee of Tianjin Children's Hospital (Tianjin University Children's Hospital), and the study approval reference number was 2016021. Written informed consent of the patient was obtained from his parents, and the parents gave their own written informed consent to take part in this study. All study procedures adhered to the tenets of the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

DL and CC conceived the project. XZ drafted the manuscript. JS designed the experiments. WS, YY, QA, and CY performed the experiments and analyzed the data. XY provided patient care and relevant clinical information. All authors read, critically reviewed, and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China [grant number 81771589] and the Public Health and Technology project of Tianjin [grant number TJWJ2021ZD007].

ACKNOWLEDGMENTS

We thank patient and his parents for their participation in this study.

REFERENCES

Altarelli, M., Ben-Hamouda, N., Schneider, A., and Berger, M. M. (2019). Copper Deficiency: Causes, Manifestations, and Treatment. *Nutr. Clin. Pract.* 34 (4), 504–513. doi:10.1002/ncp.10328

Anczuków, O., Buisson, M., Léoné, M., Coutanson, C., Lasset, C., Calender, A., et al. (2012). BRCA2 Deep Intronic Mutation Causing Activation of a Cryptic Exon: Opening toward a New Preventive Therapeutic Strategy. *Clin. Cancer Res.* 18 (18), 4903–4909. doi:10.1158/1078-0432.CCR-12-1100

Anna, A., and Monika, G. (2018). Splicing Mutations in Human Genetic Disorders: Examples, Detection, and Confirmation. *J. Appl. Genet.* 59 (3), 253–268. doi:10.1007/s13353-018-0444-7

Bonati, M. T., Verde, F., Hladnik, U., Cattelan, P., Campana, L., Castronovo, C., et al. (2017). A Novel Nonsense ATP7A Pathogenic Variant in a Family Exhibiting a Variable Occipital Horn Syndrome Phenotype. *Mol. Genet. Metab. Rep.* 13, 14–17. doi:10.1016/j.ymgmr.2017.07.007

Burdick, K. J., Cogan, J. D., Rives, L. C., Robertson, A. K., Koziura, M. E., Brokamp, E., et al.Undiagnosed Diseases Network (2020). Limitations of Exome Sequencing in Detecting Rare and Undiagnosed Diseases. *Am. J. Med. Genet.* 182 (6), 1400–1406. doi:10.1002/ajmg.a.61558

Cai, C., Yu, C., and Shu, J. (2021). Recommendations on the Clinical Significance Grading of Genetic Variation in Clinical Gene Testing. *Tianjin Med. J.* 49 (6), 561–569.

Caicedo-Herrera, G., Candelo, E., Pinilla, J. F., Vidal, A., Cruz, S., and Pachajoa, H. (2018). Novel ATP7A gene Mutation in a Patient with Menkes Disease. *Tacg* Vol. 11, 151–155. doi:10.2147/TACG.S180087

Cao, B., Yang, X., Chen, Y., Huang, Q., Wu, Y., Gu, Q., et al. (2017). Identification of Novel ATP7A Mutations and Prenatal Diagnosis in Chinese Patients with Menkes Disease. *Metab. Brain Dis.* 32 (4), 1123–1131. doi:10.1007/s11011-017-9985-4

Chen, J., Jiang, Y., Shi, H., Peng, Y., Fan, X., and Li, C. (2020). The Molecular Mechanisms of Copper Metabolism and its Roles in Human Diseases. *Pflugers Arch. - Eur. J. Physiol.* 472 (10), 1415–1429. [Journal Article; Research Support, Non-U.S. Gov't; Review]. doi:10.1007/s00424-020-02412-2

Cummings, B. B., Marshall, J. L., Tukiainen, T., Lek, M., Donkervoort, S., Foley, A. R., Bolduc, V., Waddell, L. B., Sandaradura, S. A., O'Grady, G. L., Estrella, E., Reddy, H. M., Zhao, F., Weisburd, B., Karczewski, K. J., O'Donnell-Luria, A. H., Birnbaum, D., Sarkozy, A., Hu, Y., Gonorazky, H., Claeys, K., Joshi, H., Bournazos, A., Oates, E. C., Ghaoui, R., Davis, M. R., Laing, N. G., Topf, A., Kang, P. B., Beggs, A. H., North, K. N., Straub, V., Dowling, J. J., Muntoni, F., Clarke, N. F., Cooper, S. T., Bönnemann, C. G., and MacArthur, D. G. Genotype-Tissue Expression Consortium (2017). Improving Genetic Diagnosis in Mendelian Disease with Transcriptome Sequencing. *Sci. Transl. Med.* 9 (386). [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't]. doi:10.1126/scitranslmed.aal5209

de Gemmis, P., Enzo, M. V., Lorenzetto, E., Cattelan, P., Segat, D., and Hladnik, U. (2017). 13 Novel Putative Mutations in ATP7A Found in a Cohort of 25 Italian Families. *Metab. Brain Dis.* 32 (4), 1173–1183. doi:10.1007/s11011-017-0010-8

Elliott, A. M. (2020). Genetic Counseling and Genome Sequencing in Pediatric Rare Disease. *Cold Spring Harb Perspect. Med.* 10 (3), a036632. doi:10.1101/cshperspect.a036632

Frésard, L., Smail, C., Smail, C., Ferraro, N. M., Teran, N. A., Li, X., Smith, K. S., Bonner, D., Kernohan, K. D., Marwaha, S., Zappala, Z., Balliu, B., Davis, J. R., Liu, B., Prybol, C. J., Kohler, J. N., Zastrow, D. B., Reuter, C. M., Fisk, D. G., Grove, M. E., Davidson, J. M., Hartley, T., Joshi, R., Strober, B. J., Utiramerur, S., Lind, L., Ingelsson, E., Battle, A., Bejerano, G., Bernstein, J. A., Ashley, E. A., Boycott, K. M., Merker, J. D., Wheeler, M. T., and Montgomery, S. B.Undiagnosed Diseases Network; Care4Rare Canada Consortium (2019). Identification of Rare-Disease Genes Using Blood Transcriptome Sequencing and Large Control Cohorts. *Nat. Med.* 25 (6), 911–919. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't]. doi:10.1038/s41591-019-0457-8

Gonorazky, H. D., Naumenko, S., Ramani, A. K., Nelakuditi, V., Mashouri, P., Wang, P., et al. (2019). Expanding the Boundaries of RNA Sequencing as a Diagnostic Tool for Rare Mendelian Disease. *Am. J. Hum. Genet.* 104 (3), 466–483. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't]. doi:10.1016/j.ajhg.2019.01.012

Hartwig, C., Zlatic, S. A., Wallin, M., Vrailas-Mortimer, A., Fahrni, C. J., and Faundez, V. (2019). Trafficking Mechanisms of P-type ATPase Copper Transporters. *Curr. Opin. Cel Biol.* 59, 24–33. [Journal Article; Research Support, N.I.H., Extramural; Review]. doi:10.1016/j.ceb.2019.02.009

Horn, N., and Wittung-Stafshede, P. (2021). ATP7A-Regulated Enzyme Metalation and Trafficking in the Menkes Disease Puzzle. *Biomedicines* 9 (4), 391. [Journal Article; Review]. doi:10.3390/biomedicines9040391

Jelin, A. C., and Vora, N. (2018). Whole Exome Sequencing. *Obstet. Gynecol. Clin. North America* 45 (1), 69–81. doi:10.1016/j.ogc.2017.10.003

Kaler, S. G., and DiStasio, A. T. (1993). *ATP7A-Related Copper Transport Disorders*. [Review; Book Chapter].

Kaler, S. G., Ferreira, C. R., and Yam, L. S. (2020). Estimated Birth Prevalence of Menkes Disease and ATP7A-Related Disorders Based on the Genome

Aggregation Database (gnomAD). *Mol. Genet. Metab. Rep.* 24, 100602. doi:10.1016/j.ymgmr.2020.100602

Kodama, H., Fujisawa, C., and Bhadhprasit, W. (2012). Inherited Copper Transport Disorders: Biochemical Mechanisms, Diagnosis, and Treatment. *Cdm* 13 (3), 237–250. [Journal Article; Research Support, Non-U.S. Gov't; Review]. doi:10.2174/138920012799320455

Kremer, L. S., Bader, D. M., Mertes, C., Kopajtich, R., Pichler, G., Iuso, A., et al. (2017). Genetic Diagnosis of Mendelian Disorders via RNA Sequencing. *Nat. Commun.* 8, 15824. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1038/ncomms15824

Lappalainen, T., Scott, A. J., Brandt, M., and Hall, I. M. (2019). Genomic Analysis in the Age of Human Genome Sequencing. *Cell* 177 (1), 70–84. [Journal Article; Research Support, N.I.H., Extramural; Research Support, Non-U.S. Gov't; Review]. doi:10.1016/j.cell.2019.02.032

Lee, H., Huang, A. Y., Wang, L.-k., Yoon, A. J., Renteria, G., Eskin, A., Signer, R. H., Dorrani, N., Nieves-Rodriguez, S., Wan, J., Douine, E. D., Woods, J. D., Dell'Angelica, E. C., Fogel, B. L., Martin, M. G., Butte, M. J., Parker, N. H., Wang, R. T., Shieh, P. B., Wong, D. A., Gallant, N., Singh, K. E., Tavyev Asher, Y. J., Sinsheimer, J. S., Krakow, D., Loo, S. K., Allard, P., Papp, J. C., Palmer, C., Martinez-Agosto, J. A., and Nelson, S. F. Undiagnosed Diseases Network (2020). Diagnostic Utility of Transcriptome Sequencing for Rare Mendelian Diseases. *Genet. Med.* 22 (3), 490–499. [Journal Article; Research Support, N.I.H., Extramural]. doi:10.1038/s41436-019-0672-1

Maung, M. T., Carlson, A., Olea-Flores, M., Elkhadragy, L., Schachtschneider, K. M., Navarro-Tito, N., et al. (2021). The Molecular and Cellular Basis of Copper Dysregulation and its Relationship with Human Pathologies. *FASEB j.* 35 (9), e21810. doi:10.1096/fj.202100273RR

Møller, L. B., Mogensen, M., and Horn, N. (2009). Molecular Diagnosis of Menkes Disease: Genotype-Phenotype Correlation. *Biochimie* 91 (10), 1273–1277. [Journal Article; Review]. doi:10.1016/j.biochi.2009.05.011

Møller, L. B., Tümer, Z., Lund, C., Petersen, C., Cole, T., Hanusch, R., et al. (2000). Similar Splice-Site Mutations of the ATP7A Gene lead to Different Phenotypes: Classical Menkes Disease or Occipital Horn Syndrome. *Am. J. Hum. Genet.* 66 (4), 1211–1220. [Comparative Study; Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1086/302857

Pan, Y., Lu, T., Peng, L., Zeng, Q., Huang, X., Yao, X., et al. (2021). Functional Analysis of Ectodysplasin-A Mutations in X-Linked Nonsyndromic Hypodontia and Possible Involvement of X-Chromosome Inactivation. *Stem Cell Int.* 2021, 1–10. [Journal Article]. doi:10.1155/2021/7653013

Parada, G. E., Munita, R., Cerda, C. A., and Gysling, K. (2014). A Comprehensive Survey of Non-canonical Splice Sites in the Human Transcriptome. *Nucleic Acids Res.* 42 (16), 10564–10578. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1093/nar/gku744

Patel, P., Prabhu, A. V., and Benedek, T. G. (2017). The History of John Hans Menkes and Kinky Hair Syndrome. *JAMA Dermatol.* 153 (1), 54. [Biography; Historical Article; Journal Article]. doi:10.1001/jamadermatol.2016.0163

Perge, P., and Igaz, P. (2019). Family Screening and Genetic Counseling. *Exp. Suppl.* 111, 29–32. [Journal Article; Review]. doi:10.1007/978-3-030-25905-1_3

Petris, M. J., Strausak, D., and Mercer, J. F. (2000). The Menkes Copper Transporter Is Required for the Activation of Tyrosinase. *Hum. Mol. Genet.* 9 (19), 2845–2851. [Journal Article; Research Support, Non-U.S. Gov't]. doi:10.1093/hmg/9.19.2845

Prithvi, A., Sharawat, I. K., Saini, A. G., and Singh, P. (2019). Epilepsy and Neurodegeneration: Clues in the Hair and Blood Vessels!. *J. Pediatr.* 206, 293. [Case Reports; Journal Article]. doi:10.1016/j.jpeds.2018.09.069

Pu, D., Wu, J., and Liu, J. (2010). Skewed X Chromosome Inactivation May Be Not Associated with Premature Ovarian Failure. *Gynecol. Endocrinol.* 26 (6), 423–428. [Journal Article; Meta-Analysis]. doi:10.3109/09513591003632217

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al.ACMG Laboratory Quality Assurance Committee (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 (5), 405–424. [Consensus Development Conference; Guideline; Journal Article; Research Support, N.I.H., Extramural]. doi:10.1038/gim.2015.30

Sato, K., Uehara, S., Hashiyada, M., Nabeshima, H., Sugawara, J.-i., Terada, Y., et al. (2004). Genetic Significance of Skewed X-Chromosome Inactivation in Premature Ovarian Failure. *Am. J. Med. Genet.* 130A (3), 240–244. [Journal Article]. doi:10.1002/ajmg.a.30256

Skjørringe, T., Amstrup Pedersen, P., Salling Thorborg, S., Nissen, P., Gourdon, P., and Birk Møller, L. (2017). Characterization of ATP7A Missense Mutants Suggests a Correlation between Intracellular Trafficking and Severity of Menkes diseaseResearch Support, Non-U.S. Gov't]. *[Journal Articlesci Rep.* 7 (1), 757. doi:10.1038/s41598-017-00618-6

Skjørringe, T., Tümer, Z., and Møller, L. B. (2011). Splice Site Mutations in the ATP7A Gene. *PLoS One* 6 (4), e18599. doi:10.1371/journal.pone.0018599

Smpokou, P., Samanta, M., Berry, G. T., Hecht, L., Engle, E. C., and Lichter-Konecki, U. (2015). Menkes Disease in Affected Females: The Clinical Disease Spectrum. *Am. J. Med. Genet.* 167 (2), 417–420. doi:10.1002/ajmg.a.36853

Thouin, M. M., Giron, J. M., and Hoffman, E. P. (2003). Detection of Nonrandom X Chromosome Inactivation. *Curr. Protoc. Hum. Genet.* 35, t7–t9. Chapter 9. doi:10.1002/0471142905.hg0907s35

Tümer, Z. (2013). An Overview and Update ofATP7AMutations Leading to Menkes Disease and Occipital Horn Syndrome. *Hum. Mutat.* 34 (3), 417–429. doi:10.1002/humu.22266

Yang, M., and Kim, J.-W. (2018). Principles of Genetic Counseling in the Era of Next-Generation Sequencing. *Ann. Lab. Med.* 38 (4), 291–295. doi:10.3343/alm.2018.38.4.291

Yoon, S. H., Choi, Y. M., Hong, M. A., Kang, B. M., Kim, J. J., Min, E. G., et al. (2008). X Chromosome Inactivation Patterns in Patients with Idiopathic Premature Ovarian Failure. *Hum. Reprod.* 23 (3), 688–692. doi:10.1093/humrep/dem415

Zhang, J., Yao, Y., He, H., and Shen, J. (2020). Clinical Interpretation of Sequence Variants. *Curr. Protoc. Hum. Genet.* 106 (1), e98. doi:10.1002/cphg.98

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.

frontiers | Frontiers in Genetics

METHODS
published: 04 April 2022
doi: 10.3389/fgene.2022.852472

The Economic, Medical and Psychosocial Consequences of Whole Genome Sequencing for the Genetic Diagnosis of Patients With Intellectual Disability: The DEFIDIAG Study Protocol

Catherine Lejeune[1,2], Charley Robert-Viard[1,3], Nicolas Meunier-Beillard[1,3], Myriam Alice Borel[4], Léna Gourvès[5], Stéphanie Staraci[6], Anne-Laure Soilly[3], Francis Guillemin[7], Valerie Seror[8], Hamza Achit[7], Marion Bouctot[1], Marie-Laure Asensio[1], Anne-Sophie Briffaut[1], Christelle Delmas[9], Ange-Line Bruel[10], Alexia Benoit[11], Alban Simon[12], Bénédicte Gerard[11], Hamza Hadj Abdallah[13,14], Stanislas Lyonnet[13,14], Laurence Faivre[10], Christel Thauvin-Robinet[10], Sylvie Odent[15], Delphine Heron[6], Damien Sanlaville[16], Thierry Frebourg[17,18], Jean Muller[11,12,19], Yannis Duffourd[10], Anne Boland[20], Jean-François Deleuze[20], Hélène Espérou[9], Christine Binquet[1] and Hélène Dollfus[12] for the DEFIDIAG Study group*

[1]CHU Dijon Bourgogne, Inserm, Université de Bourgogne, CIC 1432, Module Épidémiologie Clinique, Dijon, France, [2]Inserm, Université Bourgogne-Franche-Comté, UMR 1231, EPICAD, Dijon, France, [3]CHU Dijon Bourgogne, Délégation à la Recherche Clinique et à l'Innovation, USMR, Dijon, France, [4]Observatoire Régional de Santé Bourgogne Franche-Comté, Dijon, France, [5]CHU Dijon Bourgogne, Direction de la Recherche Clinique, Dijon, France, [6]Unité Fonctionnelle de Génétique Médicale et Centre de Référence « Déficiences Intellectuelles de Causes Rares », APHP Sorbonne Université, Groupe Hospitalier Pitié-Salpêtrière et Hôpital Trousseau, Paris, France, [7]CIC1433-Epidémiologie Clinique, Centre Hospitalier Régional et Universitaire, Inserm, Université de Lorraine, Nancy, France, [8]Aix Marseille Univ, IRD, APHM, SSA, VITROME, IHU-Méditerranée Infection, Marseille, France, [9]Inserm, Pôle de Recherche Clinique, Paris, France, [10]CHU Dijon Bourgogne, Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Dévelopment (TRANSLAD), Inserm, Université Bourgogne-Franche-Comté, UMR1231, Équipe GAD, Dijon, France, [11]Laboratoires de Diagnostic Génétique, Institut de Génétique Médicale d'Alsace (IGMA), Hôpitaux Universitaires de Strasbourg, Strasbourg, France, [12]Inserm UMRS_1112, Institut de Génétique Médicale d'Alsace, Université de Strasbourg, France et Service de Génétique Médicale Hôpitaux Universitaires de Strasbourg, Strasbourg, France, [13]Inserm, IHU Imagine—Institut des Maladies Génétiques, Université Paris Cité, Paris, France, [14]Fédération de Génétique et Médecine Génomique, Hôpital Necker-Enfants Malades, GHU APHP. Centre-Université Paris Cité, Paris, France, [15]Service de Génétique Clinique, Centre de Référence Anomalies du Dévelopment CLAD- Ouest, CNRS, IGDR UMR6290 (Institut de Génétique et Dévelopment de Rennes), ERN ITHACA, Université de Rennes, Rennes, France, [16]Hospices Civils de Lyon, GHE, Service de Génétique, Université Claude Bernard Lyon 1, Lyon, France, [17]CHU de Rouen, Service de Génétique, Rouen, France, [18]Inserm, UMR1245, Centre de Génomique et de Médecine Personnalisée, Université de Normandie, Rouen, France, [19]Unité Fonctionnelle de Bioinformatique Médicale Appliquée au Diagnostic (UF7363), Hôpitaux Universitaires de Strasbourg, Strasbourg, France, [20]CEA, Centre National de Recherche en Génomique Humaine (CNRGH), Université Paris-Saclay, Evry, France

OPEN ACCESS

Edited by:
Bo Xiong, Huazhong University of Science and Technology, China

Reviewed by:
Massimo Carella, Home for Relief of Suffering (IRCCS), Italy
Conrad O. Iyegbe, Icahn School of Medicine at Mount Sinai, United States

***Correspondence:**
Catherine Lejeune
catherine.lejeune@u-bourgogne.fr

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *11 January 2022*
Accepted: *08 March 2022*
Published: *04 April 2022*

Citation:
Lejeune C, Robert-Viard C, Meunier-Beillard N, Borel MA, Gourvès L, Staraci S, Soilly A-L, Guillemin F, Seror V, Achit H, Bouctot M, Asensio M-L, Briffaut A-S, Delmas C, Bruel A-L, Benoit A, Simon A, Gerard B, Hadj Abdallah H, Lyonnet S, Faivre L, Thauvin-Robinet C, Odent S, Heron D, Sanlaville D, Frebourg T, Muller J, Duffourd Y, Boland A, Deleuze J-F, Espérou H, Binquet C and Dollfus H (2022) The Economic, Medical and Psychosocial Consequences of Whole Genome Sequencing for the Genetic Diagnosis of Patients With Intellectual Disability: The DEFIDIAG Study Protocol. Front. Genet. 13:852472. doi: 10.3389/fgene.2022.852472

Abbreviations: HAS, French national health authority; ANPGM, national association of molecular genetics practitioners; CMA, chromosomal microarray analysis; DNA, deoxyribonucleic acid; DRG, diagnosis related groups; ENC, echelle nationale des coûts; WES, whole exome sequencing; e-CRF, e-case report form; GPS, gene panel sequencing; 44GPS, 44 gene panel strategy; HSS, humanities and social sciences; ICER, incremental cost-effectiveness ratio; ID, intellectual deficiency; MRI, magnetic resonance imaging; NGS, next generation sequencing; NMB, net monetary benefit; PFMG 2025, plan france médecine génomique 2025; WGS, whole genome sequencing; WGS_T, trio whole genome sequencing; WGS_S, solo whole genome sequencing.

Introduction: Like other countries, France has invested in a national medical genomics program. Among the four pilot research studies, the DEFIDIAG project focuses on the use of whole genome sequencing (WGS) for patients with intellectual disability (ID), a neurodevelopmental condition affecting 1–3% of the general population but due to a plethora of genes. However, the access to genomic analyses has many potential individual and societal issues in addition to the technical challenges. In order to help decision-makers optimally introduce genomic testing in France, there is a need to identify the socio-economic obstacles and leverages associated with the implementation of WGS.

Methods and Analysis: This humanities and social sciences analysis is part of the DEFIDIAG study. The main goal of DEFIDIAG is to compare the percentage of causal genetic diagnoses obtained by trio WGS (including the patient and both parents) (WGS_T) to the percentage obtained using the minimal reference strategy currently used in France (Fragile-X testing, chromosomal microarray analysis, and gene panel strategy including 44 ID genes) for patients with ID having their first clinical genetics consultation. Additionally, four complementary studies will be conducted. First, a cost-effectiveness analysis will be undertaken in a subsample of 196 patients consulting for the first time for a genetic evaluation; in a blinded fashion, WGS_T and solo (index case, only) genomic analysis (WGS_S) will be compared to the reference strategy. In addition, quantitative studies will be conducted: the first will estimate the cost of the diagnostic odyssey that could potentially be avoidable with first-line WGS_T in all patients previously investigated in the DEFIDIAG study; the second will estimate changes in follow-up of the patients in the year after the return of the WGS_T analysis compared to the period before inclusion. Finally, through semi-directive interviews, we will explore the expectations of 60 parents regarding genomic analyses.

Discussion: Humanities and social sciences studies can be used to demonstrate the efficiency of WGS and assess the value that families associate with sequencing. These studies are thus expected to clarify trade-offs and to help optimize the implementation of genomic sequencing in France.

Ethics Statement: The protocol was approved by the Ethics Committee Sud Méditerranée I (June 2019)—identification number: 2018-A00680-55 and the French data privacy commission (CNIL, authorization 919361).

Clinical Trial Registration: (ClinicalTrials.gov), identifier (NCT04154891).

Keywords: intellectual disability, genome sequencing, cost-effectiveness, qualitative study, micro-costing

INTRODUCTION

Genomics in medicine is profoundly modifying our understanding and our clinical practices. To ensure that access to these new technologies is equitably distributed throughout the country, France launched in 2016 a national plan for genomic medicine (Aviesan, 2025; PFMG 2025), similarly to what had already been implemented in other countries (Collins et al., 2003; Zhao et al., 2004; Peplow, 2016; Aviesan, 2025). This national plan aims to change the way patients are diagnosed, followed-up, and treated by 2025 in various medical specialties such as cancer and rare diseases, and to extend this to common diseases (Zhao et al., 2004). By setting up routine genome sequencing, the ambition is to enable more personalized diagnosis and therapeutic management of patients.

Rare diseases are considered to be at the forefront for the implementation of next generation sequencing (NGS) in France as part of the developing genomic strategies. Among rare diseases, intellectual deficiency (ID) is the most common reason for referral in genetic centers. This neurodevelopmental disorder is characterized by intellectual quotients (IQ) under 70 before the age of 18. It affects between 1 and 3% of the general population, with around 15 per 1,000 persons with mild ID and around 3 per 1000 with severe ID (Buntinx et al., 2016). ID is extremely genetically heterogeneous, which makes it difficult to obtain a genetic diagnosis. The combination of

these characteristics has made ID a major public health challenge, and there is an urgent need to improve the rate of diagnosis for individuals with ID in order to offer them optimal care.

The emergence of NGS is a real technological breakthrough for the molecular diagnosis of ID. It was used to develop gene panel sequencing (GPS), in which a selection of genes is captured and sequenced, whole exome sequencing (WES), and more recently whole genome sequencing (WGS). In France, the sequencing strategy for rare diseases is still under debate, and studies are needed to define the diagnostic yield of the different NGS techniques as well as replacement of chromosomal microarray analysis (CMA) by WGS for instance. The evaluation of the efficiency of GPS and WES compared to the strategy using Fragile-X (Fra-X) testing, CMA and a gene panel of 44 selected ID genes, which has been historically used in France, is ongoing through an economic research program funded by the French Ministry of Health. But WGS constitutes another apparent opportunity for patients with ID. Its diagnostic yield is estimated to be 60–68% (Van Nimwegen, 2017) compared with 43% using WES (Gilissen et al., 2014) and up to only 32% using GPS (Redin et al., 2014; Chérot et al., 2018).

To determine the optimal conditions for the implementation and the generalization of WGS in clinical practice in a context of limited resources, decision-makers need to be provided with complete and comprehensive information. In this context, one of the four pilot studies of the PFMG 2025, the DEFIDIAG study, was dedicated specifically to ID. Beyond the technical challenge of improving the diagnostic yield, this study represents an opportunity to explore other humanities and social issues, which concern the patients and their families, the clinicians, the payers and decision-makers, and the health care system as a whole. Studies have been already been published on the complete cost and/or efficiency of WGS (Jegathisawaran et al., 2020; Yuen et al., 2018; Tslipova et al., 2017; Van Nimwegen, 2017; Ontario Health, 2020; Schwartze et al., 2020). Some recent studies also focused on the diagnostic costs avoided by the use of first-line genetic WES (Monroe et al., 2016; Stark et al., 2017; Vissers et al., 2017) and more rarely WGS (Ontario Health, 2020). But heterogeneous clinical presentation, sample sizes, and differences in methodologies used make it difficult to compare these results and to generalize to other settings. Finally, authors already explored the perception that families have of diagnostic results (Foster et al., 2009; Kohler et al., 2017; Mollison et al., 2020), but to the best of our knowledge, no study has explored their experience of the entire process of care (from the prescription of genetic analysis until the disclosure of results and a post-result period). A longitudinal investigation may help to better understand their needs and expectations as well as their perception of the utility of the results for themselves and for their relatives.

In this context, this article aims to present the humanities and social sciences (HSS) dimensions of WGS in the French context based on the population of the DEFIDIAG project. Four dimensions will be explored: 1) efficiency of WGS, 2) impact of WGS on cost savings, 3) impact of WGS on the medical, medico-social, rehabilitative and psychological follow-up of patients presenting ID and 4) investigation of the experience of the parents concerning the health care pathway as a whole.

METHODS AND ANALYSES

The HSS study is part of the DEFIDIAG project, which is a prospective multicenter diagnostic study (Binquet et al., 2022).

Summary of the DEFIDIAG Study

The main goal of the DEFIDIAG project (NCT04154891) is to compare the percentage of ID causal diagnoses obtained using 2 different strategies: WGS using a trio strategy (WGS_T) and the reference strategy [*i.e.*, use of the French guidelines based on the ANPGM (National Association of Molecular Genetics Practitioners)] applied blindly to consecutive patients with no obvious diagnosis. The population is composed of patients between 0 and 5 years meeting stringent criteria (severe delayed development in terms of motor skills, language, and/or sociability) or patients older than 6 years whatever the ID severity (but with proven ID by *ad hoc* neuropsychological testing) and the associated manifestations, and without any obvious diagnosis identified during a genetic consultation.

A total of 1,275 patients will be then included in one of the 14 participating clinical centers, with 50% of patients coming for their first consultation (patients never explored) and the other 50% having had a previous genetic exploration. Each included patient will be their own control; they will benefit from the two main strategies in parallel. WGS_T and the reference strategy will be compared in 7 subgroups: 3 subgroups defined according to age (<2 years old/2–5 years/>5 years), 4 subgroups of patients defined according the severity of ID, and/or with associated manifestations (ID associated with major non-cerebral manifestation, moderate to severe ID, mild ID associated with another sign, ID associated with epilepsy). The diagnostic yield of WGS using a simplex strategy (WGS_S) will be also investigated in parallel to the two main strategies, but only in a randomized subgroup of the overall population consulting a geneticist for the first time.

Objectives of the HSS Study

The HSS study consists of four parts, which have been designed in conjunction with the DEFIDIAG study:

- a *cost-effectiveness analysis* that will be undertaken in the population consulting a geneticist for the first time.
- a study aiming to estimate *the cost of the diagnostic odyssey* which could have potentially been avoided with a first-line WGS_T in previously investigated patients.
- a study whose goal will be to estimate the *frequency and nature of changes in patient follow-up* in the first year after the WGS_T analyses compared to the period before inclusion for both previously investigated patients and the population consulting for the first time.
- a *qualitative study* which will explore parent expectations regarding the genomic analyses, how they feel about the results and how they perceive the future.

The whole DEDIDIAG study, including the HSS analyses, has been approved by ad hoc ethics committee (Identification number 2018-A00680-55). The inclusion of patients is ongoing.

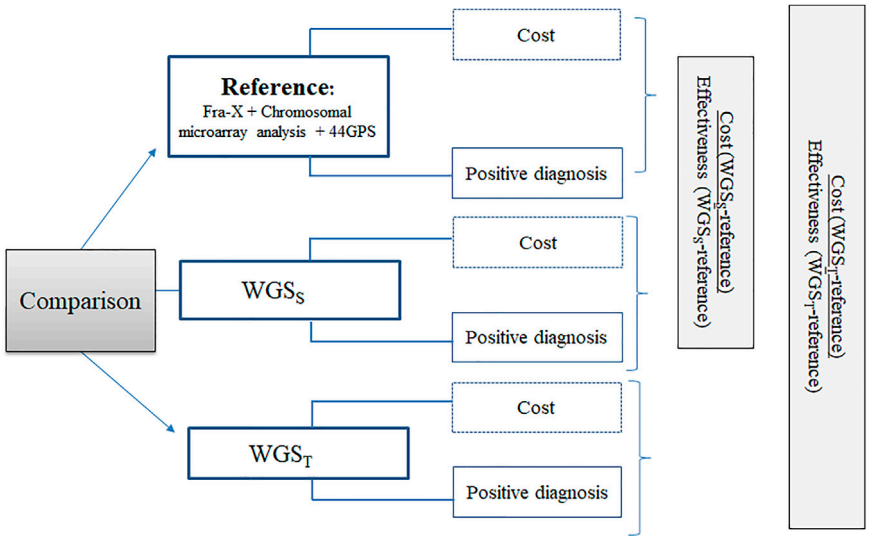


FIGURE 1 | Design of the efficiency study: This figure illustrates the design of the cost-effectiveness study. Three strategies will be compared: the solo Whole Genome Sequencing strategy (WGS_S), the trio Whole Genome Sequencing strategy (WGS_T) and the reference strategy. Comparisons will be made simultaneously in terms of cost and effectiveness (positive diagnosis).

Methods of the HSS Studies
Efficiency Study

Compared Strategies
The aim of this efficiency study is to compare the following three strategies in terms of cost and effectiveness for the causal diagnosis of ID in the first investigated population: the French core minimal reference strategy combining Fra-X testing, CMA, and a panel of genes commonly known to be involved in ID (44 GPS), WGS_T and WGS_S (**Figure 1**).

Sample Size
Overall 9 comparisons had to be performed in the main project given the 7 subgroups to account for and the comparison between WGS_T and WGS_S. Thus, the alpha risk was set to 0,00278 and the target power to 80%. A difference of 7% was expected between WGS_T and WGS_S (66–68% as reported by Gilissen et al., 2014 for WGS_T and 60% by Lionel et al., 2017 for WGS_S) and less than 0.1% of diagnosis identified by WGS_S and not by WGS_T. Assuming these assumptions, the sample size required was 189 patients. We planned to include 7 additional patients (3.5%) to account for unusable samples or other technical problems.

Effectiveness
The effectiveness criterion used in the cost-effectiveness analysis will be the primary study end point of the DEFIDIAG study, *i.e.*, the identification of a causal diagnosis of ID defined as the identification of one class 4 (likely pathogenic) or 5 (pathogenic) variant that explains the symptoms presented by the patient.

Identification and Measurements of Data for the Estimation of Costs
The planned time frame for the cost-effectiveness analysis is 12 months, which is the estimated maximum time required to perform from the WGS_T and WGS_S analyses and interpret the data, and to return the results to the patient and their parents.

The economic evaluation will be conducted from the point of view of the "collectivity," as recommended by the French National Health Authority (HAS), meaning that all the stakeholders involved in the decision are considered. In this perspective, the production costs of strategies should be identified, measured and valued independently of their current or envisaged sources of funding. Thinking in terms of production costs makes it possible to have an in-depth analysis of the resources used and allows the stakeholders to think about their participation in terms of funding, and to determine or adjust tariffs.

Only direct medical costs will be considered. They will include: (i) consultation with the clinical geneticist in the recruiting centers, as well as exams preceding the genetic analysis and inclusion in the DEFIDIAG project (brain magnetic resonance imaging (MRI) and neuropsychological assessment carried out according to the French recommendations); (ii) first blood sample; (iii) Fra-X testing, CMA, 44GPS, WGS_T/WGS_S and (iiii) possible complementary (imaging) and confirmatory exams (such as Sanger and qPCR: quantitative Polymerase Chain Reaction). The first four categories of costs will be recorded for each patient. To identify the complementary examinations, investigator physicians will be required to indicate the additional examinations/visits that they would consider prescribing for each patient in the e-Case Report Form (e-CRF) at the time of inclusion. They must complete

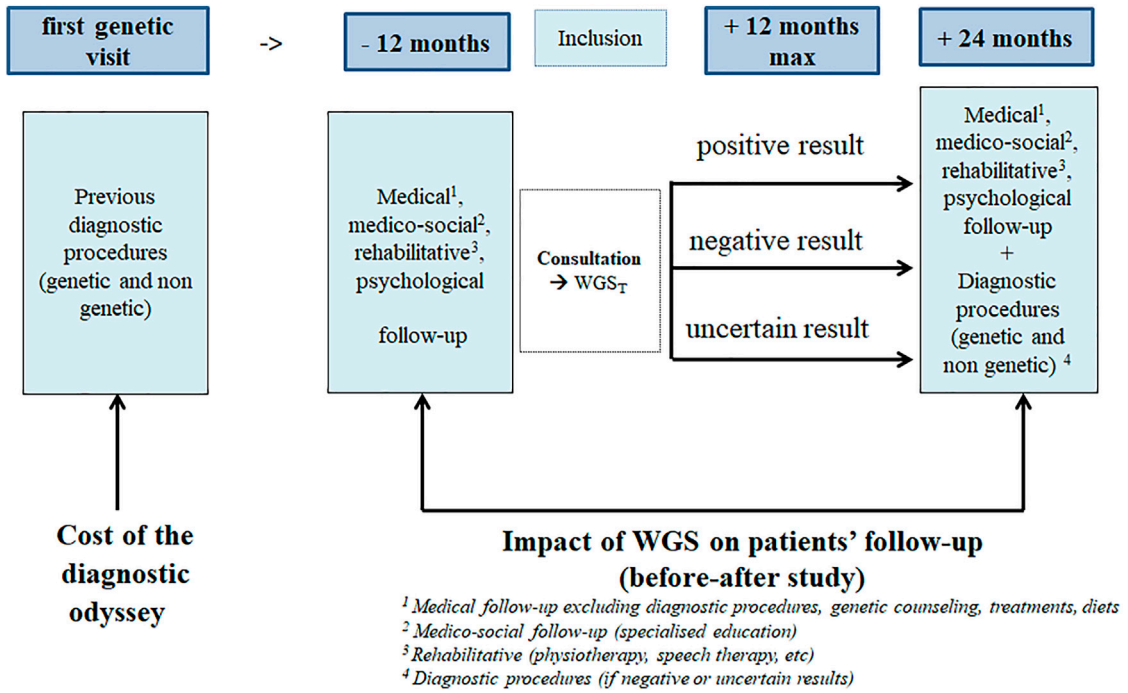


FIGURE 2 | Design of the impact of genome sequencing on the diagnostic odyssey and follow-up: This figure illustrates the design of the before-after study aiming at assessing the impact of WGS on patients' follow-up (12 months after the reporting of WGS results) compared to the period of 1 year preceding the inclusion.

the form for each of the three strategies independently of each other. Physicians will also be asked to indicate which confirmatory techniques they judge useful to perform for each strategy before unblinding and to record this information in the e-CRF during the multi-disciplinary meeting (MDM). The economic analysis will also consider the fact that these analyses could have been performed or not. If they are performed, the date will be collected. The use of a declarative method to identify the examinations is justified by the blinded design of the study.

Monetary Valuation of Ressources

The transport costs for the blood samples will be valued on the basis of invoices. Existing tariffs will be used for valuation of all other medical, biological and technical exams. The use of tariffs is considered acceptable by the HAS because they are considered as the counterpart of the resources consumed for carrying out these analyses. Otherwise, the microcosting method will be used to estimate the production costs of WGS_T and WGS_S (Frick, 2009). The cost will be obtained by considering the volumes of each mobilized resource (labor, disposable and reagents, material, equipment, etc.) and their associated monetary value (mean gross wages, purchase price of consumables, software, and equipment, etc.). The microcosting method will require the development of grids for collecting the resources used during the pre-analytical (DNA extraction and quality control), analytical (sample bank preparation, production, quality control and primary bioinformatic analysis including data-storage, bioinformatic and the post-analytical (interpretation by biologists, multidisciplinary meeting discussions and disclosure of results to the parents and patients) steps. The cost of the re-analysis of variants of unknown significance will not be taken into consideration because it goes beyond the 12-months time horizon of the analysis.

Cost of the Diagnostic Odyssey

The cost-effectiveness analysis will be completed by the estimation of the costs related to the iterative search for a diagnosis in the *previously investigated* population. The search could potentially include genetic investigations, biological tests, imaging procedures, specialized consultations, and/or hospitalization (**Figure 2**). The collection of previous examinations will be based on the investigation of the patient's medical file and the consultation including the medical geneticist, the patient and their parents. In case of discordance between the data sources, the interviews will be considered the definitive source to be used for the analysis. All previous examinations will be valued with the current tariffs. In case of hospitalisation, the cost of daily hospital stays will be estimated from the Diagnosis Related Groups (DRG) associated with each inpatient stay. They will be identified using the patient's main diagnosis, the medical procedures performed during the hospitalization, and the status of the health care centre (public, private). The cost of each DRG will be then estimated using a National Cost Survey Sample, named "*Echelle Nationale des Coûts*" (ENC), which provides costs data based on a sample of centres in France. The identification of DRGs and their monetary

TABLE 1 | Template of the qualitative study.

Interview 1: After the inclusion in the DEFIDIAG study		
Trajectory of care	Experience of the care pathway	Can you tell me about your child's care pathway? -Support -Care, type of relationship with professionals -Organization for daily life
	Representation of difficulties (historicity)	What were the difficulties encountered? -Access to knowledge -Information flow -Feelings at diagnosis/diagnosis delay
Positioning in relation to genetic research	Représentation de la génétique	What are your expectations regarding genetic research? -How long have you been searching for the reasons for your child's disability? *Reminder: here there will be a distinction between de novo versus parents who have been searching for years and differences in perception* -Fears and hopes possibly in the process of recovery
	Representation of genetics	- Could you tell me why you agreed to participate in the DEFIDIAG study? -What information did you receive about the study?
	Anticipation genetic results	- How do you feel about waiting for the results? -How do you think you will react? -What support would you like at this time?
	Secondary data	-What were the reasons for your decision to seek secondary data? →Do you have any particular expectations?
End		-Could you tell me three words that describe how you feel now?
Interview 2: After the results' disclosure		
Introductory question: Which results did you receive?		
The experience of the announcement of the results	Subjective dimension	Could you tell me how you felt when you got your results? - personal reactions - reactions of those around you - reactions to the people around you
	Context of the announcement	-Could you tell me how the results were announced to you?
Impact of the announcement of the results	Reception and appropriation of results	-What did you retain from the information given? -Did it raise any questions? -With whom did you discuss it?
Perception of the future	Expected changes	-Do you think these results will change things? -Which ones? -In the short/medium/long term - Do you think it will change the relationship you have with your child?
	Feeling	-Do these results lead to changes in perspective? -For you -For your child -Did you have any knowledge of the diseases mentioned? -Representations -Did you share these results with other family members? How did they react to this information?
Primary and secondary data		What if you had to do it again?
End		-Could you tell me three words that describe how you feel now?
Interview 3: 1 year after the disclosure of the results		
Changes since the announcement of the results	Organization of daily life	-Have there been any changes since the results? →concretely?
	Subjective dimension	-How do you feel since the results? -How often do you think about the results? -How have you experienced the period after the results? -Are there any changes in the way you think about the disease/treatment/follow-up? -Are there any changes in your relationships with your family and friends?
Evolution of the perception of genetics		-What do you think of the care provided by the genetics team? -Do you want to know the full results of the study?
If secondary data disclosure		-What did you learn about the secondary data

(Continued on following page)

TABLE 1 | (*Continued*) Template of the qualitative study.

Interview 1: After the inclusion in the DEFIDIAG study		
	Appropriation of results and feelings	-Have you discussed the secondary data with others? -Did the results change anything in your daily life? -Have you had/are you considering further tests? What if you had to do it again?
End	Perception of the future	-How do you see the future? -What kind of support would you like for the future? -Could you tell me three words that describe how you feel now?

valuation will be performed by each health care centre included in the DEFIDIAG project.

Impact of WGS on Patient Follow-Up

A before-after study is planned aiming to assess the frequency and nature of changes in medical follow-up (treatments, diets, medical supervision, genetic counseling to the patients and relatives), as well as medico-social (education, type of institution attended), rehabilitative (physiotherapy, speech therapy, etc.), and psychological follow-up in the first year after the reporting of WGS_T analyses (whether the results are positive, negative or uncertain), and to compare the findings to a period of 1 year maximum preceding the inclusion (**Figure 2**). In the "before" period, the consultation between the medical geneticist, the patient and their parents will allow us to collect the data which could be completed by medical files. In the "after" period, parents will be asked to fill in a diary. Quarterly, a clinical technician will call the parents to fill the e-CFR with the recorded data.

Qualitative Study

A qualitative study will be based on semi-directive sociological and psychological interviews with the parents of the patients at three different times: at inclusion, a few days after the WGS results are disclosed and 1 year after the results are disclosed. At inclusion, the family history, the representation of genetics, and the expectations regarding the genomic analysis will be explored. We will also explore how they anticipate the waiting time for the results (how do they think they will react and what support they would like at this time). After the disclosure of results, the objective is to assess the reaction of the family and how they anticipate the future, and at 1 year, the possible changes that the families have experienced, how they feel about the genomic analysis, and their expectations about the future.

Parents will be included in two main centres participating in the DEFIDIAG study: Dijon University Hospital and Pitié-Salpétrière University Hospital (Paris). The choice of these centers is justified by the need to have a diversity of parent profiles in terms of both socio-economic and cultural background. The qualitative study will be performed in a sub-sample of the whole population of the efficacy study. We hypothesized that 60 interviews (30 in sociology and 30 in psychology) will be sufficient to achieve data saturation in each approach (Hennink et al., 2017) since the qualitative template has been built by HSS researchers and that questions will be similar whatever their HSS background (**Table 1**). Inclusions of patients coming for a first consultation have to be as consecutive as possible to meet the standards of a diagnostic study, therefore guaranteeing the heterogeneity of the demographic, and also cultural and socio-economic profiles of parents. If possible, the parents will be stratified in subgroups (first investigation population vs. previously investigated population) and according to the clinical profile of the patient (mild, moderate or severe/profound ID).

Data concerning the family situation, the number of children (with and without disability), and the deprivation level (working status and education level) will be recorded at inclusion. Their phone number and current address will also be collected in order to provide this information to the sociologist and the psychologist, who will contact the participants to define a date and a place for the interview. The interview can take place at the health care center, at home or in a neutral place at the convenience of the participants. A phone call or a videoconference will be also possible. The parents can choose whether they prefer to be interviewed together or not.

Analysis

Efficiency Analysis

The efficiency criterion will be based on the estimation of an incremental cost-effectiveness ratio (ICER), expressed in terms of cost per additional positive diagnosis. The results will not be discounted given the time horizon. A deterministic analysis will be used to consider evolutions in the technological field, such as the use of different generations of sequencing machines and the automatization of some steps which could modify the relative part of labor in the cost estimation. Another analysis involves the completion or non-completion of the complementary and confirmatory examinations that will be carried out at the time of the inclusion (the main analysis will consider only the prescribed acts that were actually performed). A probabilistic analysis based on a non-parametric bootstrap analysis will be also performed in order to manage the uncertainty associated with sampling and to estimate the 95% confidence interval of the ICERs (Briggs and Alastair, 1999). Given the number of strategies compared (> 2), we will estimate the net monetary benefit (NMB) associated with each of the strategies of the study. The NMB represents the value of an intervention in monetary terms when a willingness to pay threshold (λ) for a unit of benefit (E) is known (York Health Economics Consortium, 2016). A positive NMB indicates that the intervention is cost-effective compared with the

alternative at the given willingness-to-pay threshold. The NMB is calculated for each selected value of λ, making it possible to plot the acceptability curves for each strategy. This analysis will enable decision-makers to further view the results of this evaluation in terms of their budgetary reality.

Cost of the Diagnostic Odyssey

The costs of the diagnostic odyssey will be described only in the population of previously investigated patients. The results will be expressed as means and standard deviations in case of normal distribution, and as medians with interquartile ranges otherwise. A subgroup analysis may be performed to specifically target ID characteristics. Mean costs will be then compared by an analysis of variance or by a Kruskall-Wallis test according to the conditions of application. A *p*-value below 0.05 is considered as significant.

Impact of WGS on Patient Follow-Up

The frequency at which changes are made to patient follow-up between the period prior to inclusion and the period following the results will be calculated with the 95% CI. A global analysis will be then performed, whatever the result of WGS_T (positive, negative, non-conclusive). Sub analyses will be then conducted according to the result of WGS_T. We will also provide a description of the new medical diagnostic procedures performed among patients whose results are negative or uncertain with the WGS.

Qualitative Study

The analysis of the interviews will be based on the following steps: open codification of transcribed interviews to identify as many topics as possible in the initial corpus; categorization of codified elements; careful re-reading of the corpus as a whole with the aim of clearly defining each category; linking categories; writing more detailed memos and designing explanatory diagrams; integration of the previous steps in order to identify the main points of the phenomenon; and theorization: the meticulous and exhaustive construction of the "multidimensionality" and the "multicausality" of the phenomenon of the relationships between needs, expectations, hopes, suffering, and the results of genetic analysis. For the psychological aspect, the interviews will be analysed using the general inductive method (David and Thomas, 2006), which uses the first three steps described above.

DISCUSSION

This part of the DEFIDIAG project illustrates the complementarity of the various HSS methodologies and their ability to extend beyond the primary goal of the initial study. Our ambition is to consider the economic, medical, sociological, and psychological dimensions of genomics in order to provide French decision-makers with in-depth information about the advantages and constraints of WGS. In the economic field, one recent study demonstrated that WGS was about US$ 1,000 less expensive than standard testing (which included CMA, Fra-X, targeted single-gene tests and GPS) and was more than two-fold effective (in number of molecular diagnoses) (Ontario Health, 2020). Such data is lacking in the field of ID in France. Another goal is to provide WGS complete costs. Schwarze et al. provided the first data in the United Kingdom in the field of a rare disease trio case based on the microcosting method. The cost was estimated to be £ 7,050 (US $ 9,330) per genome (Schwartze et al., 2020). Other results which have already been published presented various estimations from € 1,421 (US $ 1,602) (Van Nimwegen, 2016) to CAN $ 6,435 (US $ 4,975) (Jegathisawaran et al., 2020), mainly explained by differences in terms of methodology. Results are also dependent from local organizations. In France, this kind of data is essential to obtain to contribute to the determination of tariffs to be reimbursed by the national health insurance. This result will be then used in our cost-effectiveness analysis. These results could also be used in a decision-analysis model to consider the improvement in diagnostic performance with WGS.

One potential limit relative to this planned economic evaluation as part of the DEFIDIAG project is that effectiveness will be expressed in terms of the number of positive diagnoses. Our choice is justified by the need to remain consistent with the primary goal of the study. Cost-utility analyses are commonly used alongside economic evaluations. In these studies, effectiveness is expressed in terms of QALYs (Quality-adjusted Life-years), a bidimensional criterion based on the quantity of life lived weighted by an "utility" score which represents the satisfaction that patients attribute to a health state. One QALY equals 1 year of life in perfect health. But the first challenge of ID is to explain the disease by providing the parents with a causal diagnosis and decreasing the negative impact or burden of the impairment (Adithyan et al., 2017). In the state of the art, the questionnaires used today to assess utility (*e.g.,* EQ-5D™) are not totally adapted to ID. Moreover, economic evaluations can not yet combine the point of view of the parents and the patients. However, it is important to consider how genomic testing impacts families. As an alternative, we decided to obtain a longitudinal comprehensive view of the experience of the families during the genetic and care pathway of their children in the first year after the disclosure of the results. These data are fundamental to adjust, if necessary, the support provided to these families. Previously published studies on the perception that families have of the diagnostic results generally focused on the concept of "personal utility" (Foster et al., 2009; Kohler et al., 2017; Mollison et al., 2020). We choose not to base the interview questions on a specific concept. Concepts and models will be identified from the interviews using an inductive approach. The collected parameters, such as the age of the parents, their socio-economic level as well as the demographic and clinical profile of their child will be useful for the interpretation phase. Religious and spiritual beliefs and practices will not be collected, but the interviews make it possible to consider these aspects given the fact that the question of the value associated with genetics will be explored. Given the fact that eligible patients have to be as consecutive as possible to meet the standards of a diagnostic study, heterogeneity of the sampling of parents for the interviews will be guaranteed. Finally, the HSS DEFIDIAG study planned to assess the changes in the follow-up of the patients 1 year after the results. To our knowledge, the evaluation of a similar objective was conducted only in one

monocentric retrospective study. It compared the pre-WES to the post-WES costs among patients with ID, but only from a diagnostic point of view; neither the impact on treatments nor the impact for patients presenting negative result was considered (Vrijenhoek et al., 2018).

To conclude, decision-makers need to be given a clear demonstration of the efficiency of WGS, to be informed about how WGS will affect the medical care pathway, and to fully understand the fear and expectations of families, all of these factors are among the conditions required for its successful generalisation.

TRIAL STATUS

Recruiting is ongoing (1 224/1 275 patients included as of 03/01/2022).

FULL LIST OF CO-INVESTIGATORS OF THE DEFIDIAG STUDY GROUP

Centre National de Recherche en Génomique Humaine et al., Centre National de Recherche en Génomique Humaine: MEYER Vincent; CHU d'Angers: BONNEAU Dominique, BARTH Magalie, TESSARECH Marine, ZIEGLER Alban; CHU de Bordeaux: GOIZET Cyril, LACOMBE Didier, LEGENDRE Marine, MARGOT Henri, MICHAUD Vincent, NAUDION Sophie, ROORYCK THAMBO Caroline; **Dijon-Bourgogne and Inserm UMR1231- Equipe**, CHU Dijon-Bourgogne & Inserm UMR1231- Equipe GAD: BOURNEZ Marie, BRUEL Ange-Line, COLIN Estelle, DELANNE Julian, DENOMME-PICHON Anne-Sophie, GARDE Aurore, MOUTTON Sébastien, NAMBOT Sophie, PHILIPPE Christophe, SAFRAOU Hana, SORLIN Arthur, THAUVIN Christel, TRAN-MAU-THEM Fréderic, VITOBELLO Antonio; **Grenoble-Alpes et al.**, CHU Grenoble-Alpes: DIETRICH Klaus, DURAND Chantal, MAREY Isabelle, N'GUYEN-MOREL Marie-Ange, THEVENON Julien; **CHU de Lille et al.**, CHU de Lille: BOUTE Odile, CAUMES Roseline, COLSON Cindy, DIEUX Anne, GHOUMID Jamal, MARSILI Luisa, PETIT Florence, VANLERBERGHE Clémence, VINCENT-DELORME Catherine; **Hospices Civils de Lyon et al.**, Hospices Civils de Lyon: ARMAND Thibaud, CHATRON Nicolas, CURIE Aurore, DES PORTES Vincent, EDERY Patrick, HAYE Damien, LABALME Audrey, LESCA Gaëtan, MONIN Pauline, PONS Linda, PUTOUX Audrey, ROSSI Massimiliano, ROUGEOT Christelle, TILL Marianne; **CHU de Montpellier et al.**, CHU de Montpellier: BLANCHET Patricia, COUBES Christine, DEILLER Caroline, GENEVIEVE David, PINSON Lucile, WELLS Constance, WILLEMS Marjolaine; **CHU de Nantes et al.**, CHU de Nantes: ISIDOR Bertrand, MERCIER Sandra, NIZON Mathilde, VINCENT Marie; **Hôpital Necker-Enfants Malades**, Hôpital Necker-Enfants Malades (AP-HP): AMIEL Jeanne, BARCIA Giulia, BAUJAT Geneviève, CORMIER Valérie, GUIMIER Anne, HADJ ABDALLAH Hamza, MALAN Valérie, MARLIN Sandrine, MARZIN Pauline, MICHOT Caroline, ORMIERES Clothilde, RIO Marlène, ROMANA Serge; **Groupe Hospitalier Pitié-Salpêtrière et al.**, Groupe Hospitalier Pitié-Salpêtrière (AP-HP): AFENJAR Alexandra, BURGLEN Lydie, CHARLES Perrine, COURTIN Thomas, HEIDE Solveig, KEREN Boris, LEHALLE Daphné, MIGNOT Cyril, MOUTHON Linda, WHALEN Sandra; **CHU de Rennes et al.**, CHU de Rennes: FRADIN Mélanie, JEAN-MARCAIS Nolwenn, LAVILLAUREIX Alinoë, MOREL Godelieve, PASQUIER Laurent, QUELIN Chloé, RIOU Audrey, UGOLIN Mélissa; **CHU de Rouen: BREHIN Anne-Claire et al.**, CHU de Rouen: BREHIN Anne-Claire, CASSINARI Kévin, CHAMBON Pascal, GOLDENBERG Alice, GUERROT Anne-Marie, JOLY-HELAS Géraldine, LECOQUIERRE François, LEMEUR Nathalie, NICOLAS Gaël, SAUGIER-VEBER Pascale, VERA Gabriella; Hôpitaux Universitaires de Strasbourg: EL CHEHADEH Salima, CALMELS Nadège, HAUSHALTER Virginie, MAILLARD Pierre-Yves, MULLER Jean, PHILIPPE Anaïs, PITON Amélie, SCHAEFER Elise, SCHEIDECKER Sophie, SCHLUTH-BOLARD Caroline; Hôpital de la Timone (Hôpitaux Universitaires de Marseille): BUSA Tiffany, PHILIP-SARLES Nicole, RICCARDI Florence, SIGAUDY Sabine; Institut Imagine: NITSCHKE Patrick.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee Sud Méditerranée I (June 2019)—identification number: 2018-A00680-55 and the French data privacy commission (CNIL, authorization 919361). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

CL: design, coordination of the SHS studies, drafting the protocol, supervision of the study, writing the manuscript. CR-V, A-LS: supervision of the study, reviewing the manuscript. NM-B: design, co-coordination with CL of the qualitative study—reviewing the manuscript. MAB, LG, SS: conducting interviews of the qualitative study, reviewing the manuscript. FG, VS, HA: design of the medico-economic study, reviewing the manuscript. M-LA, MB: project administration, support to the coordination of the project, reviewing the manuscript. A-SB: data management, reviewing the manuscript. CD: project administration, drafting the DEFIDIAG protocol, resources management, supervision of the DEFIDIAG study, reviewing the manuscript. A-LB: design of the qualitative study, reviewing the manuscript. ABe, HA: providing data for the

micro-costing study. AS, BG, CT-R, JM, YD, ABo, J-FD: providing data for the micro-costing study, reviewing the manuscript.. SL, SO, DS, TF: design the DEFIDIAG study, reviewing the manuscript. LF, DH: enrolment of parents of the qualitative study, reviewing the manuscript. HE: representing the sponsor (Inserm) of the study, reviewing the manuscript. CB: design of the diagnostic study methodology, reviewing the manuscript. HD: design and coordination of the DEFIDIAG study, drafting the protocol, supervision of the study, reviewing the manuscript. This study makes use of data generated by the DEFIDIAG study sponsored by Institut national de la santé et de la recherche médicale (Inserm). A full list of centres who contributed to the generation of the data is available from https://defidiag.inserm.fr/ and *via* email from c16-110.coordo.isp@inserm.fr.

FUNDING

TheDEFIDIAG study is funded by grants from the French Ministry of Health in the framewok of the national French initiative for genomic medicine. Institut national de la santé et de la recherche médicale (Inserm) is the sponsor of the DEFDIAG study.

ACKNOWLEDGMENTS

The authors would like to thank the following clinical research technicians and genetic counsellors: AKLOUL Linda, BAUDIER Marie-Pierre, BAURAND Amandine, BELLENGIER Laurence BERNARD Céline, BOREL Myriam, CONSOLINO Emilie, CRANTELLE Laura, DE NADAI Narimène, DORIAN Virginie, EL AMRANI Lamia, FAUDET Anne, FOURNIER Chloé, GALIVEL-VOISINE Annastasia, GAUDILLAT Léa, GAUTHIER Marjolaine, GLAZUNOVA Olga, GOETZ Nathalie, GONDE Delphine, GOURVES-NOIZET Lena, GUEGAN Caroline, GUYON Laura, HAQUET Emmanuelle, KASTNER Claire, KHALIL Mirna, LAURENT Manon, LEXTREYT Barbara, MALLET Audrey, MOUKOSSI Caroline, OUMESSOUM Amina, NGOM Ndeye-Fatou, PELLEN Anne-Sophie, PELLETIER Valérie, POIREL Elisabeth, PROUTEAU Clément, RUBECK Coralie, STARACI Stéphanie, SAWKA Caroline, TEMPE Laurine, WYREBSKI Antoine, ZORDAN Cécile. We also thank the members of the Scientific Committee of the DEFIDIAG project: BECKMANN Jacques, FITZPATRICK David, MATTHIJS Gert, RAYMOND Lucy, RIESS Olaf, ABRAHAMOWICZ Michal, BUCHANAN James.

REFERENCES

Adithyan, G. S., Sivakami, M., and Jacob, J. (2017). Positive and Negative Impacts on Caregivers of Children with Intellectual Disability in India. *Dcid* 28 (2), 74–94. doi:10.5463/dcid.v28i2.595

Aviesan (2025). France Médecine Génomique. . https://aviesan.fr/ (Accessed November 8, 2021).

Binquet, C., Lejeune, C., Faivre, L., Bouctot, M., Asensio, M. A., Simon, A., et al. (2022). Genome Sequencing for Genetics Diagnosis of Patients with Intellectual Disability: The DEFIDIAG Study. *Front. Genet* 12, 1–13. doi:10.3389/fgene.2021.766964

Briggs, A. H., and Gray, A. M. (1999). Methods in Health Service Research: Handling Uncertainty in Economic Evaluations of Healthcare Interventions. *BMJ* 319 (7210), 635–638. doi:10.1136/bmj.319.7210.635

Buntinx, W., Cans, C., Colleaux, L., Courbois, Y., Debbané, M., Desportes, et al. (2016). *Déficiences Intellectuelles. Expertise Collective*. Institut national de la santé et de la recherche médicale INSERM.

Chérot, E., Keren, B., Dubourg, C., Carré, W., Fradin, M., Lavillaureix, A., et al. (2018). Using Medical Exome Sequencing to Identify the Causes of Neurodevelopmental Disorders: Experience of 2 Clinical Units and 216 Patients. *Clin. Genet.* 93, 567–576. doi:10.1111/cge.13102

Collins, F. S., Morgan, M., and Patrinos, A. (2003). The Human Genome Project: Lessons from Large-Scale Biology. *Science* 300 (5617), 286–290. doi:10.1126/science.1084564

Foster, M. W., Mulvihill, J. J., and Sharp, R. R. (2009). Evaluating the Utility of Personal Genomic Information. *Genet. Med.* 11 (8), 570–574. doi:10.1097/GIM.0b013e3181a2743e

Frick, K. D. (2009). Microcosting Quantity Data Collection Methods. *Med. Care* 47 (1), S76–S81. doi:10.1097/MLR.0b013e31819bc064

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., et al. (2014). Genome Sequencing Identifies Major Causes of Severe Intellectual Disability. *Nature* 511, 344–347. doi:10.1038/nature13394

Hennink, M. M., Kaiser, B. N., and Marconi, V. C. (2017). , 27, 591–608. doi:10.1177/1049732316665344Code Saturation versus Meaning Saturation: How Many Interviews Are Enough?*Qual. Health Res.*4

Jegathisawaran, J., Tsiplova, K., Hayeems, R., and Ungar, W. J. (2020). Determining Accurate Costs for Genomic Sequencing Technologies-A Necessary Prerequisite. *J. Community Genet.* 11 (2), 235–238. doi:10.1007/s12687-019-00442-7

Kohler, J. N., Turbitt, E., Lewis, K. L., Wilfond, B. S., Jamal, L., Peay, H. L., et al. (2017). Defining Personal Utility in Genomics: A Delphi Study. *Clin. Genet.* 92 (3), 290–297. doi:10.1111/cge.12998

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 (4), 435–443. doi:10.1038/gim.2017.119

Mollison, L., O'Daniel, J. M., Henderson, G. E., Berg, J. S., and Skinner, D. (2020). Parents' Perceptions of Personal Utility of Exome Sequencing Results. *Genet. Med.* 22 (4), 752–757. doi:10.1038/s41436-019-0730-8

Monroe, G. R., Frederix, G. W., Savelberg, S. M. C., de Vries, T. I., Duran, K. J., van der Smagt, J. J., et al. (2016). Effectiveness of Whole-Exome Sequencing and Costs of the Traditional Diagnostic Trajectory in Children with Intellectual Disability. *Genet. Med.* 18 (9), 949–956. doi:10.1038/gim.2015.200

Ontario Health (Quality) (2020). Genome-Wide Sequencing for Unexplained Developmental Disabilities or Multiple Congenital Anomalies: A Health Technology Assessment. *Ont Health Technol. Assess. Ser.* 20 (11), 1–178.

Peplow, M. (2016). The 100 000 Genomes Project. *BMJ* 353, i1757. doi:10.1136/bmj.i1757

Redin, C., Gérard, B., Lauer, J., Herenger, Y., Muller, J., Quartier, A., et al. (2014). Efficient Strategy for the Molecular Diagnosis of Intellectual Disability Using Targeted High-Throughput Sequencing. *J. Med. Genet.* 51 (11), 724–736. doi:10.1136/jmedgenet-2014-102554

Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., et al. (2020). The Complete Costs of Genome Sequencing: a Microcosting Study in Cancer and Rare Diseases from a Single center in the United Kingdom. *Genet. Med.* 22 (1), 85–94. doi:10.1038/s41436-019-0618-7

Stark, Z., Schofield, D., Alam, K., Wilson, W., Mupfeki, N., Macciocca, I., et al. (2017). Prospective Comparison of the Cost-Effectiveness of Clinical Whole-Exome Sequencing with that of Usual Care Overwhelmingly Supports Early Use and Reimbursement. *Genet. Med.* 19 (8), 867–874. doi:10.1038/gim.2016.221

Thomas, D. R., and Thomas, A. (2006). A General Inductive Approach for Analyzing Qualitative Evaluation Data. *Am. J. Eval.* 27 (2), 237–246. doi:10.1177/1098214005283748

Tsiplova, K., Zur, R. M., Marshall, C. R., Stavropoulos, D. J., Pereira, S. L., Merico, D., et al. (2017). A Microcosting and Cost-Consequence Analysis of Clinical Genomic Testing Strategies in Autism Spectrum Disorder. *Genet. Med.* 19 (11), 1268–1275. doi:10.1038/gim.2017.47

Van Nimwegen, K. (2017). "Health Technology Assessment of Next-Generation Sequencing," ([NijmegenThe Netherlands]: Radboud University). [thesis]. doi:10.13140/RG.2.2.19492.99201

Vissers, L. E. L. M., van Nimwegen, K. J. M., Schieving, J. H., Kamsteeg, E.-J., Kleefstra, T., Yntema, H. G., et al. (2017). A Clinical Utility Study of Exome Sequencing versus Conventional Genetic Testing in Pediatric Neurology. *Genet. Med.* 19 (9), 1055–1063. doi:10.1038/gim.2017.1

Vrijenhoek, T., Middelburg, E. M., Monroe, G. R., van Gassen, K. L. I., Geenen, J. W., Hövels, A. M., et al. (2018). Whole-exome Sequencing in Intellectual Disability; Cost before and after a Diagnosis. *Eur. J. Hum. Genet.* 26 (11), 1566–1571. doi:10.1038/s41431-018-0203-6

York Health Economics Consortium (2016). Net Monetary Benefit. https://yhec.co.uk/glossary/net-monetary-benefit/ (Accessed November 8, 2021).

Yuen, T., Carter, M. T., Szatmari, P., and Ungar, W. J. (2018). Cost-effectiveness of Genome and Exome Sequencing in Children Diagnosed with Autism Spectrum Disorder. *Appl. Health Econ. Health Pol.* 16 (4), 481–493. doi:10.1007/s40258-018-0390-x

Zhao, W., Wang, J., He, X., Huang, X., Jiao, Y., Dai, M., et al. (2004). BGI-RIS: an Integrated Information Resource and Comparative Analysis Workbench for rice Genomics. *Nucleic Acids Res.* 32, 377D–382D. doi:10.1093/nar/gkh085

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.

frontiers | Frontiers in Genetics

ORIGINAL RESEARCH
published: 20 April 2022
doi: 10.3389/fgene.2022.859140

Novel Heterozygous Missense Variant in GRIA4 Gene Associated With Neurodevelopmental Disorder With or Without Seizures and Gait Abnormalities

Hua Wang[1], Jiatong Liu[1], Fuwei Li[2], Ziteng Teng[1], Mingyu Liu[2] and Weiyue Gu[2]*

[1]Department of Pediatric Neurology, Shengjing Hospital of China Medical University, Shenyang, China, [2]Chigene (Beijing) Translational Medical Research Center Co., Ltd., Beijing, China

OPEN ACCESS

Edited by:
Tianyun Wang,
University of Washington,
United States

Reviewed by:
Yuki Hitomi,
Hoshi University, Japan
Stefano Castellana,
Home for Relief of Suffering (IRCCS),
Italy

****Correspondence:***
Hua Wang
shengjingwangh1@163.com

Specialty section:
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

Received: *21 January 2022*
Accepted: *21 March 2022*
Published: *20 April 2022*

Citation:
Wang H, Liu J, Li F, Teng Z, Liu M and Gu W (2022) Novel Heterozygous Missense Variant in GRIA4 Gene Associated With Neurodevelopmental Disorder With or Without Seizures and Gait Abnormalities. Front. Genet. 13:859140. doi: 10.3389/fgene.2022.859140

Objective: Neurodevelopmental disorder with or without seizure and gait abnormalities (NEDSGA, MIM * 617864) is a newly described autosomal dominant inherited disease caused by a heterozygous variant in the GRIA4 gene. *GRIA4* plays an essential role in excitatory synaptic transmission. In this study, we presented the clinical and genetic features of a female patient carrying a novel *de novo* variant in *GRIA4* and further reviewed the previously reported five different patients.

Methods: Evaluation of the patient included a detailed history and clinical examination. Trio-whole exome sequencing (WES) was performed to identify pathogenic variants in NEDSGA. Sanger sequencing was further used to validate the variants.

Results: We described the clinical features of an infant diagnosed with NEDSGA caused by a *GRIA4* variant, who presented with severe developmental delay, limb hypertonia, generalized seizure, retinal hypoplasia, and chorioretinal hyperpigmentation. The patient developed tricuspid regurgitation, and imaging examination revealed a patent foramen ovale. Trio-WES identified a novel *de novo* heterozygous missense variant c.1918G>T, p.Ala640Ser in the GRIA4 gene. Multiple in silico tools predicted deleterious effects of p.Ala640Ser.

Conclusion: A novel heterozygous missense variant in the GRIA4 gene (c.1918G>T) identified in the proband expanded the genotypic and phenotypic spectrum of disorders associated with *GRIA4* variants. This is the first NEDSGA case reported in China. Our findings provide valuable information for the differential diagnosis of neonatal onset neurodevelopmental disorders.

Keywords: GRIA4, NEDSGA, neurodevelopmental disorder, trio-whole exome sequencing, novel heterozygous missense variant

INTRODUCTION

Neurodevelopmental disorder with or without seizure and gait abnormalities (NEDSGA, MIM * 617864) is an early onset of neurodevelopmental disorder associated with global developmental delay and variable intellectual disability. Most patients presented with irritability, stiffness, seizure, and hypertonia early in life, followed by spasticity and impaired gait. NEDSGA is caused by the gene *GRIA4* (MIM * 138246) located on chromosomes 11q22 and is inherited in an autosomal dominant manner. GRIA4 encodes glutamate ionotropic receptor AMPA type subunit 4 and plays an essential role in excitatory synaptic transmission. *GRIA4* is ubiquitous in the central nervous system and is highly expressed in the thalamus, especially in the thalamic reticular nucleus (Beyer et al., 2008). GRIA4 in the rat brain is relatively high in CA1 pyramidal cells, hippocampal dentate gyrus, cerebral cortex, and cerebellar granule cells (Keinänen et al., 1990). Since Martin first reported NEDSGA in 2017 (Martin et al., 2017), only five NEDSGA patients with *GRIA4* variants have been reported, ranging in age from 4 to 21 years. All *GRIA4* variants described are *de novo* heterozygous missense variants that cluster in the transmembrane and ligand-binding domains of the GRIA4 protein.

In this study, we reported a 9-month-old girl from a nonconsanguineous family with healthy parents. The patient presented with severe global developmental delay (HP: 0011344), limb hypertonia (HP:0002509), partial seizure (HP:0007359), retinal hypoplasia (HP:0007770), chorioretinal hyperpigmentation (HP:0040031), tricuspid regurgitation (HP:0005180), and patent foramen ovale (HP: 0001655). Using trio-whole exome sequencing (WES), we identified a novel *de novo* heterozygous missense variant in *GRIA4*, c.1918G>T, p.Ala640Ser. To the best of our knowledge, this is the first NEDSGA case reported in China. Our finding expands the genotypic and phenotypic spectrum associated with *GRIA4*.

PATIENTS AND METHODS

Patients

The Ethics Committee of the Shengjing Hospital of China Medical University approved this study. The patient's legal guardians signed the informed consent for the study.

The proband was a 9-month-old girl, the first child of healthy Chinese parents. The mother was 34 years old, and the father was 40 years old. The patient was transferred to our hospital due to the occurrence of breathing difficulties, limb hypertonia, and seizure. Electroencephalogram (EEG) and brain magnetic resonance imaging (MRI) were performed on her clinical presentation.

Variation Analysis

DNA was obtained from the peripheral venous blood of the girl and the parents and submitted to the Chigene Translational Medicine Research Center Co., Ltd., Beijing, for trio (parents and proband)-WES. Whole-exome capture was xGen Exome Research Panel v2.0 (IDT, Iowa, United States). The sequencing operation flow was standardized on the DNBSEQ-T7 (BGI, China) platform. Raw-sequencing reads were processed by fastp (https://github.com/OpenGene/fastp) for adapter removal and low-quality read filtering. High-quality sequencing data were generated and performed on the Ensemble GRCh37/hg19 reference genome using the Burrows-Wheeler Aligner (BWA, https://github.com/lh3/bwa). GATK (http://www.broadinstitute.org/gatk/) was used for base quality score recalibration and SNP and INDEL calling. Trio-WES had a mean depth of coverage of at least 172× per sample, with 98% of the exome covered 20× or greater. The sequencing depth ranged from 192×–413× coverage of the GRIA4 gene, with 100% target region coverage >10× sequencing depth. Pathogenicity of the genetic variants were predicted by bioinformatics tools such as PolyPhen (http://www.bork.embl-heidelberg.de/PolyPhen/), Mutation Taster (http://www.mutationtaster.org), REVEL (https://sites.google.com/site/revelgenomics/), and CADD (http://cadd.gs.washington.edu/). Finally, the variants were classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards et al., 2015).

Pathogenic variants were detected using trio-WES, followed by Sanger sequencing validation. Primer 5.0 primer software was used to design the *GRIA4* primers: forward primer (5′- GCTAAAGCCCATGGTATAATTGTT G-3′) and reverse primer (5′-GCATGGTGAATTGACGGT ATTTCTT-3′). Sanger sequencing was further performed using the 3730xl DNA Analyzer (Applied Biosystems, United States). The identified rare variants have been submitted to the ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/) (Accession number: SCV002055968).

The GRIA4 protein is highly conserved among vertebrates, showing 86% sequence identity among 10 different species, including primates, rodents, laurasiatheria, placental mammals, sauropsida, and fish (http://asia.ensembl.org/index.html). Multiple protein sequence alignments were conducted on MEGA X (Kumar et al., 2018). 3D modeling of structural effects was performed using the GRIA4 protein structure (AlphaFold, AF-P48058-F1) (Varadi et al., 2022). The models were visualized using Pymol (www.pymol.org) (Schrodinger, 2015).

RESULT

Clinical Features

The girl was born by an uncomplicated delivery at 38 weeks' gestation, with a normal birth weight of 3,200 g and a body length of 50 cm. Neonatally she presented with irritability, breathing difficulties, limb hypertonia, and partial seizure. Cardiac ultrasonography showed tricuspid regurgitation and patent foramen ovale. Fundus examinations revealed hypoplasia of the retina and chorioretinal hyperpigmentation. Brain magnetic resonance imaging (MRI) was normal. Her electroencephalogram (EEG) demonstrated multifocal sharp waves and low waves during

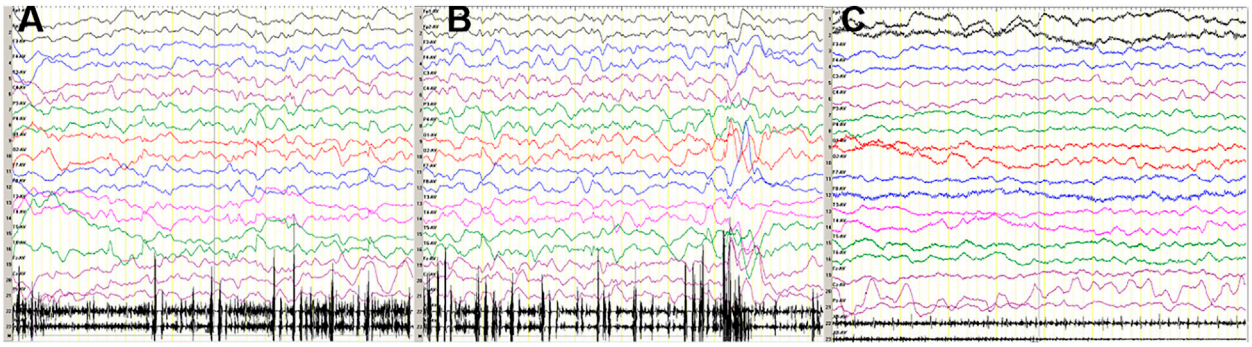


FIGURE 1 | (A, B) At the age of 3 months, sleep EEG of the patient revealed multifocal sharp waves and low waves. **(C)** Wake EEG was normal.

sleep (**Figures 1A–C**). Laboratory tests showed hyperlactemia and hyperammonemia.

In the follow-up, she suffered from severe developmental delay. At 4 months, she started taking lysine, inosite and vitamin B12 oral solution, GABA compound nutritious solid drink, and cerebroprotein hydrolysate oral solution. Limb hypertonia persisted, but seizures were well controlled.

To the best of our knowledge, only one publication (**Table1**) (Martin et al., 2017) reported a total of five cases with *GRIA4* variants. Our patient is the sixth case of NEDSGA, in which neurodevelopmental disorders with seizures and abnormal gait were the most common phenotypes caused by pathogenic variants in *GRIA4*.

All six patients (2 females and 4 males, aged 9 months to 21 years) had neurodevelopmental disorders, with mild (2/6; 33%) to severe (4/6; 67%) developmental delay. Four patients (4/6; 67%) had movement disorders, including clumsy or stiff gait and inability to walk. Three patients (3/6; 50%) presented with ocular anomalies, including strabismus, nystagmus, optic nerve hypoplasia, hypoplasia of the retina, and chorioretinal hyperpigmentation. Five patients (5/6, 83%) had poor speech or aphasia. Seizures occurred in four of six patients (67%). Patient 2 had sudden muscle cramps/seizures lasting up to 1 hour after trauma, while patient 6 had no seizures. The age of seizure onset ranged from 1 day to 14 months, with a median age of 7 months. Of note, seizure disorders were found in all patients with severe developmental delay. A broad spectrum, including generalized seizures, febrile seizures, and seizure-like episodes, was reported. Three patients (3/4; 75%) achieved seizure control by therapies. However, patient 3 developed refractory seizures and status epilepticus. All four patients with seizures exhibited abnormal EEG and two of them had abnormal MRI.

Our patient presented with severe developmental delay, limb hypertonia, partial seizure, retinal hypoplasia, chorioretinal hyperpigmentation, tricuspid regurgitation, and patent foramen ovale, whereas the girl had no signs of craniofacial or MRI abnormalities. In addition, our patient was too young to assess her ability to walk or speak. Together with our clinical findings, half of the affected patients had ocular abnormalities, so the patient's ocular examination should not be ignored.

Variation Analysis

Trio-WES identified a *de novo* heterozygous variant in *GRIA4* in the patient: chr11:105797537G>T(hg19), c.1918G>T transition. The c.1918G>T is predicted to result in the substitution of the alanine residue p.Ala640Ser. This variant has not been reported in public databases (gnomAD v2.1.1, http://gnomad.broadinstitute.org/ and the 1000 Genomes Project, http://www.internationalgenome.org). In addition, it was predicted as pathogenic by multiple bioinformatic tools (SIFT: damaging; PolyPhen: probably damaging; Mutation Taster: disease_causing; REVEL: deleterious; CADD: 26.3). According to the ACMG guidelines, we confirmed the variant to be pathogenic (PS2+PM1+PM2+PP2+PP3) (Richards et al., 2015; Harrison et al., 2019). Furthermore, we did not detect pathogenic or likely pathogenic variants in genes known to be associated with neurodevelopmental disorder in probands using Trio-WES. The variants of the *GRIA4* gene were confirmed by Sanger sequencing (**Figure 2A**).

Sequence alignment among multiple vertebrate species suggested that p.Ala640Ser was located at a highly conserved site (**Figure 2B**) and in the transmembrane domain (**Figure 2C**). 3D structural analysis of the *GRIA4* protein showed that the mutated Ser640 residue formed a new hydrogen bond between Ser640 and the neighboring Ser636 compared to the WT model (**Figures 3A–C**). The mutation of Ala640Ser, located in an alpha-helix, causes the change of hydrogen bond between residues, and further affects protein folding.

DISCUSSION

NEDSGA is a newly recognized rare neurodevelopmental disorder caused by a heterozygous variant in the *GRIA4* gene. All the reported *GRIA4* variants were heterozygous missense, located in the transmembrane and ligand-binding domains (**Figure 2C**) (Martin et al., 2017).

TABLE 1 | Clinical features of individuals with *de novo GRIA4* variants.

Patient	1 (This study)	2 (Martin et al., 2017)	3 (Martin et al., 2017)	4 (Martin et al., 2017)	5 (Martin et al., 2017)	6 (Martin et al., 2017)	Total
Variant	c.1918G>T,p.Ala640Ser	c.1915A>T,p.Thr639Ser	c.1921A>G,p.Asn641Asp	c.1928C>G,p.Ala643Gly	c.1931C>T,p.Ala644Val	c.2090G>C,p.Arg697Pro	—
Provean	Deleterious (-2.68)	Deleterious (–3.57)	Deleterious (–4.47)	Deleterious (–3.58)	Deleterious (–3.58)	Deleterious (–2.69)	—
SIFT	Damaging (0.014)	Damaging (0.001)	Damaging (0.001)	Damaging (0.001)	Damaging (0.0)	Damaging (0.005)	—
PolyPhen	Probably damaging (1.0)	Probably damaging (1.0)	Probably damaging (1.0)	Probably damaging (1.0)	Probably damaging (1.0)	Benign (0.026)	—
Mutation Taster	Disease_causing (1)	Disease_causing (1)	Disease_causing (1)	Disease_causing (1)	Disease_causing (1)	Disease_causing (0.985836)	—
CADD	Deleterious (26.3)	Deleterious (24.7)	Deleterious (26.3)	Deleterious (29.2)	Deleterious (29.8)	Deleterious (24.1)	—
Gender	Female	Male	Male	Male	Male	Female	—
Age at last exam	9 months	15 years	21 years	4 years	4 years	4 years	—
Seizures	Partial seizures	No	Intractable generalized seizures	Seizure-like episodes	Febrile seizures	No	—
Age at epilepsy onset	1 day	—	5 weeks	14 months	13 months	—	—
Seizure outcome	Seizure control	—	Refractory	Seizure control	Seizure control	—	—
EEG	Multifocal sharp waves, low waves during sleep	Unremarkable	Diffuse cerebral disturbance without electrographic correlates to the seizures	Generalized slowing, no epileptiform discharges	Generalized spikes and waves during sleep	Unremarkable	4/6
Brain MRI	Unremarkable	Unremarkable	Bilateral symmetric extensive atrophy of frontal lobes, mild frontal ventriculomegaly, thin corpus callosum	Optic nerve hypoplasia	Unremarkable	Unremarkable	2/6
Muscle	Hypertonia	Hyperekplexia with exaggerated head-retraction reflex, stiffness, and hypertonia	Severe spastic quadriplegia and hypertonia with contractures	Spasticity	Mild muscular hypotonia (neonatal)	Unremarkable	5/6
Developmental delay	Moderate to severe	Mild to moderate	Severe	Severe	Moderate to severe	Mild to moderate	6/6
Motor development	Too young to evaluate	Clumsy or stiff gait	Inability to walk	Clumsy or stiff gait	Clumsy or stiff gait	Normal	4/5
Speech impairment	Too young to evaluate	Speech with dysarthria	Absent speech	Absent speech	Absent speech	Poor speech	5/5
Eye features	Hypoplasia of the retina, chorioretinal hyperpigmentation	NA	Strabismus	Nystagmus, optic nerve hypoplasia	NA	NA	3/3
Heart	Tricuspid regurgitation, patent foramen ovale	NA	NA	NA	NA	NA	1/1
Additional features	Hyperlactemia, hyperammonemia	Sleeping problems, dysmorphic features	Feeding difficulties, apneas, choreiform movements, dysmorphic features	Dysmorphic features	Stereotypic hand movements	Hyporeflexia, simian crease on both hands	—

EEG, electroencephalograph; MRI, magnetic resonance imaging; NA, not available.

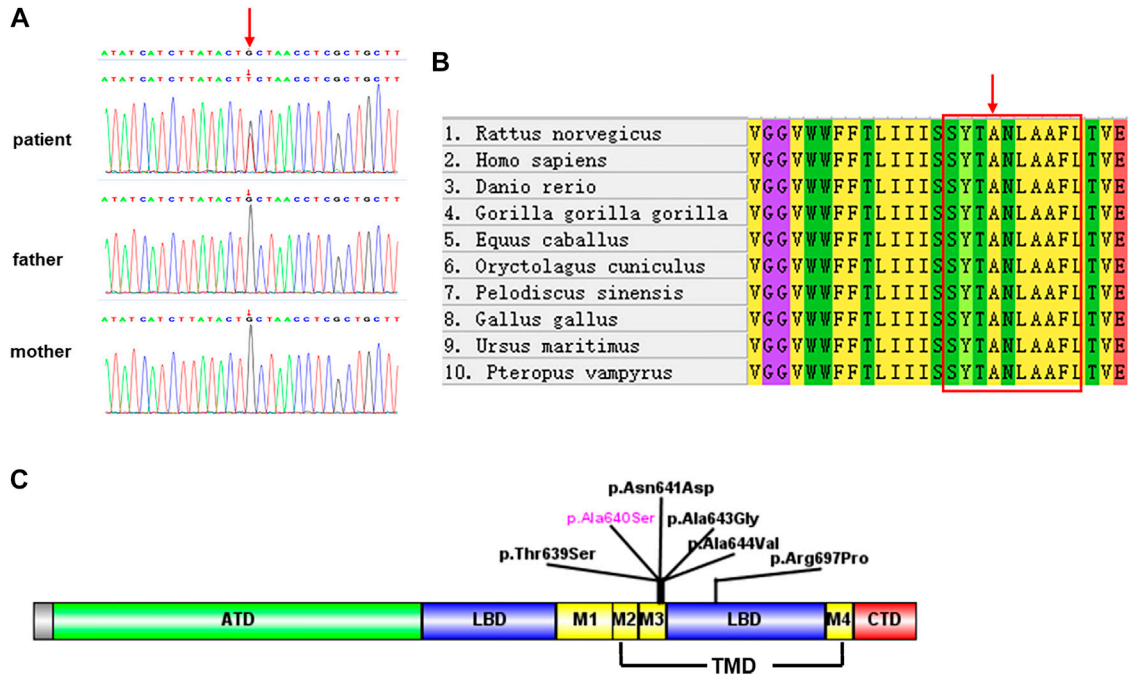


FIGURE 2 | (A) Sanger sequencing validation of the variant c.1918G>T in the proband and parents. **(B)** Multiple species sequence alignment. The mutated alanine A640 (red arrow) falls within a highly conserved SYTANLAAF motif (red box). **(C)** Schematics depicting the location of GRIA4 variants. Amino-terminal domain (ATD, in green), ligand binding domain (LBD, in blue), transmembrane domain (TMD, in yellow), and carboxyl-terminal domain (CTD, in red). TMD contains three transmembrane domains (M1, M3, and M4) and re-entrant membrane loop M2.

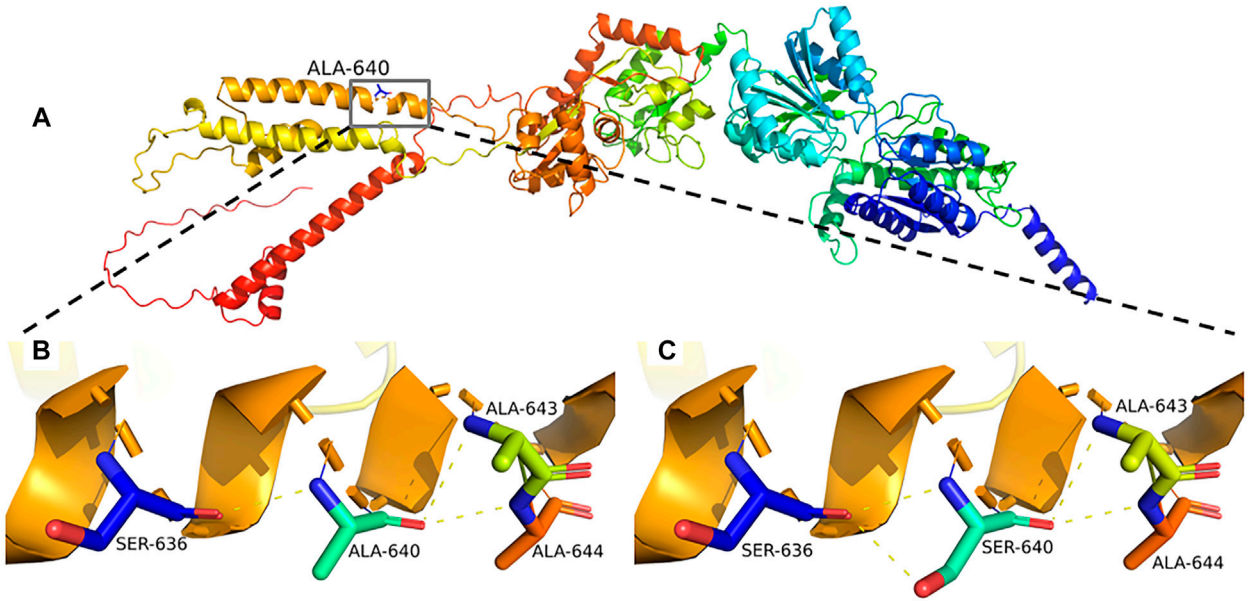


FIGURE 3 | (A–C) Predicted mutational impact of p.Ala640Ser on the GRIA4 protein structure. Compared to the WT model **(B)**, the new hydrogen bond formed between the mutant Ser640 **(C)** and the neighboring Ser636 hydrogen bonds, yellow dashed lines.

AMPARs (α-amino-3-hydroxy-5-methyl-4-isoxazolepropionate receptors) consist of four subunits GluR1-GluR4, mediate fast excitatory neurotransmission in the central nervous system, and play a critical role in learning, memory formation, and brain development. Each iGluR subunit comprises an amino-terminal domain (ATD), a ligand-binding domain (LBD), a transmembrane domain (TMD), and a carboxyl-terminal domain (CTD) (**Figure 2C**) (Traynelis et al., 2010; Sobolevsky, 2015). TMD contains three transmembrane domains (M1, M3, and M4) and re-entrant membrane loop M2 (Twomey et al., 2017). M3 transmembrane segment contains a nine amino acid sequence, SYTANLAAF motif, that plays a crucial role in channel activation and gating, and is highly sensitive to a conservative amino acid change (Salpietro et al., 2019). The variant p.Ala640Ser is located in the SYTANLAAF motif that is highly conserved throughout all members of the glutamate receptor family (Jones et al., 2002). Furthermore, multiple sequence alignment analysis revealed that p.Ala640 is highly conserved in different species (**Figure 2B**). For the variant p.Ala640Ser, the presence of the serine acid residue is predicted to lead to the formation of a new hydrogen bond with the p.Ala636 residue. p.Ala640Ser is in hydrophobic membrane-spanning helix M3, mutation of a critical hydrophobic residue (Ala) to a hydrophilic one (Ser) is predicted to destroy the hydrophobic interactions and further disrupt the gating mechanism (Sobolevsky et al., 2003; Sobolevsky, 2015). In the previous study, Salpietro et al. identified the variant p.Ala639Ser in *GRIA2* in two unrelated infants with uncontrolled seizures from the first days of life (Salpietro et al.). And the variant p.Ala639Ser in *GRIA2* is located in the same position (in the motif of SYTANLAAF) as the variant p.Ala640Ser in *GRIA4* in our patient. An *in vitro* functional study showed that the variant p.Ala639Ser in *GRIA2* reduced agonist-induced current amplitude, resulting in a significantly reduced cell-surface expression of *GRIA2* (Salpietro et al.). Therefore, we speculated that the variant p.Ala640Ser in GRIA4 had a high probability of pathogenicity, as it was also located in the same important functional domain. Moreover, further investigation by the electrophysiology experiments and biotinylation assay will be necessary to determine whether p.Ala640Ser in GRIA4 affects the channel synthesis or trafficking and its effect on currents.

In conclusion, using trio-WES, we identified a novel *de novo* heterozygous missense variant in the *GRIA4* gene and diagnosed the sixth NEDSGA patient with severe developmental delay, limb hypertonia, partial seizure, hypoplasia of the retina, chorioretinal hyperpigmentation, and other clinical characteristics. Our findings enriched the phenotypic spectrum of genetic disorders associated with *GRIA4* variants. And genetic evidence further supports the association of rare and newly reported NEDSGA caused by the *GRIA4* gene. Our data will be helpful in diagnosing NEDSGA, especially in the affected newborns.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/, SCV002055968.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Shengjing Hospital of China 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

HW drafted the initial manuscript. FL, WG, and ML were responsible for genetic analysis. HW, JL, and ZT collected clinical information. HW designed the study, and critically reviewed the manuscript. All the authors read and approved the final manuscript.

REFERENCES

Beyer, B., Deleuze, C., Letts, V. A., Mahaffey, C. L., Boumil, R. M., Lew, T. A., et al. (2008). Absence Seizures in C3H/HeJ and Knockout Mice Caused by Mutation of the AMPA Receptor Subunit Gria4. *Hum. Mol. Genet.* 17 (12), 1738–1749. doi:10.1093/hmg/ddn064

Harrison, S. M., Biesecker, L. G., and Rehm, H. L. (2019). Overview of Specifications to the ACMG/AMP Variant Interpretation Guidelines. *Curr. Protoc. Hum. Genet.* 103 (1), 93. doi:10.1002/cphg.93

Jones, K. S., VanDongen, H. M. A., and VanDongen, A. M. J. (2002). The NMDA Receptor M3 Segment Is a Conserved Transduction Element Coupling Ligand Binding to Channel Opening. *J. Neurosci.* 22 (6), 2044–2053. doi:10.1523/jneurosci.22-06-02044.2002

Keinänen, K., Wisden, W., Sommer, B., Werner, P., Herb, A., Verdoorn, T. A., et al. (1990). A Family of AMPA-Selective Glutamate Receptors. *Science* 249 (4968), 556–560. doi:10.1126/science.2166337

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* 35 (6), 1547–1549. doi:10.1093/molbev/msy096

Martin, S., Chamberlin, A., Shinde, D. N., Hempel, M., Strom, T. M., Schreiber, A., et al. (2017). De Novo Variants in GRIA4 Lead to Intellectual Disability with or without Seizures and Gait Abnormalities. *Am. J. Hum. Genet.* 101 (6), 1013–1020. doi:10.1016/j.ajhg.2017.11.004

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 (5), 405–424. doi:10.1038/gim.2015.30

Salpietro, V., Dixon, C. L., Guo, H., Bello, O. D., Vandrovcova, J., Efthymiou, S., et al. (2019). AMPA Receptor GluA2 Subunit Defects Are a Cause of Neurodevelopmental Disorders. *Nat. Commun.* 10 (1), 3094–10910. doi:10.1038/s41467-019-10910-w

Schrodinger, L. L. C. (2015). The AxPyMOL Molecular Graphics Plugin for Microsoft PowerPoint. Version 1.8.

Sobolevsky, A. I. (2015). Structure and Gating of Tetrameric Glutamate Receptors. *J. Physiol.* 593 (1), 29–38. doi:10.1113/jphysiol.2013.264911

Sobolevsky, A. I., Yelshansky, M. V., and Wollmuth, L. P. (2003). Different Gating Mechanisms in Glutamate Receptor and K+Channels. *J. Neurosci.* 23 (20), 7559–7568. doi:10.1523/jneurosci.23-20-07559.2003

Traynelis, S. F., Wollmuth, L. P., McBain, C. J., Menniti, F. S., Vance, K. M., Ogden, K. K., et al. (2010). Glutamate Receptor Ion Channels: Structure, Regulation, and Function. *Pharmacol. Rev.* 62 (3), 405–496. doi:10.1124/pr.109.002451

Twomey, E. C., Yelshanskaya, M. V., Grassucci, R. A., Frank, J., and Sobolevsky, A. I. (2017). Channel Opening and Gating Mechanism in AMPA-Subtype Glutamate Receptors. *Nature* 549 (7670), 60–65. doi:10.1038/nature23479

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., et al. (2022). AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res.* 50 (D1), D439–D444. doi:10.1093/nar/gkab1061

Conflict of Interest: FL and WG were employed by Chigene (Beijing) Translational Medical Research Center Co., Ltd.

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.

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.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 14 June 2022
doi: 10.3389/fnins.2022.892768

Splicing Interruption by Intron Variants in *CSNK2B* Causes Poirier–Bienvenu Neurodevelopmental Syndrome: A Focus on Genotype–Phenotype Correlations

Wen Zhang[1,2,3], *Fanghua Ye*[1], *Shimeng Chen*[1,2,3], *Jing Peng*[1,2,3], *Nan Pang*[1,2,3]* and *Fei Yin*[1,2,3]*

[1] Department of Pediatrics, Xiangya Hospital, Central South University, Changsha, China, [2] Hunan Intellectual and Developmental Disabilities Research Center, Changsha, China, [3] Clinical Research Center for Children Neurodevelopmental Disabilities of Hunan Province, Xiangya Hospital, Central South University, Changsha, China

OPEN ACCESS

Edited by:
Xiaoli Chen,
Capital Institute of Pediatrics, China

Reviewed by:
Shengnan Wu,
Shanghai Children's Hospital, China
Madelyn Gillentine,
Seattle Children's Hospital,
United States

***Correspondence:**
Nan Pang
nanpang@csu.edu.cn
Fei Yin
yf2323@hotmail.com

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

Received: 09 March 2022
Accepted: 18 May 2022
Published: 14 June 2022

Citation:
Zhang W, Ye F, Chen S, Peng J, Pang N and Yin F (2022) Splicing Interruption by Intron Variants in CSNK2B Causes Poirier–Bienvenu Neurodevelopmental Syndrome: A Focus on Genotype–Phenotype Correlations. Front. Neurosci. 16:892768. doi: 10.3389/fnins.2022.892768

CSNK2B has recently been identified as the causative gene for Poirier–Bienvenu neurodevelopmental syndrome (POBINDS). POBINDS is a rare neurodevelopmental disorder characterized by early-onset epilepsy, developmental delay, hypotonia, and dysmorphism. Limited by the scarcity of patients, the genotype–phenotype correlations in POBINDS are still unclear. In the present study, we describe the clinical and genetic characteristics of eight individuals with POBINDS, most of whom suffered developmental delay, generalized epilepsy, and hypotonia. Minigene experiments confirmed that two intron variants (c.367+5G>A and c.367+6T>C) resulted in the skipping of exon 5, leading to a premature termination of mRNA transcription. Combining our data with the available literature, the types of POBINDS-causing variants included missense, nonsense, frameshift, and splicing, but the variant types do not reflect the clinical severity. Reduced casein kinase 2 holoenzyme activity may represent a unifying pathogenesis. We also found that individuals with missense variants in the zinc finger domain had manageable seizures (p = 0.009) and milder intellectual disability (p = 0.003) than those with missense variants in other domains of *CSNK2B*. This is the first study of genotype–phenotype correlations in POBINDS, drawing attention to the pathogenicity of intron variants and expanding the understanding of neurodevelopmental disorders.

Keywords: ***CSNK2B*, Poirier–Bienvenu neurodevelopmental syndrome, genotype, phenotype, intron variants**

INTRODUCTION

Poirier–Bienvenu neurodevelopmental syndrome [Online Mendelian Inheritance in Man (OMIM) #618732, POBINDS] is a recently described rare autosomal dominant neurodevelopmental disorder characterized by early-onset seizures, developmental delay, hypotonia, and dysmorphism (Ernst et al., 2021). POBINDS is caused by a *de novo* variant in the *CSNK2B* gene

(located at 6p21.33), which encodes the casein kinase 2 enzyme β subunit (CK2β). CK2 is a ubiquitous protein serine/threonine kinase, a heterotetrameric enzyme consisting of two catalytic CK2α or CK2α′ subunits and two regulatory β subunits (Niefind et al., 2001). CK2β is a highly conserved regulatory subunit, and excessive synthesis of CK2β will form dimers that disrupt the complete structure and function of CK2 (Graham and Litchfield, 2000; Niefind et al., 2001). In previous studies, CK2 was found to be widespread in the brain (Guerra et al., 1999), and conditional knockout of *CSNK2B* mice induced age-dependent reduction of grip strength and impaired neuromuscular transmission and synaptic gene expression (Eiber et al., 2019). Knockdown of *CSNK2B* in mouse embryonic neural stem cells (NSC) impaired cell differentiation and reduced dendritic length and branching (Yang et al., 2018). These studies suggest a pivotal role for *CSNK2B* in neurodevelopmental disorders. To date, more than 40 individuals with POBINDS have been reported in the literature (Li et al., 2019; Ernst et al., 2021; Yang et al., 2021), but the genotype–phenotype correlations of POBINDS are still unclear.

In this study, we describe the clinical and genetic characteristics of eight *de novo* individuals with POBINDS, most of whom suffered developmental delay, generalized epilepsy, and hypotonia. Two missense, two nonsense, one frameshift, and three intron variants were detected, with only c.58G>T being reported before (Ernst et al., 2021). Minigene experiments confirmed that two intron variants (c.367+5G>A and c.367+6T>C) resulted in skipping of exon 5, leading to premature termination of mRNA transcription. Combining our data with the available literature, missense, nonsense, frameshift, and splicing site variants were found to cause POBINDS. The variant types do not reflect the severity of POBINDS. Reduced CK2 holoenzyme activity may represent a unifying pathogenesis. Missense variants in the zinc finger domain result in manageable seizures (p = 0.009) and milder intellectual disability (ID) (p = 0.003) than missense in other domains of *CSNK2B*. This is the first study of the genotype–phenotype correlations of *CSNK2B* variants and expands the knowledge of POBINDS.

MATERIALS AND METHODS

Subject

In this study, a total of eight individuals contributed clinical information and genetic data. This study was reviewed and approved by the Ethics Committee of Xiangya Hospital of Central South University following the Helsinki Declaration. The parents or legal guardian of each individual provided informed consent.

Next-Generation Sequencing and Sanger Sequencing Verification

Trio-whole-exome sequencing was performed on each family in this study and the methods have been described in detail in previous studies (Peng et al., 2018). All variants reported were confirmed with Sanger sequencing, and the interpretation of variants was conducted under the American College of Medical Genetics guidelines (Richards et al., 2015).

Minigene Construction

Minigene plasmids covering wild-type "*CSNK2B*_WT (exon 4–6)" and mutants "*CSNK2B*_Mu (exon 4–6)" (Mu1: c.367+2T>C, Mu2: c.367+5G>A, and Mu3: c.367+6T>C) were constructed using the pcDNA3.1 vector, respectively. The constructed minigene plasmids were verified by Sanger sequencing.

Cell Culture and Transfection

HEK293T cells and Hela cells were obtained from the Kunming Cell Bank of the Chinese Academy of Sciences (Kunming, China), and they were cultured in Dulbecco's Modified Eagle Medium (Hyclone, United States) supplemented with 10% fetal bovine serum (Gibco, United States), 100 U/ml penicillin, and 100 μg/ml streptomycin (Gibco, United States). Cell culture was performed at 37°C in an incubator filled with 5% CO_2 and 95% air. The minigene plasmids were transfected into HEK293T cells or Hela cells by LipoMax (Sudgen, China) according to the DNA transfection protocol. The cells were analyzed 48 h after transfection.

Reverse Transcription-Polymerase Chain Reaction and Sequence Analysis

RNA extraction methods have been described in detail in previous studies (Kong et al., 2015). Total RNA was reverse transcribed using Hifair^{TM} 1st Strand cDNA Synthesis SuperMix for PCR (YEASEN, China). PCR was run on the T100 PCR system (Bio-Rad, United States) using PrimerStarMAX DNA Polymerase (TaKaRa Bio, Japan) according to the manufacturer's protocol. The primers were synthesized by Tsingke Company (Target exon 4 – exon 6: 5′-CTAGAGAACCCACTGCTTAC, 3′-TAGAAGGCACAGTCGAGG). The purified PCR products were separated and purified *via* 2% agarose and sequenced as described above.

Systematic Literature Search of Reported *CSNK2B* Variants

We reviewed 57 POBINDS individuals combined with our cases and reported cases (PUBMED, Human Gene Mutation Database, ClinVar, and DECIPHER) involving *CSNK2B*. A total of 82 POBINDS-related variants are summarized, and variants without any detailed clinical information were only aggregated in the information summary (**Supplementary Table 1**, Patient 58 to Patient 82). Chromosomal microdeletion/microduplications were not involved. The annotation of predicted variants effect was carried by combined annotation dependent depletion (CADD).[1] Based on the way of affecting protein function, we classified the variants into two categories: loss-of-function (LOF, including start loss, frameshift insertion/deletion, nonsense, and splicing variant), non-LOF (including in-frame insertion/deletion and missense). In the non-LOF group, we also analyzed genotype–phenotype associations based on protein domains. MetaDome[2] was used for bioinformatics mutation tolerance analysis (Wiel et al., 2019).

[1]https://cadd.gs.washington.edu/snv

[2]https://stuart.radboudumc.nl/metadome/dashboard

Statistical Analyses

Groups were compared using the Chi-square test, Fisher's exact test was used where 20% or more of the gird in the Chi-square table were expected to count <5. Significant findings ($p < 0.05$) were calculated as the dominance ratio with 95% confidence intervals. Statistical analysis was performed using SPSS (version 18.0, United States) software.

RESULTS

Identification of *de novo CSNK2B* Variants

Eight *CSNK2B* (NM_001320.6) variants were detected in eight individuals; all variants were *de novo* and absent in the gnomAD and ClinVar databases. Only one variant (c.58G>T) was reported before (Ernst et al., 2021). Variant types included nonsense (c.58G>T and c.142C>T), frameshift (c.462_465del), missense (c.325T>C and c.497T>G), and intron (c.367+5G>A, c.367+6T>C, and c.292−1G>A) variants; variant information shown in **Supplementary Table 1**. Sanger sequencing data of these variants in each family is shown in **Supplementary Figure 1**.

Clinical Features of the Eight Individuals

Phenotypes are summarized in **Table 1**. In the present study, three individuals were male and five were female. All individuals (8/8) had varying degrees of seizures as first signs. Five individuals (5/8) had tonic–clonic seizures, and abnormal electroencephalograms (EEGs) were reported in all individuals (8/8). One individual's (1/8) MRI showed scattered abnormal signals in the parenchyma. Aside from individual 1, the other seven individuals (7/8) had abnormal neurodevelopmental phenotypes ranging from mild to profound ID. Three individuals (3/8) had mild dysmorphism, which manifested as partial restricted abduction and uplift of eyes, bilateral cryptorchidism, small penis, and a small jaw and prominent forehead, respectively. All individuals (8/8) had epilepsy or global developmental delay starting in the first year of life. Six individuals (6/8) had received one anti-epileptic drug (AED), and two individuals (2/8) received two AEDs. Seizures were completely controlled in seven individuals (7/8), and only one individual (1/8) had medical refraction. The AEDs most frequently reported to be helpful were levetiracetam (4/8), valproic acid (2/8), and oxcarbazepine (2/8). During the follow-up, individuals with global developmental delay (7/8) showed satisfactory improvement in motor, language, and cognitive symptoms but were still distinguishable from their healthy peers.

Minigene Splicing Assay for Intron Variants (c.367+5G>A and c.367+6T>C)

To verify that c.367+5G>A and c.367+6T>C alter the splicing of *CSNK2B*, we performed a minigene assay. We constructed the minigene of *CSNK2B*_WT and *CSNK2B*_Mu (Mu1: c.367+2T>C, Mu2: c.367+5G>A, and Mu3: c.367+6T>C), consisting of exons 4–6 of *CSNK2B* expressed in the pcDNA3.1 vector (**Figure 1A**). *CSNK2B*_Mu1 was used as a positive control for the minigene assay, which has been confirmed to skip exon 5 in primary cultured fibroblasts (Poirier et al., 2017). Next, we transfected the *CSNK2B* gene into HEK293T and Hela cells, respectively, and reverse transcription-polymerase chain reaction (RT-PCR) results detected a single band of predicted size (**Figure 1B**). After sequencing, we found that all three splicing variants resulted in skipping of exon 5 (**Figures 1C,D**). This skipping leads to the insertion of a premature termination codon at p.Leu98Alafs*11, confirmed by RT-PCR (Poirier et al., 2017). These data suggested that these variants resulted in the loss of function of *CSNK2B*.

Genotype–Phenotype Association Analysis for Reported Cases

In the combined analysis of the available literature, we summarized the details of genotypes and phenotypes from 57 cases. Among 57 cases, 48 variants were reported, including 16 missense variants (p.Asp32Asn and p.His165Arg were found in 3 patients and p.Cys137Phe and p.Arg111Pro in 2 patients), 11 splicing variants (including c.558-3T>C, c.367+5G>A, and c.367+6T>C, all confirmed alter splicing), 9 frameshift variants, 8 nonsense variants (c.58G>T, c.139C>T, and c.303C>A were found in two individuals, respectively), 3 start loss variants, and 1 in-frame deletion variant. Phenotypes of *CSNK2B* clinical as **Table 2**, individuals with the same variant showed varying degrees of phenotypic heterogeneity. A total of 82 POBINDS-related variants are summarized (**Figures 2A,B**), 25 of them without detailed clinical information are excluded in the phenotype–genotype linkage analysis (details in **Supplementary Table 1**, Patients 58 to Patient 82). Based on the way they affected protein function, we classified the 57 variants with detailed phenotypes into LOF or non-LOF groups. There was no statistically significant difference in the severity of the POBINDS (epilepsy, refractory epilepsy, moderate to profound ID, dysmorphism, and hypotonia) among mutations of LOF compared to non-LOF, except dysmorphisms ($p = 0.041$) (**Supplementary Table 2**). Among 23 missense cases, the proportion of refractory epilepsy (Fisher $p = 0.009$) and the severity of ID (Fisher $p = 0.003$) were significantly lower when the mutations were located in the zinc finger region compared to other regions, respectively (**Supplementary Table 3**). Similar correlations were not found between other domains and phenotypes. This matched the relative tolerance calculated by the "MetaDome" algorithm (**Figure 2B**). These results suggest that individuals with zinc finger region variants have controllable seizures and a milder degree of ID.

DISCUSSION

CSNK2B was first reported by Poirier et al. (2017) in two individuals with early-onset epilepsy, developmental delay, hypotonia, and malformations, and was documented as the causative gene for POBINDS by OMIM in 2020.[3] A total of

[3]https://omim.org

TABLE 1 | Genotype and phenotype details for individuals with *CSNK2B* variants.

Patient	Sex/age	Variant, NM_001320.6 (variant type)	First signs/symptoms	Seizure onset/types	AEDs/efficacy	ID	Dysmorphism	Hypotonia	Behavior features	Other features
1	F/3.5 years	c.367+5G>A, (SS)	10 months/seizure	10 months/complex febrile seizure	VPA/seizure free	Normal	Eyes cannot move outwards and upward	–	–	Abnormal EEG
2	F/21 months	c.367+6T>C, (SS)	3 months/seizure	3 months/focal onset, GTC	LEV/seizure free	Mild	–	+	–	Abnormal EEG
3	M/19 months	c.58G>T p.E20* (NS)	3 months/seizure	3 months/GTC	LEV, OXC/seizure free	Moderate	–	+	Introverts, less interaction	Abnormal EEG
4	M/4 years	c.325T>C p.C109R (MS)	8 months/seizure	8 months/GTC	VPA/occasional petit mal	Mild	–	+	Autistic features	Abnormal EEG, MRI at 1 year: abnormal signals scattered in the parenchyma; speech delay, slurred speech
5	F/4 years	c.142C>T p.Q48* (NS)	10 months/seizure	10 months/myoclonic	LEV/seizure free	Mild	–	–	–	Abnormal EEG; motor delay, speak slowly
6	F/30 months	c.462_465del p.Asp155Alafs*70 (FS)	4 months/seizure Delayed neck control	4 months/GTC	LEV, VPA/medically refractory	Profound	–	+	–	Poor head control, poor pursuit of light and objects, cannot sit (6 months). Abnormal EEG
7	M/4 years	c.497T>G p.M166R (MS)	4 months/seizure	4 months/tonic clonic	LEV/seizure free	Moderate	Bilateral cryptorchidism, small penis?	+	–	Abnormal EEG, speech slowly, unable to express long sentences
8	F/5 years	c.292-1G>A, (SS)	4 months/seizure	4 months/focal seizures	OXC/seizure free	Mild	Small jaw, prominent forehead	+	–	Boundary EEG; speech delay; motor delay, motor delay

**, *termination; AED, anti-epileptic drug; EEG, electroencephalogram; FS, frame shift; GTC, generalized tonic–clonic; ID, intellectual disability; LEV, levetiracetam; MRI, magnetic resonance imaging; MS, missense; NS, nonsense; OXC, oxcarbazepine; SS, splice site; VPA, valproic acid.*

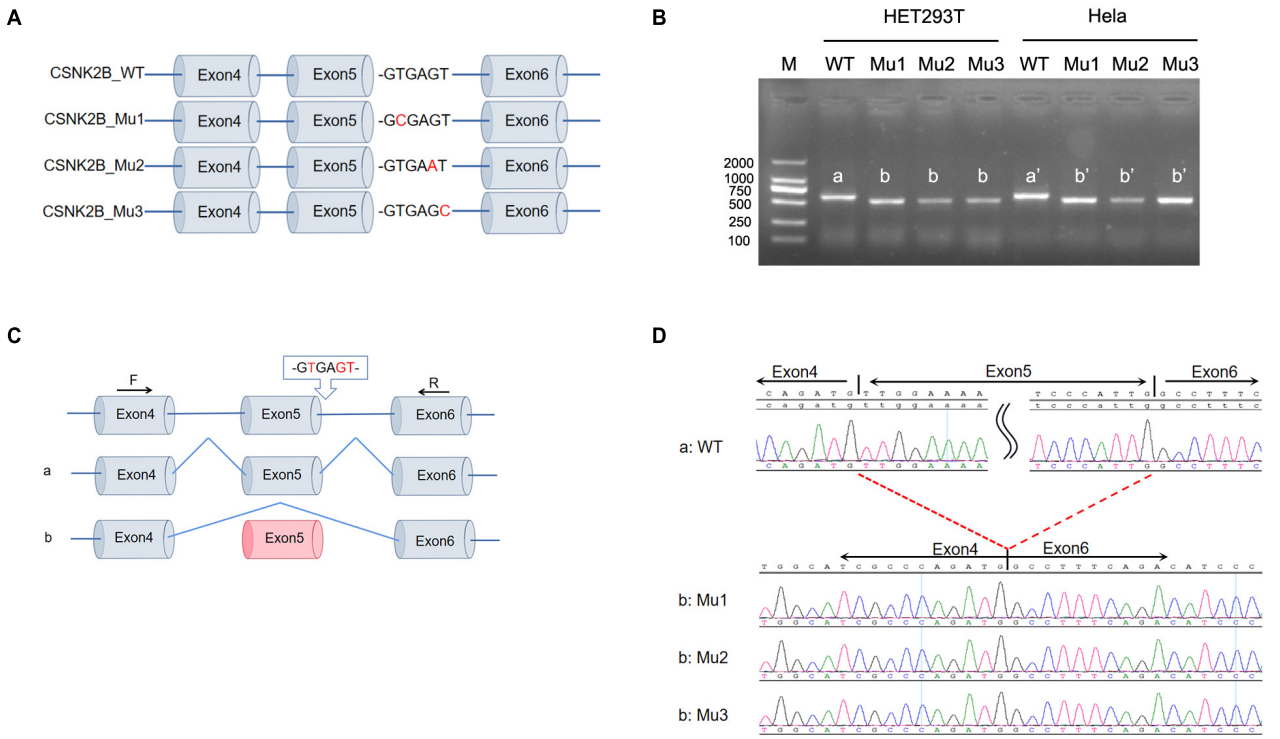


FIGURE 1 | Minigene splicing assay for intron variants. **(A)** Schematic diagram of minigene plasmid construction. Red bases are targeted sites. **(B)** Minigene was transfected into HEK293 cells and Hela cells, respectively. A single band was detected by RT-PCR with a lower molecular weight than wild-type *CSNK2B*. **(C,D)** Sequencing of the above bands revealed that three splicing variants (c.367+2T>C, c.367+5G>A, and c.367+6T>C) all resulted in exon 5 skipping and early termination of the amino acids.

57 individuals with *CSNK2B* mutations have been reported to date (including eight cases in this article), and the largest sample study to date was published in *Epilepsia* by Ernst et al. (2021). The clinical phenotype of POBINDS is complex, with great inter-individual variability and uneven severity of phenotype. Although these studies have expanded the phenotype and genotype spectrum of POBINDS, the genotype–phenotype correlations in POBINDS are still unclear.

Two splicing variants (c.367+5G>A and c.367+6T>C) were located in non-canonical splicing sites. The minigene assays confirmed that these two atypical splice variants lead to skipping of exon 5, resulting in the insertion of a premature stop codon in p.Leu98Alafs*11. A total of 11 splice variants (including three in this article) have been reported in the current literature (11/56), of which six (c.175+2T>G, c.367+2T>C, c.367+5G>A, c.367+6T>C, c.292−2A>T, and c.558−3T>C) have been experimentally confirmed to lead to abnormal splicing of *CSNK2B* (Poirier et al., 2017; Yang et al., 2021). The pathogenicity of atypical splicing site variants is difficult to decipher, and our work confirmed that these variants indeed affect splicing, which is valuable for subsequent research.

Frameshift variants often bring in a premature termination codon, leading to nonsense-mediated decay (NMD) (Shoemaker and Green, 2012). However, among frameshift variants of *CSNK2B*, six out of nine [p.(Met132Leufs*110), p.(Asp155Alafs*71), p.(Leu167Serfs*60), p.(Pro179Tyrfs*49), p.(Asn181Thrfs*46), and p.(Lys208Glnfs*38)], induced a prolonged termination codon downstream of the wild-type termination codon, not upstream. In eukaryotes, mRNAs with these variants may not be directly identified by NMD mechanisms. NMD escape was found in other protein-coding extending like COQ8A, but enzyme activities were significantly reduced (Liu et al., 2014). The functional effects of these six frameshift variants in *CSNK2B* need further verification, but extensive amino acid sequence disruption cannot be ignored. Individuals with these six variants did not show any phenotypic pattern. Another 3 frameshift and 11 nonsense variants were predicted to introduce a stop codon, none of which were in the last exon or the last 50 base pairs of the penultimate exon. In this study, these variants were ranked into LOF variants. Individuals with LOF variants did not show any clear correlations in symptoms, except dysmorphisms (p = 0.041). There may be some objective reasons for this phenomenon, such as a small sample, varying phenotype details, and different indicators of developmental assessment in the age strata.

Interestingly, we found some duplicated mutations in the cohort, which may be hotspot mutations or regions. Ernst et al. (2021) and Yang et al. (2021) have reported three start losses with different variants (c.1A>G, c.2T>A, and c.3G>A). The effect of the start loss mutations is still unpredictable. Firstly, the phenotypes of these three individuals are polytropic, c.1A>G showed only mild speech and motor developmental

delay. c.2T>A had mild ID, while c.3G>A had profound ID. c.2T>A and c.3G>A were both diagnosed with epilepsy, c.2T>A showed drug-refractory epilepsy, while c.3G>A showed seizure-free. Second, the start loss of *CSNK2B* was also recorded in the gnomAD database[4] with a relative high population frequencies. Considering that the start loss variant may theoretically affect the open reading frame and leads to gene silencing (Chen et al., 2020), or re-initiate translation by another "ATG" triple nucleotide and mainly affects phenotypic penetrance (Benkirane et al., 2021), whether those variants are pathogenic requires further study. There were eight individuals with c.G94 mutation, and detailed phenotypes were available for four individuals. In these four individuals, some mutual characteristics were found, such as moderate to severe ID, and malformations. Three individuals had seizures and two were diagnosed with epilepsy, both of which manifested as absence epilepsy and drug-refractory. These phenotype profiles suggest that mutations at the c.G94 may be involved in signaling a variety of developmental processes, such as organogenesis (Bandyopadhyay et al., 2016). The untranslated region (UTR) of exon 5 (c.292-367) may be a hotspot region for *CSNK2B* mutations. A total of 11 individuals were reported in this region, 5 individuals with detailed phenotypes available. Four individuals were diagnosed with epilepsy, all with effective antiepileptic treatment. Three with seizures before 6 months of age. The degree of ID varied from mild to severe, except one individual, had normal development (c.367+5G>A), and three cases had malformations (c.292-1G>A, c.367+2T>C, and c.367+5G>A). Although all five splice site mutations have been experimentally confirmed to cause skipping of exon 5 (except for c.292-1G>A), the varied clinical phenotypes suggest that *CSNK2B* may influence POBINDS onset and progression through some unknown pathways.

Regarding the location of the mutation, the individuals with variants in the zinc finger domain (residues 105–140) had manageable seizures and milder ID, which was inextricably linked to the structure and function of CK2β. Zinc finger structures play a crucial role in the regulation of transcription, adaptation of protein interactions, and post-transcriptional regulation in eukaryotes (Laity et al., 2001; Ecco et al., 2017). Zinc finger mutations (cys109, 114) result in the inability of CK2β to form dimers and loss of the interaction with the catalytic subunit CK2α, which is accompanied by increased degradation of the mutant protein (Canton et al., 2001). The individual in our cohort (c.325T>C) was mildly affected and had manageable epilepsy, mild ID, and mainly language delay. This does not appear to be parallel to the severity of the mutation. However, the situation *in vivo* is complex, and developmental delay may be a direct cause of seizures or due to genetic pleiotropy. Increased degradation of CK2β degradation may trigger a compensatory mechanism that regulates CK2β protein levels. Our data reveal corresponding associations from both clinical and bioinformatics analysis, suggesting that the mutation in this region may be less harmful, and also worth further functional study.

In addition to the zinc finger region, CK2β includes the N-terminal (residues 1–54) containing the autophosphorylation site, an acidic groove region (residues 55–77), and a C-terminal (residues 171–215) that stabilizes the CK2β-CK2α interaction (Reed et al., 1994; Bidwai et al., 1995). We compared missense variants across domains, and the results were not convincing due to the small number of variants. The C-terminal-truncated CK2β loses its ability to bind stably to the catalytic subunit (Sarno et al., 2000). p.(Pro179Tyrfs*49)-mutated CK2β is unable to bind to the CK2α, although it is still expressed (Nakashima et al., 2019). The frameshift of *CSNK2B* (c.499del) resulted in decreased protein expression (Yang et al., 2021). These studies suggest that the *CSNK2B* mutation is likely to affect the function of CK2 in a haploinsufficient manner. However, we cannot exclude a dominant-negative effect of CK2β. It was previously shown that although *CSNK2B* heterozygous mice did not differ significantly from wild-type mice in CK2β expression levels, some of the heterozygous mice failed to survive (Blond et al., 2005). This may have been due to the mutated CK2β forming non-functional tetramers that inhibited the normal catalytic activity of the wild-type CK2α. Such phenomena are not uncommon in proteins that perform functions in multimeric forms (Ahn et al., 2003). The reduced CK2 holoenzyme activity may represent a uniform pathogenesis.

[4]www.gnomad-sg.org

TABLE 2 | Phenotypes of *CSNK2B* clinical and epilepsy characteristics (*n* = 57).

	All (*n* = 57)	LOF (*n* = 34)	Non-LOF (*n* = 23)
Seizures	51	31	19
Epilepsy	48	31	17
Onset of epilepsy			
0–6 months	30	18	12
7–12 months	8	6	2
13–24 months	4	3	1
25–36 months	3	2	1
>36 months	3	2	1
Epilepsy types			
Generalized tonic or tonic–clonic seizures	27	17	10
Absence seizures	6	2	4
Myoclonic seizures	14	10	4
Focal onset seizures	9	5	4
Multiple seizure types	18	11	7
Antiepileptic drugs used	38	25	13
Effective epilepsy control	28	18	11
GDD/ID	43	29	20
Motor delay	37	26	16
Speech delay	37	25	17
Moderate to profound	21	14	10
Hypotonia	17	13	8
Dysmorphisms	17	10	13
Problematic behavior	10	8	3

GDD, global developmental delay; ID, intellectual disability; LOF, loss of function, including start loss, frameshift insertion/deletion, nonsense, and splicing variants; non-LOF, including in-frame insertion/deletion.

There was no statistically significant difference in the characteristics of the POBINDS among mutations of LOF compared to non-LOF, except dysmorphisms (details in ***Supplementary Table 2****). References and databases for all mutation information are detailed in* ***Supplementary Table 1****.*

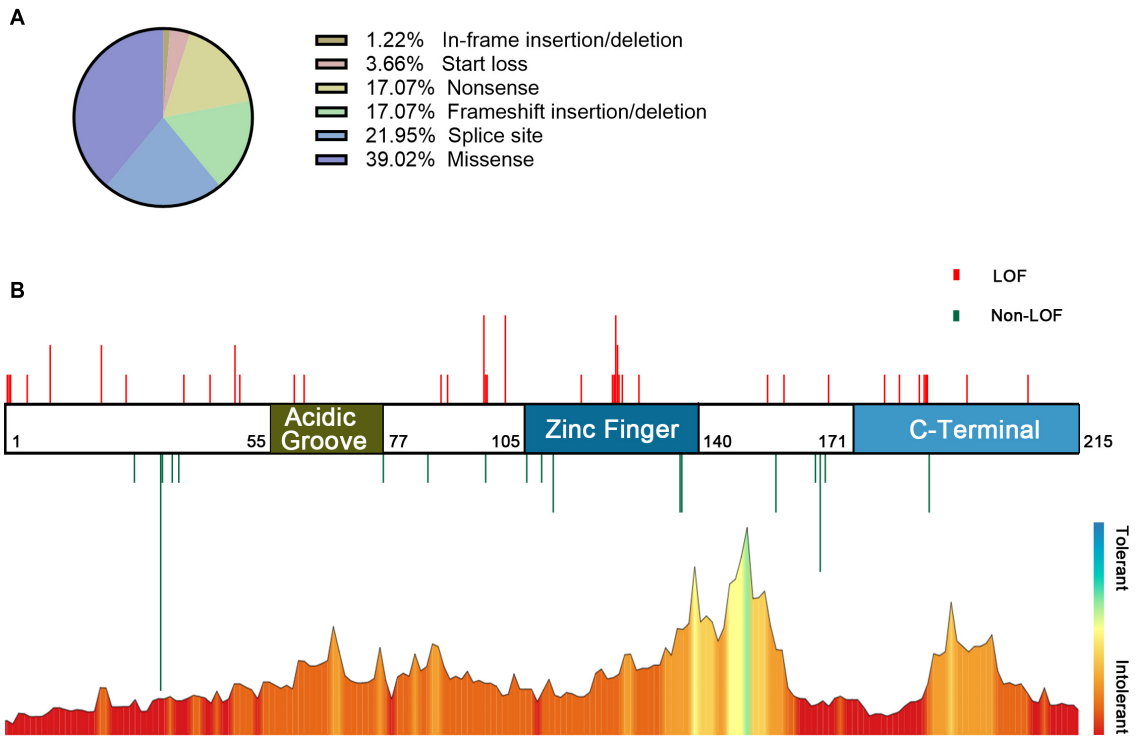


FIGURE 2 | Information on POBINDS-causing CSNK2B mutations (n = 82). **(A)** Distribution of mutation types. **(B)** Schematic representation of the mutation distribution (top) and "MetaDome" tolerance level through whole gene (blew). Red vertical lines represent LOF mutations (including start loss, frameshift insertion/deletion, nonsense, and splicing site), and green vertical lines represent non-LOF mutations (including missense and in-frame insertion/deletion). The length of vertical lines represents the number of reported cases. The zinc finger domain is more tolerant than other regions, which is revealed by the "MetaDome" algorithm, this is also consistent with the clinical findings that missense mutation in the zinc finger domain leads to milder phenotypes. References and databases for all sources of mutation information are detailed in **Supplementary Table 1**.

The severity of the neurological phenotype ultimately depends on the spatiotemporal distribution and catalytic activity of CK2. The CK2 holoenzyme is a heterotetramer composed of two catalytic subunits (α/α′) and two regulatory subunits (β), which can bind as α2β2, αα′β2, or α′2β2 forms. During early development (embryonic and fetal stages), the mRNA levels of CK2β were significantly higher than those in childhood and adulthood (Yang et al., 2018). CK2 activity was higher in cortical pyramidal neurons and large striatal neurons than in other neuronal populations (Castello et al., 2017). Interestingly, in different brain regions of adult rats, the expression of each subunit of CK2 is different at the transcriptional level and the protein level (Castello et al., 2017). This may form distinguishing CK2 heterotetramers in different brain regions, resulting in different enzymatic activity of CK2.

In addition, CK2 is involved in protein phosphorylation, and these affected proteins/signaling pathways may be the main cause of the progression of neurodevelopmental disorders. CK2 catalyzes the phosphorylation of Kv3.1 and calmodulin (Miller, 2000; Macica and Kaczmarek, 2001), the latter of which directly affects the stability of KCNQ2 channels (Benaim and Villalobo, 2002). CK2 is abundantly expressed in the postsynaptic density (Soto et al., 2004), and inhibition of CK2 blocks *N*-methyl-D-aspartate receptor-dependent synaptic transmission and long-term potentiation (Kimura and Matsuki, 2008). Meanwhile, knockdown of *CSNK2B* not only impaired cell differentiation of mouse embryonic NSC (Yang et al., 2018) but also disrupted the migration of mouse immortalized neuronal GN11 cells (Lettieri et al., 2019). The deletion of *CSNK2B* in embryonic NSC impaired the proliferation of forebrain NSCs and inhibited the differentiation of NSCs into oligodendrocyte precursor cells, resulting in defects in brain development (Huillard et al., 2010). Conditional knockout *CSNK2B* mice showed an age-dependent reduction in grip strength, impaired neuromuscular transmission, and inordinate synaptic gene expression (Eiber et al., 2019). Abnormal emotional behavior, including autistic behavior, irritability, depression, and hyperactivity, may be associated with abnormalities in the CK2-mediated dopamine signaling pathway (Rebholz et al., 2009; DiCarlo and Wallace, 2022). CK2 is also involved in signaling for a variety of developmental processes, such as embryogenesis and organogenesis (Bandyopadhyay et al., 2016). The above CK2 signal involvements underlie the phenotypes of neurodevelopmental disorders such as epilepsy, developmental delay, hypotonia, and dysmorphic disorders.

Limitations of this study include the heterogeneity of the clinical data. When analyzing the reported cases, we relied heavily on the authors' descriptions of the individuals, and the data lacked a standardized assessment. Especially concerning the judgment of developmental delay, age stratification needs to

be introduced to standardize the determination of the degree of developmental delay. The reported age of some patients is too young, and some late-onset neurodevelopmental symptoms need to be determined by long-term follow-up. Additionally, the statistical power of our Chi-square and Fisher's exact test was generally below 0.8, which may omit some true positive results. This question requires more samples to solve. Moreover, the mechanism by which *CSNK2B* variants cause CK2 dysfunction remains unclear, especially the effect of missense variants on CK2, which requires more experimental models. With the increase of the functional study and refinement of phenotypic profiles, the clarity of genotype–phenotype correlations is expected to be further improved.

In conclusion, this study further enriched the phenotype–genotype profile of POBINDS. Minigene experiments confirmed that two intron variants resulted in skipping exon 5, leading to a premature termination of mRNA transcription. The pathogenicity of intronic variants requires more experiments to decipher. Taken together, our analysis results showed that missense, nonsense, frameshift, and splicing variants were common forms of *CSNK2B* variants, and individuals with variants in the zinc finger domain had manageable seizures and milder ID. The variant types do not reflect the severity of the POBINDS. Determining further genotype–phenotype correlations in POBINDS requires more functional research and refinement of phenotypic profiles.

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 authors.

ETHICS STATEMENT

This study was reviewed and approved by the Ethics Committee of Xiangya Hospital of Central South University. 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

FY, WZ, and NP designed the research. WZ, FHY, JP, and NP performed the research. WZ, NP, and FY performed the data analysis. WZ, FHY, NP, and FY wrote the manuscript. All authors read, edited, and approved the manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (81701541 and 81771408), the Hunan Key Research and Development Program (No. 2019SK2081), and the Fundamental Research Funds for the Central Universities of Central South University (2019zzts347).

ACKNOWLEDGMENTS

We thank the Clinical Research Center for Children Neurodevelopmental Disabilities of Hunan Province for providing the platform and the personnel who work there for giving support and guidance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.892768/full#supplementary-material

REFERENCES

Ahn, W., Lee, M. G., Kim, K. H., and Muallem, S. (2003). Multiple effects of SERCA2b mutations associated with darier's disease. *J. Biol. Chem.* 278, 20795–20801. doi: 10.1074/jbc.M301638200

Bandyopadhyay, M., Arbet, S., Bishop, C. P., and Bidwai, A. P. (2016). Drosophila protein kinase CK2: genetics, regulatory complexity and emerging roles during development. *Pharmaceuticals* 10:4. doi: 10.3390/ph10010004

Benaim, G., and Villalobo, A. (2002). Phosphorylation of calmodulin. Functional implications. *Eur. J. Biochem.* 269, 3619–3631. doi: 10.1046/j.1432-1033.2002.03038.x

Benkirane, M., Marelli, C., Guissart, C., Roubertie, A., Ollagnon, E., Choumert, A., et al. (2021). High rate of hypomorphic variants as the cause of inherited ataxia and related diseases: study of a cohort of 366 families. *Genet. Med.* 23, 2160–2170. doi: 10.1038/s41436-021-01250-6

Bidwai, A. P., Reed, J. C., and Glover, C. V. (1995). Cloning and disruption of CKB1, the gene encoding the 38-kDa beta subunit of Saccharomyces cerevisiae casein kinase II (CKII). Deletion of CKII regulatory subunits elicits a salt-sensitive phenotype. *J. Biol. Chem.* 270, 10395–10404. doi: 10.1074/jbc.270.18.10395

Blond, O., Jensen, H. H., Buchou, T., Cochet, C., Issinger, O., and Boldyreff, B. (2005). Knocking out the regulatory beta subunit of protein kinase CK2 in mice: gene dosage effects in ES cells and embryos. *Mol. Cell. Biochem.* 274, 31–37. doi: 10.1007/s11010-005-3117-x

Canton, D. A., Zhang, C., and Litchfield, D. W. (2001). Assembly of protein kinase CK2: investigation of complex formation between catalytic and regulatory subunits using a zinc-finger-deficient mutant of CK2beta. *Biochem. J.* 358(Pt 1), 87–94. doi: 10.1042/0264-6021:3580087

Castello, J., Ragnauth, A., Friedman, E., and Rebholz, H. (2017). CK2-An emerging target for neurological and psychiatric disorders. *Pharmaceuticals* 10:7. doi: 10.3390/ph10010007

Chen, S., Xie, W., Liu, Z., Shan, H., Chen, M., Song, Y., et al. (2020). CRISPR start-loss: a novel and practical alternative for gene silencing through base-editing-induced start codon mutations. *Mol. Ther. Nucleic Acids* 21, 1062–1073. doi: 10.1016/j.omtn.2020.07.037

DiCarlo, G. E., and Wallace, M. T. (2022). Modeling dopamine dysfunction in autism spectrum disorder: from invertebrates to vertebrates. *Neurosci. Biobehav. Rev.* 133:104494. doi: 10.1016/j.neubiorev.2021.12.017

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. *Development* 144, 2719–2729. doi: 10.1242/dev.132605

Eiber, N., Rehman, M., Kravic, B., Rudolf, R., Sandri, M., and Hashemolhosseini, S. (2019). Loss of protein kinase Csnk2b/CK2β at neuromuscular junctions affects morphology and dynamics of aggregated nicotinic acetylcholine receptors, neuromuscular transmission, and synaptic gene expression. *Cells* 8:940. doi: 10.3390/cells8080940

Ernst, M. E., Baugh, E. H., Thomas, A., Bier, L., Lippa, N., Stong, N., et al. (2021). CSNK2B: a broad spectrum of neurodevelopmental disability and epilepsy severity. *Epilepsia* 62, e103–e109. doi: 10.1111/epi.16931

Graham, K. C., and Litchfield, D. W. (2000). The regulatory beta subunit of protein kinase CK2 mediates formation of tetrameric CK2 complexes. *J. Biol. Chem.* 275, 5003–5010. doi: 10.1074/jbc.275.7.5003

Guerra, B., Siemer, S., Boldyreff, B., and Issinger, O. G. (1999). Protein kinase CK2: evidence for a protein kinase CK2beta subunit fraction, devoid of the catalytic CK2alpha subunit, in mouse brain and testicles. *FEBS Lett.* 462, 353–357. doi: 10.1016/s0014-5793(99)01553-7

Huillard, E., Ziercher, L., Blond, O., Wong, M., Deloulme, J., Souchelnytskyi, S., et al. (2010). Disruption of CK2beta in embryonic neural stem cells compromises proliferation and oligodendrogenesis in the mouse telencephalon. *Mol. Cell. Biol.* 30, 2737–2749. doi: 10.1128/MCB.01566-09

Kimura, R., and Matsuki, N. (2008). Protein kinase CK2 modulates synaptic plasticity by modification of synaptic NMDA receptors in the hippocampus. *J. Physiol.* 586, 3195–3206. doi: 10.1113/jphysiol.2008.151894

Kong, H., Yin, F., He, F., Omran, A., Li, L., Wu, T., et al. (2015). The effect of miR-132, miR-146a, and miR-155 on MRP8/TLR4-Induced Astrocyte-Related inflammation. *J. Mol. Neurosci.* 57, 28–37. doi: 10.1007/s12031-015-0574-x

Laity, J. H., Lee, B. M., and Wright, P. E. (2001). Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struc. Biol.* 11, 39–46. doi: 10.1016/s0959-440x(00)00167-6

Lettieri, A., Borgo, C., Zanieri, L., D'Amore, C., Oleari, R., Paganoni, A., et al. (2019). Protein kinase CK2 subunits differentially perturb the adhesion and migration of GN11 cells: a model of immature migrating neurons. *Int. J. Mol. Sci.* 20:5951. doi: 10.3390/ijms20235951

Li, J., Gao, K., Cai, S., Liu, Y., Wang, Y., Huang, S., et al. (2019). Germline de novo variants in CSNK2B in Chinese patients with epilepsy. *Sci. Rep.* 9:17909. doi: 10.1038/s41598-019-53484-9

Liu, Y. T., Hersheson, J., Plagnol, V., Fawcett, K., Duberley, K. E. C., Preza, E., et al. (2014). Autosomal-recessive cerebellar ataxia caused by a novel ADCK3 mutation that elongates the protein: clinical, genetic and biochemical characterisation. *J. Neurol. Neurosurg. Psychiatry* 85, 493–498. doi: 10.1136/jnnp-2013-306483

Macica, C. M., and Kaczmarek, L. K. (2001). Casein kinase 2 determines the voltage dependence of the Kv3.1 channel in auditory neurons and transfected cells. *J. Neurosci.* 21, 1160–1168. doi: 10.1523/JNEUROSCI.21-04-01160.2001

Miller, C. (2000). An overview of the potassium channel family. *Genome Biol.* 1:S4. doi: 10.1186/gb-2000-1-4-reviews0004

Nakashima, M., Tohyama, J., Nakagawa, E., Watanabe, Y., Siew, C. N. G., Kwong, C. S., et al. (2019). Identification of de novo CSNK2A1 and CSNK2B variants in cases of global developmental delay with seizures. *J. Hum. Genet.* 64, 313–322. doi: 10.1038/s10038-018-0559-z

Niefind, K., Guerra, B., Ermakowa, I., and Issinger, O. G. (2001). Crystal structure of human protein kinase CK2: insights into basic properties of the CK2 holoenzyme. *EMBO J.* 20, 5320–5331. doi: 10.1093/emboj/20.19.5320

Peng, J., Wang, Y., He, F., Chen, C., Wu, L. W., Yang, L. F., et al. (2018). Novel West syndrome candidate genes in a Chinese cohort. *CNS Neurosci. Ther.* 24, 1196–1206. doi: 10.1111/cns.12860

Poirier, K., Hubert, L., Viot, G., Rio, M., Billuart, P., Besmond, C., et al. (2017). CSNK2B splice site mutations in patients cause intellectual disability with or without myoclonic epilepsy. *Hum. Mutat.* 38, 932–941. doi: 10.1002/humu.23270

Rebholz, H., Nishi, A., Liebscher, S., Nairn, A. C., Flajolet, M., and Greengard, P. (2009). CK2 negatively regulates Galphas signaling. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14096–14101. doi: 10.1073/pnas.0906857106

Reed, J. C., Bidwai, A. P., and Glover, C. V. (1994). Cloning and disruption of CKB2, the gene encoding the 32-kDa regulatory beta'-subunit of Saccharomyces cerevisiae casein kinase II. *J. Biol. Chem.* 269, 18192–18200.

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

Sarno, S., Marin, O., Boschetti, M., Pagano, M. A., Meggio, F., and Pinna, L. A. (2000). Cooperative modulation of protein kinase CK2 by separate domains of its regulatory beta-subunit. *Biochemistry* 39, 12324–12329. doi: 10.1021/bi0011431

Shoemaker, C. J., and Green, R. (2012). Translation drives mRNA quality control. *Nat. Struct. Mol. Biol.* 19, 594–601. doi: 10.1038/nsmb.2301

Soto, D., Pancetti, F., Marengo, J. J., Sandoval, M., Sandoval, R., Orrego, F., et al. (2004). Protein kinase CK2 in postsynaptic densities: phosphorylation of PSD-95/SAP90 and NMDA receptor regulation. *Biochem. Bioph. Res. Commun.* 322, 542–550. doi: 10.1016/j.bbrc.2004.07.158

Wiel, L., Baakman, C., Gilissen, D., Veltman, J. A., Vriend, G., and Gilissen, C. (2019). MetaDome: pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum. Mutat.* 40, 1030–1038. doi: 10.1002/humu.23798

Yang, C., Li, X., Wu, Y., Shen, Q., Zeng, Y., Xiong, Q., et al. (2018). Comprehensive integrative analyses identify GLT8D1 and CSNK2B as schizophrenia risk genes. *Nat. Commun.* 9:838. doi: 10.1038/s41467-018-03247-3

Yang, S., Wu, L., Liao, H., Lu, X., Zhang, X., Kuang, X., et al. (2021). Clinical and genetic analysis of six Chinese children with Poirier-Bienvenu neurodevelopmental syndrome caused by CSNK2B mutation. *Neurogenetics* 22, 323–332. doi: 10.1007/s10048-021-00649

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.

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fgene.2022.940776

OPEN ACCESS

EDITED BY
Xiaoli Chen,
Capital Institute of Pediatrics, China

REVIEWED BY
Yanyan Qian,
Fudan University, China
Tianyun Wang,
Peking University, China

*CORRESPONDENCE
Niu Li,
liniu0509@163.com
Jian Wang,
Labwangjian@shsmu.edu.cn

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 10 May 2022
ACCEPTED 04 July 2022
PUBLISHED 22 July 2022

CITATION
Ding Y, Chen J, Tang Y, Chen L-N, Yao R-E, Yu T, Yin Y, Wang X, Wang J and Li N (2022), Identification and functional analysis of novel SOX11 variants in Chinese patients with Coffin-Siris syndrome 9.
Front. Genet. 13:940776.
doi: 10.3389/fgene.2022.940776

Identification and functional analysis of novel *SOX11* variants in Chinese patients with Coffin-Siris syndrome 9

Yu Ding[1†], Jiande Chen[2†], Yijun Tang[1], Li-Na Chen[3], Ru-En Yao[3,4,5], Tingting Yu[3,4,5], Yong Yin[2], Xiumin Wang[1], Jian Wang[3,4,5]* and Niu Li[3,4,5]*

[1]Department of Endocrinology and Metabolism, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Respiratory Medicine, Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China, [3]Department of Medical Genetics and Molecular Diagnostic Laboratory, Shanghai Children's Medical Center, Shanghai Jiaotong University School of Medicine, Shanghai, China, [4]Shanghai Key Laboratory of Clinical Molecular Diagnostics for Pediatrics, Shanghai, China, [5]Shanghai Clinical Research Center for Rare Pediatric Diseases, Shanghai, China

SOX11 is a transcription factor belonging to the sex determining region Y-related high-mobility group box family that plays a vital role in early embryogenesis and neurogenesis. *De novo* variants in *SOX11* have been initially reported to cause a rare neurodevelopmental disorder, mainly referred to Coffin-siris syndrome 9 (CSS9, OMIM# 615866) which is characterized with growth deficiency, intellectual disability (ID), microcephaly, coarse facies, and hypoplastic nails of the fifth fingers and/or toes. A recent large-scale cohort study suggests that *SOX11* variation would result in a clinically and molecularly distinct disease from CSS. Here, we describe three unrelated Chinese cases with variable phenotype, mainly involving developmental delay, ID, short statute, microcephaly, facial deformities (i.e., prominent forehead, arched eye brow, flat nasal bridge, broad nose and short philtrum), and cryptorchidism. Whole-exome sequencing (WES) revealed three novel heterozygous variants in the *SOX11* gene, including two missense variants of c.337T>C (p.Y113H) and c.425C>G (p.A142G), and one nonsense variant of c.820A>T (p. K142*). Luciferase reporting assay shows that the two missense variants impair the transcriptional activity of the *SOX11* target gene *GDF5*. Additionally, WES uncovered a 4,300 kb deletion involving the region of 1q24.2-q25.1 (hg19,chr1:169,433,149-173,827,682) in patient 1, which also contributes to the condition of the patient. In summary, this is the first report of Chinese cases with *de novo* variants of *SOX11*. Our study partially supports the previous observation that the phenotype caused by *SOX11* variants somewhat differs from classical CSS.

KEYWORDS

SOX11, coffin-siris syndrome, missense variants, functional study, phenotypic differences

Introduction

The sex determining region Y (SRY)-related high-mobility group (HMG) box (SOX) family encodes a group of transcription factors that play essential roles in cell fate decisions during many developmental processes (Sarkar and Hochedlinger, 2013). Thus far, a total of eight SOX subgroups (A, B1/B2, C, D, E, F, G, and H) with twenty members have been identified, all of which harbor a highly conserved DNA-binding HMG domains (Bowles et al., 2000). It is well known that dysfunction of the SOX proteins plays a critical role in the occurrence and development of multiple malignant tumors (i.e., hepatocellular carcinoma) *via* transcriptional activation or suppression of distinct downstream targets or signaling pathways (de la Rocha et al., 2014; Grimm et al., 2020; Luo et al., 2022). In addition, germline variants in twelve of SOX members can lead to kinds of human genetic disorders, including *SRY* (OMIM# 480000), *SOX2* (OMIM# 184429), *SOX3* (OMIM# 313430), *SOX4* (OMIM# 184430), *SOX5* (OMIM# 604975), *SOX6* (OMIM# 607257), *SOX8* (OMIM# 605923), *SOX9* (OMIM# 608160), *SOX10* (OMIM# 602229), *SOX11* (OMIM# 600898), *SOX17* (OMIM# 610928), and *SOX18* (OMIM# 601618) (Angelozzi and Lefebvre, 2019).

The *SOX11* gene belongs to subgroup C of SOX family, which locates at 2p25.2 and contains only one exon to encode a small protein comprising 441 amino acids (NP _003099.1; UCSC database, http://genome.ucsc.edu). In 2014, Tsurusaki et al. first demonstrated that *de novo* variants in *SOX11* can result in Coffin-siris syndrome 9 (CSS9, OMIM#615866) in two patients and animal models (Tsurusaki et al., 2014). CSS is usually characterized with growth deficiency, intellectual disability (ID), microcephaly, coarse facies, and hypoplastic nails of the fifth fingers and/or toes (Kosho et al., 2014; Sokpor et al., 2017). In 2016, Hempel et al. confirmed the neural phenotype related to CSS9 in ten patients with *de novo* single nucleotide variants (SNVs) of *SOX11* or microdeletions of the chromosome 2p25.2 containing *SOX11* (Hempel et al., 2016). Along with the identification of new patients, several new clinical features were reported, including coarctation of the aorta (Okamoto et al., 2018), cleft palate (Khan et al., 2018), and glaucoma (Diel et al., 2021). More recently, a large cohort study revealed that developmental delay (DD) or ID, microcephaly, short stature, and low body weight were common characteristics in patients with *SOX11* variants. In addition, ocular malformations (oculomotor apraxia, coloboma, and microphthalmia) and hypogonadotropic hypogonadism were also reported in patients with *SOX11* variation (Al- Jawahiri et al., 2022). These studies indicate phenotypic complexity from *SOX11* variation, and a full understanding of such disease requires the accumulation of additional cases from different ethnic groups.

Here, we reported three unrelated Chinese patients with distinct *de novo* variants in *SOX11* gene which were identified by whole-exome sequencing (WES). Phenotypic analysis showed that most of the clinical features of the three patients were consistent with those reported, but there were still some differences. Additionally, the impact of the two missense variants (Y113H and A142G) on SOX11 protein was investigated by *in vitro* experiments.

Materials and methods

Patients

Three unrelated patients (two males and one female) born to physically healthy and non-consanguineous parents were enrolled in this study. Written informed consent has been obtained from each family.

DNA sequencing

WES was performed in all three patients as previously described (Wang et al., 2014; Wang et al., 2016). Briefly, genomic DNA was extracted from patients' peripheral blood and was then sheared to create fragments of 150 to 200 bp. Sequencing library was prepared using the SureSelect XT Human All Exon V6 kit (Agilent Technologies, Santa Clara, CA, United States), and sequencing was performed by the Illumina NovaSeq 6000 System (Illumina, San Diego, CA, United States). After base calling, quality assessment, and alignment of the sequence reads to the reference human genome (GRCh37, dbSNP135), all single nucleotide variants (SNVs) and indels were saved as VCF format file which was then uploaded to the QIAGEN Clinical Insight (QCI) Interpret Translational tool (https://apps.qiagenbioinformatics.cn/) for filtering and annotation. The *SOX11* variants identified by WES were validated by Sanger sequencing using the ABI 3700 sequencer (Applied Biosystems, Foster City, CA, United States), in indicated patient and their parents (sequencing primers were available upon request).

Cell culture

HEK293T cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal bovine serum (Sigma-Aldrich, St. Louis, MO, United States) and 1% penicillin/streptomycin (Thermo Fisher Scientific, Waltham, MA, United States) in a 5% CO2 incubator at 37°C.

Plasmids construction and transfection

The wild-type (WT) and mutant *SOX11* cDNA were synthesized by BGI (Shenzhen, China), and were then cloned

into pcDNA3.1(+)-N-MYC vector, respectively. The *GDF5* promoter 5′-flanking sequence (−448/+319) was synthesized and cloned into the pGL3-basic vector. Plasmid DNA was transfected into cells using Lipofectamine™ LTX reagent (Thermo Fisher Scientific).

Immunofluorescence

HEK293T cells were cultured on cover slips in 12-well plate at 80% confluence prior for transfection. 24 h after transfection with the WT or mutant *SOX11* constructs, cells were washed by 1 x PBS and fixed using 4% paraformaldehyde for 15 min at room temperature. Samples were then washed with 1 × PBS three times and blocked in the blocking buffer (1x PBS/5% goat serum/0.3% Triton X-100) for 1 h. Coverslips were incubated with mouse anti-Myc monoclonal antibody (Cell Signaling Technology, Danvers, MA, United States) at 4°C overnight. The cover slips were then mounted on microscope slides using ProLong® Gold Antifade reagent with DAPI (Cell Signaling Technology) and analyzed using a Leica DM6000 fluorescence microscope (Leica Microsystems, Wetzlar, Germany).

Western blotting

HEK293T cells were seeded into 6-well plate and transfected with the indicated *SOX11* constructs. 24 h after transfection, cells were washed with ice-cold 1x PBS and lysed with SDS lysis buffer (100 mM pH6.8 Tris-HCl, 10% Glycerol, 1% SDS). Whole cell lysates were separated on 10% SDS-PAGE gels, transferred to polyvinylidene difluoride (PVDF) membranes. After blocking in 5% non-fat milk in TBS-T for 1 h, membranes were then incubated overnight at 4°C with mouse anti-Myc monoclonal antibody (Cell Signaling Technology) and mouse anti-beta-Actin monoclonal antibody (Sigma-Aldrich). Proteins were detected using a chemiluminescence system with a horseradish peroxidase conjugated secondary antibody.

Luciferase reporter assays

GDF5-luc plasmid was used to monitor the regulatory role of *SOX11* in modulating the transcription. Briefly, HEK293T cells were seeded into 96-well plate at a density of 3×10^4 per well. 24 h after culture, the cells were co-transfected with 28 ng *GDF5*-luc, 2 ng pRL-SV40, and 70 ng control or *SOX11* expression vectors. 24 h after transfection, luciferase activity was determined by the Dua-Glo luciferase assay system (Promega Corporation, Madison, WI, United States), and normalized to renilla luciferase.

FIGURE 1
Clinical characteristics of the patients. **(A)** Facial photographs of patients 1 and 2. **(B)** Patient 1 had short fingers, clinodactyly of the fifth finger, and abnormal knuckles activity.

Results

Clinical description of the patients

Patient 1 was a male who was born at full-term gestation with a birth weight of 2100 g (−3.1 SD) by spontaneous vaginal delivery. At the stage of pregnancy, intrauterine growth restriction (IUGR) was noticed. At 3-year old, he was referred to our hospital due to growth retardation. Physical examination showed a height of 79 cm (−4.7 SD) and weight of 9.6 kg (−3.3 SD). His hair was sparse and the anterior hairline was high. He presented with facial deformities consisting of prominent forehead, arched eye brows, flat nasal bridge and broad nose, low-set ears, short philtrum, and micrognathia (Figure 1A). He had left cryptorchidism, short fingers, and clinodactyly of the 5th finger. In addition, finger joint laxity was also noticed in the patient (Figure 1B). X-ray showed delay bone age which was 1.5-year old. No DD or ID was found in the

TABLE 1 Phenotypic comparison of our patient with reported patients.

	Patient 1	Patient 2	Patient 3	Reported patients (n = 58)
General information				
Gender	Male	Female	Male	29 female/28 male /1 neonate
Born at full term	+	+	+	12/14
IUGR	+	−	−	5/7
Age	3-years	5.5-years	8-years	—
Birth length (cm)	Unknown	50 (0.2 SD)	49 (−0.8 SD)	—
Birth weight (kg)	2.1 (−3.1 SD)	3.5 (0.7 SD)	3.2 (−0.3 SD)	—
Current height (cm)	79 (−.7 SD)	113 (−0.1 SD)	118 (−2.2 SD)	10 < 2 SD (n = 30)
Current weight (kg)	9.6 (−3.3 SD)	18.7 (−0.3 SD)	17 (−2.5 SD)	11 < 2 SD (n = 29)
Head circumference (cm)	Normal	Normal	46 (−3.0 SD)	11 < 2 SD (n = 28)
Microcephaly	−	−	+	11/28
Short stature	+	−	+	10/30
Facial features				
Prominent forehead	+	+	+	1/7
Arched eye brow	+	+	−	6/16
Flat nasal bridge	+	+	+	3/12
Broad nose	+	+	−	3 with short nose
Short philtrum	+	+	−	7/17
Open mouth	−	+	−	7/15
Abnormal ears	+	+	−	5/7
Micrognathia	+	−	−	2/10
Cleft palate	−	+	−	1/7
Cleft lip	−	+	−	0/15
Neurodevelopment				
DD	−	+	+	36/37
ID	N/R	+	+	38/40
Skeletal malformations				
Clinodactyly	5th finger	−	−	10/55
Hypoplastic nails	−	−	−	6/55
Joint laxity	Fingers	−	−	1/3
Others				
Sparse scalp hair	+	+	−	6/10
High hairline	+	+	+	5/9
Cryptorchidism	+	N/A	+	5/6

DD, developmental delay; ID, intellectual disability; SD, standard deviation; N/A, not applicable; N/R, not reported.

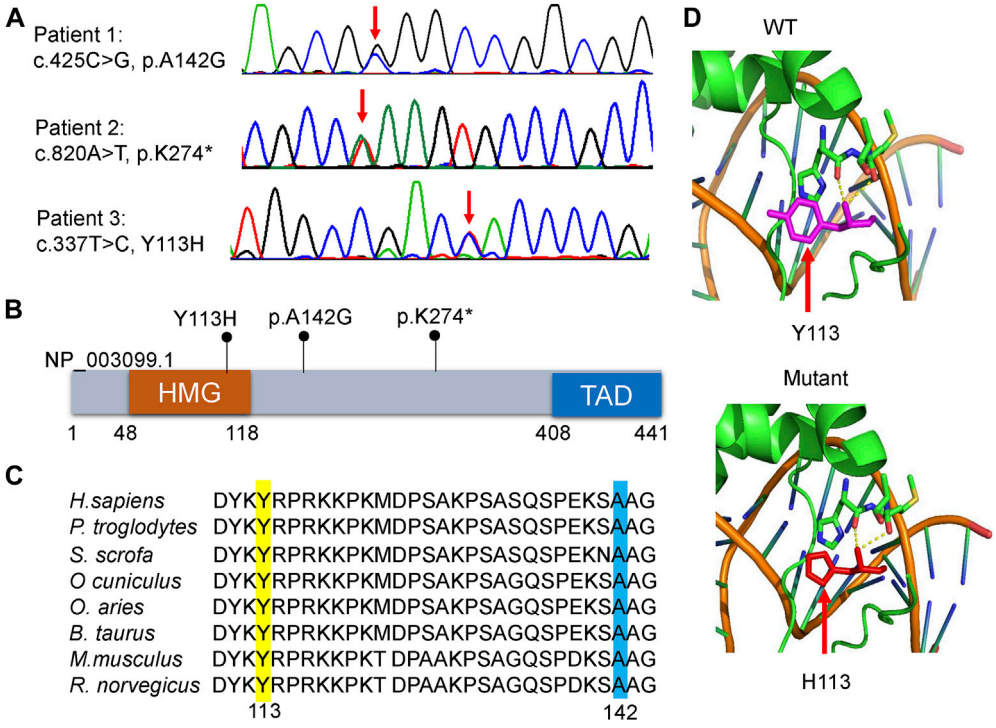


FIGURE 2
Molecular characteristics of the patients. **(A)** Sanger sequencing revealed that each patient harbored a heterozygous variant in *SOX11* gene (NM_003108.4). Red arrows indicate the variant base. **(B)** Distribution schematic of the three variants of *SOX11* gene identified in this study. Of them, the variant of Y113H localize to the high-mobility group (HMG) domain. TAD, transactivating domain. **(C)** Inter-species amino acid sequence alignment to show the missense variants of Y113H (yellow) and A142G (blue) within a highly conserved region of the protein. **(D)** Solved and predicted three-dimensional models of WT and mutant (Y113H) SOX11. The crystal structure was simulated using the mouse Sox4 HMG domain.

patient. During the last follow up, his parents denied the child have mental development problems and refused further assessment, though learning difficulties were mentioned. No special was recorded for the development of the gross motor skills of the patient. The levels of growth hormone, IGF-1, IGF-BP3, LH and FSH were normal. Cranial magnetic resonance imaging (MRI) showed his pituitary gland was thin and oblate. Abdominal ultrasound was normal.

Patient 2 was a 5.5-year old female who was born at full-term gestation with birth weight of 3500 g (0.7 SD) and birth length of 50 cm (0.2 SD) by cesarean section. She was able to sit at 6-month old and walk alone at 20-month old. At the age of 5-year old, she was found with severe speech impairment and moderate ID. Physical examination showed a height of 113 cm (−0.1 SD) and weight of 18.7 kg (−0.3 SD). She had prominent forehead, high hairline, sparse scalp hair, arched eye brows, short philtrum, flat nasal bridge and broad nose, hypoplastic right nasal alar, auricle malformation, everted upper lip, and cleft palate and lip (Figure 1A). X-ray of the bone age, cranial MRI, and the abdominal ultrasound were normal.

Patient 3 was an 8-year old male patient born at full-term gestation by cesarean section. He was admitted to the Endocrinology department due to a slow increase in height, which was less than 4 cm/year for 4 years. His birth weight was 3200 g (−0.3 SD) and birth length was 49 cm (−0.8 SD). Physical examination showed a height of 118 cm (−2.2 SD) and weight of 17 kg (−2.5 SD). He had prominent forehead, high hairline, flat nasal bridge, and bilateral cryptorchidism. Specialist assessment suggested he had DD and mild ID. X-ray showed a two-year delay in his bone age. The levels of growth hormone, IGF-1, IGF-BP3, LH and FSH were normal. Testosterone levels had a good response after HCG stimulation, increasing from less than 0.01–0.28 ng/ml. The cranial MRI and abdominal ultrasound were normal.

We compared the clinical phenotypes of our patients with those previously reported (Tsurusaki et al., 2014; Hempel et al., 2016; Khan et al., 2018; Okamoto et al., 2018; Sekiguchi et al., 2019; Diel et al., 2021; Wakim et al., 2021; Al- Jawahiri et al., 2022; Cho et al., 2022), which have been summarized by Al-Jawahiri et al. (Al- Jawahiri et al., 2022). Although the clinical features of the three Chinese patients were variable, they still fell within the spectrum of *SOX11* syndrome (Table 1). In summary,

TABLE 2 Pathogenicity Predictions for Y113H and A142G variants of *SOX11*.

	c.337T>C, Y113H		c.425C>G, A142G	
***In silico* tools**	**Score**	**Prediction**	**Score**	**Prediction**
PolyPhen-2 (v2.2)	0.999	Probably_damaging	0.904	Possibly_damaging
MutationTaster2	1	Disease_causing	0.976	Disease_causing
CADD (v1.6)	26.1	Damaging	24.4	Damaging
ClinPred	0.993	Pathogenic	0.726	Pathogenic

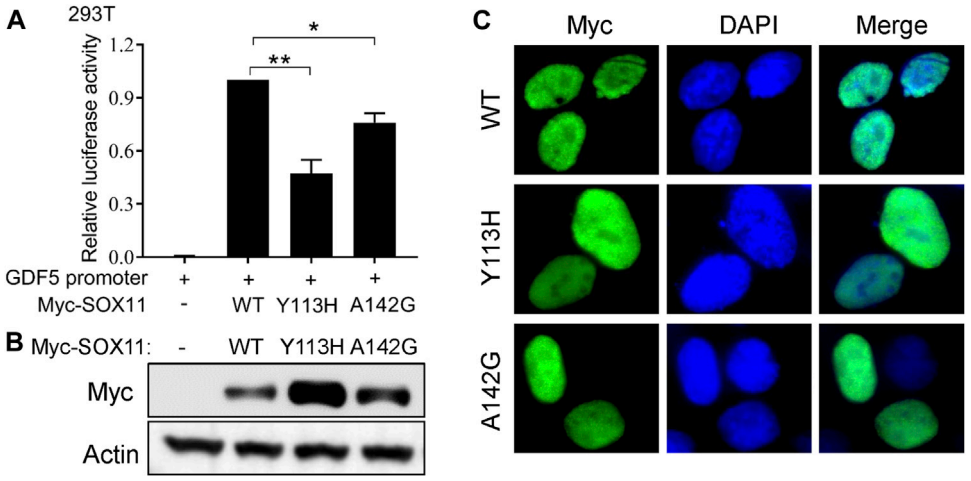


FIGURE 3
Functional study of the SOX11 missense variants. **(A)** Transcriptional activity of the *GDF5* promoter was determined by luciferase reporter assays in HEK293T cells after co-transfection of the WT or the mutant Myc-*SOX11* and the *GDF5* promoter reporter construct. Data are presented as mean values ±s.e.m. from three independent experiments. *$p < 0.05$, **$p < 0.01$, two-tailed Student's *t*-test. **(B)** The expression level of the WT and the mutant (Y113H and A142G) Myc-SOX11 were evaluated by immunoblotting in HEK293T cells. **(C)** Immunofluorescence analysis to show the subcellular localization of the WT and the mutant (Y113H and A142G) Myc-SOX11.

patients 1 and 2 had highly consistent facial deformities, comprising of prominent forehead, arched eye brow, flat nasal bridge, broad nose, short philtrum, and abnormal ears. In comparison, the facial deformity of patient 3 was much milder. Sparse scalp hair and high hairline were also noticed in both patients 1 and 2. Patients 2 and 3 had DD and ID, and both male patients had short stature and cryptorchidism. In addition, patient 1 harbored the 5th finger clinodactyly and microcephaly was only found in patient 3.

Identification and *in silico* analysis of the *SOX11* variants in the patients

WES identified three heterozygous *SOX11* variants, including the missense variant of c.425C>G, p.A142G in patient 1, the nonsense variant of c.820A>T, p.K274* in patient 2, and the missense variant of c.337T>C, p.Y113H in patient 3. Sanger sequencing confirmed these variants in the patients (Figure 2A), and also revealed the *de novo* status of each variant from the results that their parents were WT in *SOX11* gene. All of the three variants were not reported in previous cases and not included in the known public database (i.e., gnomAD, HGMD and ClinVar), suggesting they were novel.

The K274* is not located in the last exon or within the 3′-most 50 nucleotides of *SOX11*, so it is very likely to lead to nonsense-mediated mRNA decay (Abou Tayoun et al., 2018). The Y113 residue locates at the HMG domain of SOX11 protein (Figure 2B), which is highly conserved in multiple species (Figure 2C). Crystal structure analysis using the mouse Sox4 HMG domain showed the Y113H variant may not alter the secondary structure of SOX11, but is likely to affect its ability

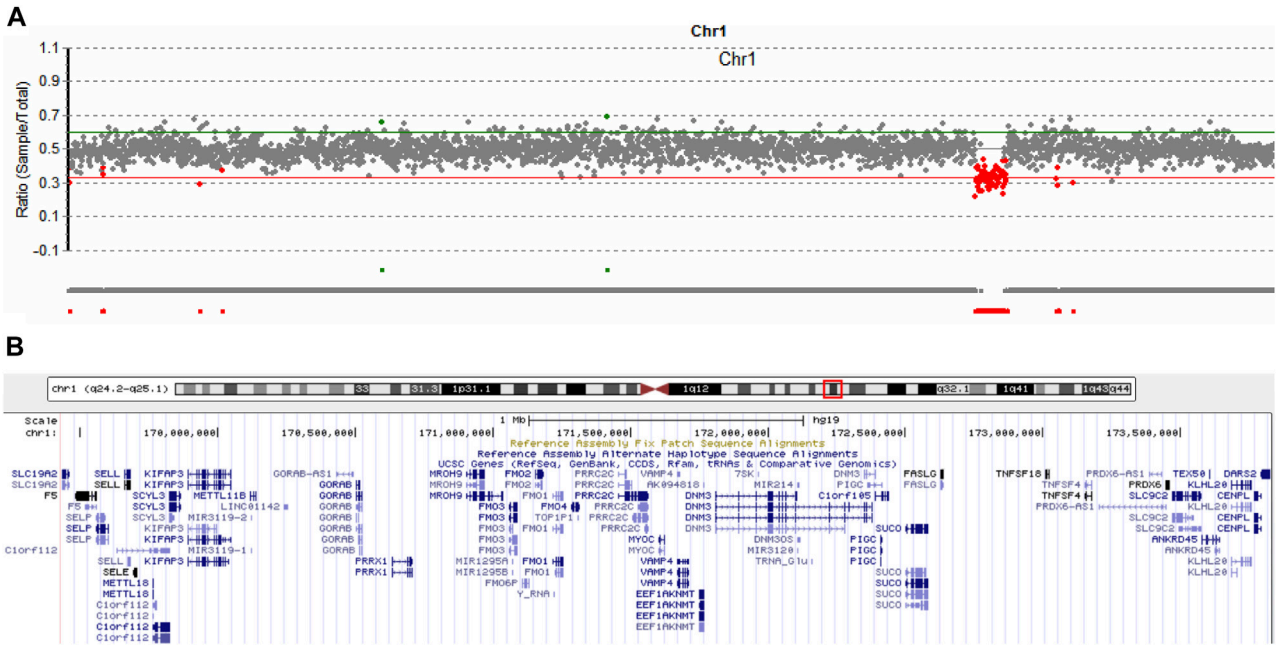


FIGURE 4
Copy number variation analysis in patient 1. **(A)** Sequencing depth analysis using the WES data showed patient 1 harbored 1q24.2-q25.1 deletion (hg19, chr1:169,433,149-173,827,682). **(B)** Involving genes in the 1q24.2-q25.1 region.

to bind DNA (Figure 2D). By contrast, the other two variants of A142G and K274* locate downstream of the HMG domain (Figure 2B), and the A142 residue is also highly conserved (Figure 2C). In addition, the two missense variants of Y113H and A142G are predicted to be deleterious according to various *in silico* tools including PolyPhen-2, MutationTaster2, CADD, and ClinPred (Table 2).

Novel missense variants of *SOX11* impair the transcriptional activity

To further evaluate the impact of the Y113H and A142G variants on SOX11 function, we first performed luciferase assay using a reporter construct containing a fragment of *GDF5* promoter which has been described previously (Tsurusaki et al., 2014). Compared to the with WT *SOX11*, both Y113H and A142G mutants result in decreased transcriptional activities (Figure 3A). *In vitro* experiments in HEK293T cells revealed that both missense variants increased SOX11 expression (Figure 3B), which were also observed in previous studies (Tsurusaki et al., 2014; Hempel et al., 2016). Immunofluorescence analysis showed that the WT and two missense mutant SOX11 protein localized in the nucleus (Figure 3C).

Identification of the genomic region deletion in patient 1

It is estimated that near 5% of patients can have more than one molecular findings, and may cause more serious phenotypic features (Posey et al., 2017). We routinely analyze copy number variation (CNV) of each patient by comparing the read-depth with the WES data from the other samples of the same batch (Yao et al., 2017; Yao et al., 2019). As shown in Figure 4A, patient 1 harbored a 4,300 kb deletion involving the region of 1q24.2-q25.1 (hg19,chr1:169,433,149-173,827,682), which includes multiple disease-causing genes, such as *EEF1AKNMT* (OMIM#617987), *FASLG* (OMIM#134638), *FM O 3* (OMIM#136132), *GORAB* (OMIM#607983), *MYOC* (OMIM#601652), *PIGC* (OMIM#601730), *PRRX1* (OMIM#167420), *SLC19A2* (OMIM#603941), and *TNFSF4* (OMIM# 603,594) (Figure 4B). Deletion of this genomic region has been described in series of patients (Burkardt et al., 2011; Chatron et al., 2015). According to the clinical interpretation of CNV by the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen) (Riggs et al., 2020), the microdeletion of 1q24.2-q25.1 in this patient was classified as pathogenic. No questionable CNVs were identified in the other two patients.

Discussion

In this study, we reported three novel *SOX11* variants (Y113H, A142G, and K274*) in Chinese patients. According to the clinical interpretation of genetic variants by the ACMG/AMP 2015 guideline (Richards et al., 2015), The K274* nonsense variant was classified as pathogenic (PVS1+PS2+PM2), *In vitro* experiments show that Y113H and A142G has no effect on the nuclear localization of SOX11 protein, but impairs the transcriptional activity of the target gene *GDF5*. Therefore, the two missense variants are also classified as pathogenic (PS2+PS3+PM2+PP3). Interestingly, rare missense variants in *SOX11* appear to increase its protein expression levels. Such observation is particularly pronounced when the missense variants locate in the HMG domain, including the Y113H variant in our patient and the previously reported variants of K50N, S60P, Y116C and P120H (Tsurusaki et al., 2014; Hempel et al., 2016). It is noticed that variants in the HMG domain more severely impair the transcriptional regulatory activity of *SOX11* than outside the HMG domain (A142G in this study and A176E in ref. 14). These results suggest that the increased protein expression caused by missense variants in HMG domain is a compensatory effect after functional loss. We here provide functional evidence to explain the phenomenon observed by Al-Jawahiri et al. in the gnomAD database that there are far fewer missense variants in the HMG box than outside the HMG domain, including the N-terminal, central, and transactivating domains in SOX11 (Al- Jawahiri et al., 2022).

In addition to the A142G variant in *SOX11*, patient 1 yet harbored microdeletion of 1q24.2-q25.1, which has been described in multiple patients (Burkardt et al., 2011; Chatron et al., 2015). The clinical features resulting from this CNV highly overlap with *SOX11*-related syndrome. For example, Chatron et al. summarized 18 patients with 1q24-q25 deletions and suggested that common clinical phenotypes include IUGR, short stature, microcephaly, delayed bone age, ID, hypertelorism, dysplastic ears, micro- and retrognathia, short hands and feet, brachydactyly, and fifth finger clinodactyly (Chatron et al., 2015). At the moment, it was difficult to determine the contribution of *SOX11* A142 variant and 1q24-q25 deletion to the phenotype of patient 1.

SOX11 is widely expressed and has an important regulatory role in tissue remodelling during early embryogenesis. Conventional knockout of *Sox11* in mice lead to birth death with severe developmental defects, involving in the malformations in brain, skeleton, eyes, spleen, lung, stomach, pancreas, and heart (Sock et al., 2004; Kato et al., 2015). In particular, *SOX11* is critical for neurogenesis, as it can drive the differentiation of embryonic stem cells into neuronal progenitor cells and the generation of mature neurons and glial cells (Bergsland et al., 2011; Kavyanifar et al., 2018). Such roles of *SOX11* in neurodevelopment have been validated in *Sox11* conditional knockout mice (Wang et al., 2013), *sox11* knockdown xenopus laevis (Hempel et al., 2016), and the *SOX11*$^{+/-}$ heterozygous human embryonic stem cell models (Turan et al., 2019). These molecular mechanism studies can well explain that patients with *SOX11* variants are mainly manifested as DD and/or ID. Consistently, patients 2 and 3 in this study had both DD and ID. However, patients with neither DD nor ID were also reported in previous studies (i.e., case 9 who harbored H109P missense variant in the HMG domain in ref. 14), which may also be true in patient 1 of this study. These indicate a high degree of heterogeneity in clinical features caused by *SOX11* variants.

Typical CSS, which is also referred to as BAFopathy, is caused by variants in the subunit of BAF complex, including *ARID1A* (OMIM#603024), *ARID1B* (OMIM#614556), *SMARCA4* (OMIM#603254), *SMARCB1* (OMIM#601607), *ARID2* (OMIM#609539), and *SMARCE1* (OMIM#603111) (Bogershausen and Wollnik, 2018). Due to the clinical phenotypic similarity, *SOX11* was considered as the new causative gene of CSS (Tsurusaki et al., 2014; Hempel et al., 2016). However, with the number of patients increased, the differences between patients with *SOX11* variation and CSS gradually became apparent. Al-Jawahiri et al. revealed that *SOX11* syndrome has distinct clinical features from *ARID1B*-related CSS. Firstly, *SOX11* mutant patients with microcephaly tend to be associated with oculomotor apraxia or abnormal eye morphology, while the *ARID1B*-related CSS patients usually have a coarse face. Secondly, deformity of the fifth fingers is an identifiable feature of CSS but do not appear to be specific to *SOX11* syndrome. Thirdly, nearly one in five of the *SOX11* mutant patients had hypogonadotrophic hypogonadism which was rare reported in CSS patients. Additionally, they uncovered a distinctive pattern of blood DNA methylation in the patients with *SOX11* variants, and thus generated a *SOX11* variation-related episignature model which was also failed to identify the BAFopathy complex samples. These results strongly suggest *SOX11* syndrome and CSS were two groups of clinically and molecularly distinct diseases (Al- Jawahiri et al., 2022). In our study, none of the patients had coarse facies and hypoplastic nails, which are the core features of CSS. Moreover, the phenotypic summary of reported patients showed only about one-third of patients had microcephaly, and one-tenth had nail dysplasia (Table 1), also suggesting differences between *SOX11*-related phenotypes and classical CSS.

In conclusion, our study identified three novel *SOX11* variants and elucidated the damage effect of two missense variants on SOX11 protein *via in vitro* experiments. This is the first report of Chinese patients with *SOX11*-related CSS9. Our study provides new evidence to support the observation that the

clinical phenotype caused by *SOX11* variants is highly heterogeneous and differs from classical CSS to some extent.

Data availability statement

The datasets for this article are not publicly available due to concerns regarding participant/patient anonymity. Requests to access the datasets should be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by Shanghai Children's Medical Center. 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

NL and JW conceived the study. YD, JC, YT, YY, and XW were responsible for recruiting patients and collection of clinical features. YD, JC, and L-NC performed the *in vitro* experiments. R-EY and TY were responsible for the sequencing work. JC and NL drafted the manuscript, tables, and figures. All authors read and approved the final manuscript.

Funding

This research was sponsored by the grant from Shanghai "Rising Stars of Medical Talents" Youth Development Program (NL), Shanghai Pujiang Program (20PJ1409900 to NL), and the Project of Clinical Research Plan of Shanghai Hospital Development Center (Grant No. SHDC2020CR3042B to JW).

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.

Acknowledgments

We are deeply grateful to the patients and their families for participating in this study.

References

Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., et al. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum. Mutat.* 39 (11), 1517–1524. doi:10.1002/humu.23626

Al-Jawahiri, R., Foroutan, A., Kerkhof, J., McConkey, H., Levy, M., Haghshenas, S., et al. (2022). SOX11 variants cause a neurodevelopmental disorder with infrequent ocular malformations and hypogonadotropic hypogonadism and with distinct DNA methylation profile. *Genet. Med.* 24, 1261–1273. doi:10.1016/j.gim.2022.02.013

Angelozzi, M., and Lefebvre, V. (2019). SOXopathies: growing family of developmental disorders due to SOX mutations. *Trends Genet.* 35 (9), 658–671. doi:10.1016/j.tig.2019.06.003

Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., Muhr, J., et al. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes. Dev.* 25 (23), 2453–2464. doi:10.1101/gad.176008.111

Bogershausen, N., and Wollnik, B. (2018). Mutational landscapes and phenotypic spectrum of SWI/SNF-related intellectual disability disorders. *Front. Mol. Neurosci.* 11, 252. doi:10.3389/fnmol.2018.00252

Bowles, J., Schepers, G., and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev. Biol.* 227 (2), 239–255. doi:10.1006/dbio.2000.9883

Burkardt, D. D., Rosenfeld, J. A., Helgeson, M. L., Angle, B., Banks, V., Smith, W. E., et al. (2011). Distinctive phenotype in 9 patients with deletion of chromosome 1q24-q25. *Am. J. Med. Genet. A* 155A (6), 1336–1351. doi:10.1002/ajmg.a.34049

Chatron, N., Haddad, V., Andrieux, J., Desir, J., Boute, O., Dieux, A., et al. (2015). Refinement of genotype-phenotype correlation in 18 patients carrying a 1q24q25 deletion. *Am. J. Med. Genet. A* 167A (5), 1008–1017. doi:10.1002/ajmg.a.36856

Cho, C. Y., Tsai, W. Y., Lee, C. T., Liu, S. Y., Huang, S. Y., Chien, Y. H., et al. (2022). Clinical and molecular features of idiopathic hypogonadotropic hypogonadism in taiwan: a single center experience. *J. Formos. Med. Assoc.* 121 (1 Pt 1), 218–226. doi:10.1016/j.jfma.2021.03.010

de la Rocha, A. M., Sampron, N., Alonso, M. M., and Matheu, A. (2014). Role of SOX family of transcription factors in central nervous system tumors. *Am. J. Cancer Res.* 4 (4), 312–324.

Diel, H., Ding, C., Grehn, F., Chronopoulos, P., Bartsch, O., Hoffmann, E. M., et al. (2021). First observation of secondary childhood glaucoma in coffin-siris syndrome: a case report and literature review. *BMC Ophthalmol.* 21 (1), 28. doi:10.1186/s12886-020-01788-0

Grimm, D., Bauer, J., Wise, P., Kruger, M., Simonsen, U., Wehland, M., et al. (2020). The role of SOX family members in solid tumours and metastasis. *Semin. Cancer Biol.* 67 (Pt 1), 122–153. doi:10.1016/j.semcancer.2019.03.004

Hempel, A., Pagnamenta, A. T., Blyth, M., Mansour, S., McConnell, V., Kou, I., et al. (2016). Deletions and de novo mutations of SOX11 are associated with a neurodevelopmental disorder with features of coffin-siris syndrome. *J. Med. Genet.* 53 (3), 152–162. doi:10.1136/jmedgenet-2015-103393

Kato, K., Bhattaram, P., Penzo-Mendez, A., Gadi, A., and Lefebvre, V. (2015). SOXC transcription factors induce cartilage growth plate formation in mouse

embryos by promoting noncanonical WNT signaling. *J. Bone Min. Res.* 30 (9), 1560–1571. doi:10.1002/jbmr.2504

Kavyanifar, A., Turan, S., and Lie, D. C. (2018). SoxC transcription factors: multifunctional regulators of neurodevelopment. *Cell Tissue Res.* 371 (1), 91–103. doi:10.1007/s00441-017-2708-7

Khan, U., Study, D., Baker, E., and Clayton-Smith, J. (2018). Observation of cleft palate in an individual with SOX11 mutation: indication of a role for SOX11 in human palatogenesis. *Cleft Palate. Craniofac. J.* 55 (3), 456–461. doi:10.1177/1055665617739312

Kosho, T., Miyake, N., and Carey, J. C. (2014). Coffin-siris syndrome and related disorders involving components of the BAF (mSWI/SNF) complex: historical review and recent advances using next generation sequencing. *Am. J. Med. Genet. C Semin. Med. Genet.* 166C (3), 241–251. doi:10.1002/ajmg.c.31415

Luo, X., Ji, X., Xie, M., Zhang, T., Wang, Y., Sun, M., et al. (2022). Advance of SOX transcription factors in hepatocellular carcinoma: From role, tumor immune relevance to targeted therapy. *Cancers (Basel)* 14 (5), 1165. doi:10.3390/cancers14051165

Okamoto, N., Ehara, E., Tsurusaki, Y., Miyake, N., and Matsumoto, N. (2018). Coffin-Siris syndrome and cardiac anomaly with a novel SOX11 mutation. *Congenit. Anom.* 58 (3), 105–107. doi:10.1111/cga.12242

Posey, J. E., Harel, T., Liu, P., Rosenfeld, J. A., James, R. A., Coban Akdemir, Z. H., et al. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. *N. Engl. J. Med.* 376 (1), 21–31. doi:10.1056/NEJMoa1516767

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 (5), 405–424. doi:10.1038/gim.2015.30

Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., et al. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the american college of medical genetics and genomics (ACMG) and the clinical genome Resource (ClinGen). *Genet. Med.* 22 (2), 245–257. doi:10.1038/s41436-019-0686-8

Sarkar, A., and Hochedlinger, K. (2013). The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* 12 (1), 15–30. doi:10.1016/j.stem.2012.12.007

Sekiguchi, F., Tsurusaki, Y., Okamoto, N., Teik, K. W., Mizuno, S., Suzumura, H., et al. (2019). Genetic abnormalities in a large cohort of coffin-siris syndrome patients. *J. Hum. Genet.* 64 (12), 1173–1186. doi:10.1038/s10038-019-0667-4

Sock, E., Rettig, S. D., Enderich, J., Bosl, M. R., Tamm, E. R., Wegner, M., et al. (2004). Gene targeting reveals a widespread role for the high-mobility-group transcription factor Sox11 in tissue remodeling. *Mol. Cell. Biol.* 24 (15), 6635–6644. doi:10.1128/MCB.24.15.6635-6644.2004

Sokpor, G., Xie, Y., Rosenbusch, J., and Tuoc, T. (2017). Chromatin remodeling BAF (SWI/SNF) complexes in neural development and disorders. *Front. Mol. Neurosci.* 10, 243. doi:10.3389/fnmol.2017.00243

Tsurusaki, Y., Koshimizu, E., Ohashi, H., Phadke, S., Kou, I., Shiina, M., et al. (2014). De novo SOX11 mutations cause coffin-siris syndrome. *Nat. Commun.* 5, 4011. doi:10.1038/ncomms5011

Turan, S., Boerstler, T., Kavyanifar, A., Loskarn, S., Reis, A., Winner, B., et al. (2019). A novel human stem cell model for coffin-siris syndrome-like syndrome reveals the importance of SOX11 dosage for neuronal differentiation and survival. *Hum. Mol. Genet.* 28 (15), 2589–2599. doi:10.1093/hmg/ddz089

Wakim, V., Nair, P., Delague, V., Bizzari, S., Al-Ali, M. T., Castro, C., et al. (2021). SOX11-related syndrome: report on a new case and review. *Clin. Dysmorphol.* 30 (1), 44–49. doi:10.1097/MCD.0000000000000348

Wang, J., Yu, T., Wang, Z., Ohte, S., Yao, R. E., Zheng, Z., et al. (2016). A new subtype of multiple synostoses syndrome is caused by a mutation in GDF6 that decreases its sensitivity to noggin and enhances its potency as a BMP signal. *J. Bone Min. Res.* 31 (4), 882–889. doi:10.1002/jbmr.2761

Wang, J., Zhang, W., Jiang, H., and Wu, B. L. (2014). Mutations in HFM1 in recessive primary ovarian insufficiency. *N. Engl. J. Med.* 370 (10), 972–974. doi:10.1056/NEJMc1310150

Wang, Y., Lin, L., Lai, H., Parada, L. F., and Lei, L. (2013). Transcription factor Sox11 is essential for both embryonic and adult neurogenesis. *Dev. Dyn.* 242 (6), 638–653. doi:10.1002/dvdy.23962

Yao, R., Yu, T., Qing, Y., Wang, J., and Shen, Y. (2019). Evaluation of copy number variant detection from panel-based next-generation sequencing data. *Mol. Genet. Genomic Med.* 7 (1), e00513. doi:10.1002/mgg3.513

Yao, R., Zhang, C., Yu, T., Li, N., Hu, X., Wang, X., et al. (2017). Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. *Mol. Cytogenet.* 10, 30. doi:10.1186/s13039-017-0333-5

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 03 August 2022
DOI 10.3389/fnins.2022.956429

OPEN ACCESS

EDITED BY
Tianyun Wang,
Peking University, China

REVIEWED BY
Michael Zech,
Technical University of Munich, Germany
Xiaoli Chen,
Capital Institute of Pediatrics, China

*CORRESPONDENCE
Hong Zhang
zhangh@shchildren.com.cn
Shengnan Wu
wushengnan@shchildren.com.cn

†These authors have contributed equally to this work

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

RECEIVED 30 May 2022
ACCEPTED 15 July 2022
PUBLISHED 03 August 2022

CITATION
Song X, Xu W, Xiao M, Lu Y, Lan X, Tang X, Xu N, Yu G, Zhang H and Wu S (2022) Two novel heterozygous truncating variants in *NR4A2* identified in patients with neurodevelopmental disorder and brief literature review. *Front. Neurosci.* 16:956429. doi: 10.3389/fnins.2022.956429

Two novel heterozygous truncating variants in *NR4A2* identified in patients with neurodevelopmental disorder and brief literature review

Xiaozhen Song[1†], Wuhen Xu[1†], Man Xiao[1], Yanfen Lu[2], Xiaoping Lan[1], Xiaojun Tang[1], Nanjie Xu[3,4], Guangjun Yu[5], Hong Zhang[1*] and Shengnan Wu[1*]

[1]Molecular Diagnostic Laboratory, Department of Clinical Laboratory, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [2]Department of Neurology, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [3]Research Center of Translational Medicine, Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [4]Department of Anatomy and Physiology, School of Medicine, Shanghai Jiao Tong University, Shanghai, China, [5]Shanghai Children's Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China

Pathogenic variants in the nuclear receptor superfamily 4 group A member 2 (*NR4A2*) cause an autosomal dominant neurodevelopmental disorder with or without seizures. Here, we described two patients presenting with developmental delay, language impairment, and attention-deficit hyperactivity disorder. Trio-based whole exome sequencing revealed two novel heterozygous variants, c.1541-2A > C and c.915C > A, in *NR4A2*. Both variants were identified as *de novo* and confirmed by Sanger sequencing. *In vitro* functional analyses were performed to assess their effects on expression of mRNA or protein. The canonical splicing variant c.1541-2A > C caused aberrant splicing, leading to the retention of intron 7 and a truncated protein due to an early termination codon within intron 7 with decreased protein expression, while the variant c.915C > A was shown to result in a shorter protein with increased expression level unexpectedly. The clinical and genetic characteristics of the previously published patients were briefly reviewed for highlighting the potential link between mutations and phenotypes. Our research further confirms that *NR4A2* is a disease-causing gene of neurodevelopmental disorders and suggests alterations in different domains of *NR4A2* cause various severity of symptoms.

KEYWORDS

***NR4A2*, truncating, neurodevelopmental disorder, intellectual disability, language impairment, attention deficit**

Introduction

The extensive clinical use of exome sequencing has enabled increasing recognition of the genetics and pathophysiology of neurodevelopmental disorders (NDD), in which *de novo* variants play major roles (Retterer et al., 2016). The nuclear receptor superfamily 4 group A member 2 (*NR4A2*) gene, also known as nuclear receptor-related 1 (*Nurr1*), is located on chromosome 2q24.1 and encodes an orphan nuclear receptor that belongs to the nuclear steroid-thyroid hormone and retinoid receptor superfamily (Zetterstrom et al., 1997; Wang et al., 2003). *NR4A2* is widely expressed throughout the brain and in other tissues and organs, such as the cortex, hippocampus, and peripheral blood. The encoded protein is a transcription factor with essential regulatory functions in the central nervous system (CNS), including differentiation, migration, maturation, and maintenance of mesencephalic dopaminergic neurons, which are related to memory and learning. It has been shown that *NR4A2* knockout mice are unable to produce midbrain dopaminergic neurons and die shortly after birth (Safe et al., 2016).

In human pathology, earlier studies have indicated *NR4A2* may be a potential susceptibility gene for Parkinson's disease and schizophrenia by case-control association studies. The G insertion at IVS6 + 18 of *NR4A2* was observed with higher frequency in Parkinson's patients than in healthy individuals (Zheng et al., 2003; Grimes et al., 2006; Chen et al., 2007; Liu et al., 2013). Le et al. (2003) reported two mutations −291Tdel and −245T > G located in the non-coding exon 1 of the *NR4A2* gene in 10 of 107 individuals with familial Parkinson's disease, but not in unaffected controls. Besides mutations in non-coding or intronic regions, one rare missense mutation (c.709C > G) in exon 3 of *NR4A2* was detected in a patient with Parkinson's disease (Grimes et al., 2006). Additionally, missense mutations (c.289A > G and c.308A > G) in exon 3 of *NR4A2* were identified in schizophrenic patients, and a small deletion mutation (c.364_366delTAC) in an individual with manic-depressive disorder (Buervenich et al., 2000). Subsequently, *de novo* heterozygous deletions in 2q24.1 with the minimum of overlap being *NR4A2*, yielded by the application of array comparative genomic hybridization (array-CGH) and Next-Generation Sequencing (NGS), have been described in patients with intellectual disability (ID), language impairment and autism spectrum disorder (ASD), suggesting the role of *NR4A2* haploinsufficiency in NDD (Barge-Schaapveld et al., 2013; Leppa et al., 2016; Reuter et al., 2017; Shimojima et al., 2017; Levy et al., 2018). However, *NR4A2* has been considered as a monogenic disease-causing gene due to the recent genetic findings. The *NR4A2* loss-of-function variant c.326dupA was identified in a patient with epilepsy, ID, and language impairment (Ramos et al., 2019). The clinical features of mild intellectual disability at childhood and dystonia parkinsonism in early adulthood observed in two patients have been attributed to the frameshift variants in *NR4A2* (c.326dupA and c.881dupA) (Wirth et al., 2020). Eight missense and loss-of-function variants within *NR4A2* identified in a larger cohort consolidated the pivotal role of *NR4A2* in NDD (Singh et al., 2020). To date, all reported patients carrying *NR4A2* likely pathogenic/pathogenic (LP/P) variants presented with certain consistent neurodevelopmental features, including varying levels (mild to severe) of ID/developmental delay (DD), language impairment, behavioral problems, movement disorders, and epilepsy.

Here, we described two novel *NR4A2* truncating variants in two unrelated Chinese patients mainly presenting with mild ID/DD, language impairment, and attention deficit, without epilepsy. The results of our study provided more evidence for the role of *NR4A2* haploinsufficiency in NDD and updated the mutation spectrum of *NR4A2*. Meanwhile, we briefly reviewed the *NR4A2* variants and clinical manifestations and found that loss-of-function variants and missense variants occurring in the key domains of NR4A2 were generally intolerant.

Materials and methods

Patients

Two patients from unrelated families enrolled at Shanghai Children's Hospital were included in this study. Peripheral blood samples and clinical data from each patient and the parents were collected. Written informed consent was obtained from the parents for the use of clinical and genetic information. This study was approved by the Ethics Committee of Shanghai Children's Hospital (2020R007-F01).

Genetic analysis

To elucidate the cause of the disorder, we performed child–parent trio whole-exome sequencing (TWES) on the two patients and their parents due to the known efficiency of TWES for NDD. Genomic DNA for TWES and subsequent Sanger sequencing was extracted from peripheral blood using a DNA Blood Mini Kit (QIAGEN, Germany), following the manufacturer's protocol. An IDTxGen® Exome Research Panel (IDT, United States) was used to capture the exons, and HiseqX10 (Illumina, United States) was used to sequence the DNA fragments. FASTQ data analysis was conducted as described in previous studies (Xiaozhen et al., 2021). The candidate pathogenic genes were confirmed by Sanger sequencing. The pathogenicity of the variations was evaluated according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology

(ACMG/AMP) guidelines and ClinGen specifications (Richards et al., 2015; Zhang et al., 2020).

Plasmids construction and functional studies of the variants in NR4A2

Wild type (WT) and mutant human *NR4A2* (transcript NM_006186.3) expression plasmids, labeled with an EGFP tag at the N-terminal, were synthesized and cloned into the pCMV-EGFP-C1 vector. The *NR4A2* mutation c.915C > A (p.Cys305*) was introduced into the WT isoform (named NR4A2-WT) using a site-directed mutagenesis kit (Santa Clara, CA, United States). To determine the effects of variant c.1541-2A > C on splicing, WT (c.1541-2A) or mutant (c.1541-2C) intron 7 containing 149 nucleobases was introduced into NA4A2-WT isoform between exon 7 and exon 8. The successfully constructed recombinant plasmid products were named NR4A2-AG and NR4A2-CG (**Supplementary Figure 1**), respectively. The primers used in the study are shown in **Supplementary Table 1**. Human embryonic kidney (HEK)-293T cells were transfected with different expression plasmids. The pCMV-EGFP-C1 vector expressing fluorescence was used as a transfection marker.

Forty-eight hours after transfection, total RNA was isolated from cultured cells using TRIzol Reagent (Thermo Fisher Scientific, Waltham, MA, United States), and 1 μg of RNA was converted to cDNA by reverse transcription using EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China). The PCR amplification of cDNA, including the target region spanning exons 6–8, was performed using DNA Polymerase (TransGen Biotech, Beijing, China). The primers were as follows: F-TGGAGATGACACCCAGCATA; R-GTGGCACCAAGTCTTCCAAT. Agarose gel electrophoresis (1%) and PCR products' sequencing were performed to evaluate the products' length and sequence.

Total protein was extracted using a Protein Extraction Kit (Sangon Biotech, Shanghai, China). Subsequently, 25 μg of total protein was electrophoresed on 8% sodium dodecyl sulfate polyacrylamide gels (SDS-PAGE) and then transferred to nitrocellulose membranes (Merck Millipore, Darmstadt, Germany). After blocking with 5% skim milk, the membranes were incubated with specific primary antibodies overnight at 4°C, followed by incubation with HRP-conjugated anti-mouse (ZSGB-BIO, Beijing, China) secondary antibodies at room temperature for 1 h. Finally, the bands were detected using ECL Prime Western Blotting Reagent (GE Healthcare, Buckinghamshire, United Kingdom). The experiments were repeated in triplicate after determining the optimal working conditions. The primary antibodies used were rabbit anti-GAPDH (Cell signaling technology; 1:3,000) and mouse anti-Nurr1binding to amino acids 2–99 of human Nurr1 (Abcam; 1:500).

Cell culture and plasmid transfection

HEK-293T cells (purchased from the National Infrastructure of Cell Line Resource, China) were plated at a density of 5×10^5 cells into six-well culture plates. The cells were maintained in Dulbecco's modified eagle's medium (DMEM; Gibco, United States), supplemented with 1% penicillin/streptomycin (Gibco, United States) and 10% fetal bovine serum (FBS; Gibco, United States), at 37°C in a humidified atmosphere incubator with 5% CO_2, and passaged every 2–3 days. Transient transfection with WT or variant *NR4A2* plasmids was performed using Lipofectamine 2000 Transfection Reagent (Thermo Fischer Scientific, United States) according to the manufacturer's recommendations. Briefly, the day before transfection, cells were plated at a density of 5×10^5 cells per well; 24 h later, 2 μg/well plasmid DNA was added, diluted with 100 μl Opti-MEM (Gibco, United States), and mixed with 6 μL of Lipofectamine 2000.

Results

Clinical description

The patients recruited in this study were the only children of their families and were born at full term to non-consanguineous parents with uneventful pregnancies and deliveries. Physical examinations found no dysmorphic features. Neither of them had a remarkable family history of neurologic diseases.

Patient 1 (P1) was an 11-year-old boy at the time of his visit to the clinic. His head circumference and length at birth were within the normal range, and his birth weight was in the 90th percentile. Motor development was in normal milestones with walking independently at 14 months. Significant delay of speech development was observed at the age of 6 years. He presented with a lack of expressive language and poor oral communication with other children. He showed behavioral abnormalities, including attention deficit hyperactivity disorder (ADHD), learning difficulties, tantrums, and aggression. He had mild hypotonia. Both electroencephalogram (EEG) activity and brain MRI revealed no abnormalities. He had no history of seizures.

Patient 2 (P2) was a 12-year-old boy. His developmental milestones were globally delayed, and his fine motor development was poor. He displayed ADHD, learning difficulties, and mild hypotonia. His brain MRI was normal. He had normal EEG and no history of seizures.

Genetic results

Child-parent trio-based whole-exome sequencing revealed two novel *NR4A2* variants, including a heterozygous canonical splice-acceptor variant in intron 7, NM_006186.3: c.1541-2A > C in patient 1 (Figure 1A), and a heterozygous nonsense variant within exon 4, NM_006186.3: c.915C > A (p.Cys305*) in patient 2 (Figure 1B), respectively; both variants occurred *de novo* (Figure 1). Neither of the variants was present in the Genome Aggregation Database (gnomAD). The variant c.1541-2A > C detected in P1 was predicted by *in silico* tools to affect splicing and potentially lead to a deficiency in NR4A2 protein. The nonsense variant, c.915C > A, introducing a premature termination, was predicted to lead to either a truncated non-functional protein or a nonsense-mediated mRNA decay (NMD). No other potentially LP/P variants were found.

Functional assay for the c.1541-2A > C and c.915C > A variants in NR4A2

The *NR4A2* gene contains eight exons in total, and the c.1541-2A > C variant is located at the splicing acceptor site of intron 7. Due to the unavailability of fresh blood or cell samples from patient 1 with c.1541-2A > C variant, we constructed an *NR4A2* recombinant plasmid, introducing intron 7 of *NA4A2* between exon 7 and exon 8 in *NR4A2* WT isoform. The recombinant plasmids

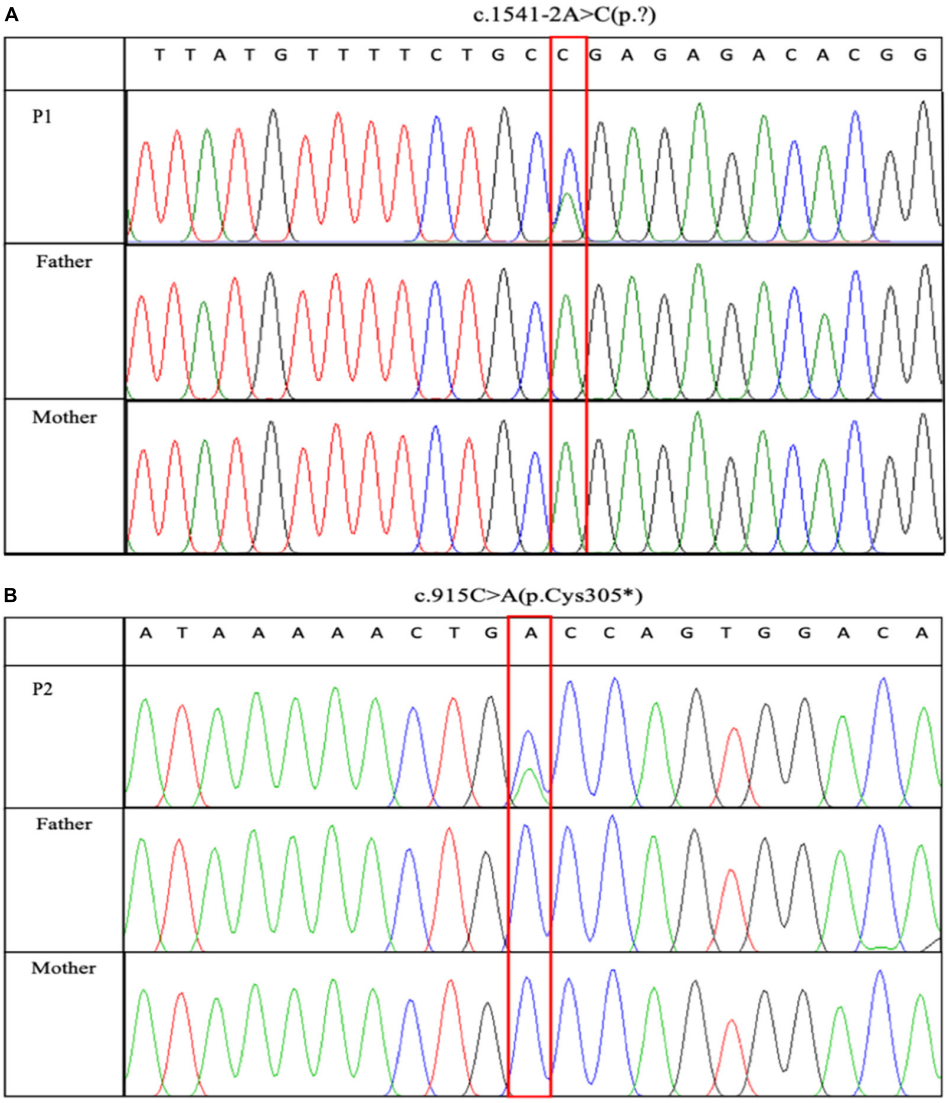


FIGURE 1
Sanger sequencing maps of two pathogenic variants. c.1541-2A > C and c.915C > A were identified in P1 and P2, respectively. Sanger sequencing verified that the two variants occurred *de novo*. **(A)** c.1541-2A > C was identified in P1. **(B)** c.915C > A was identified in P2.

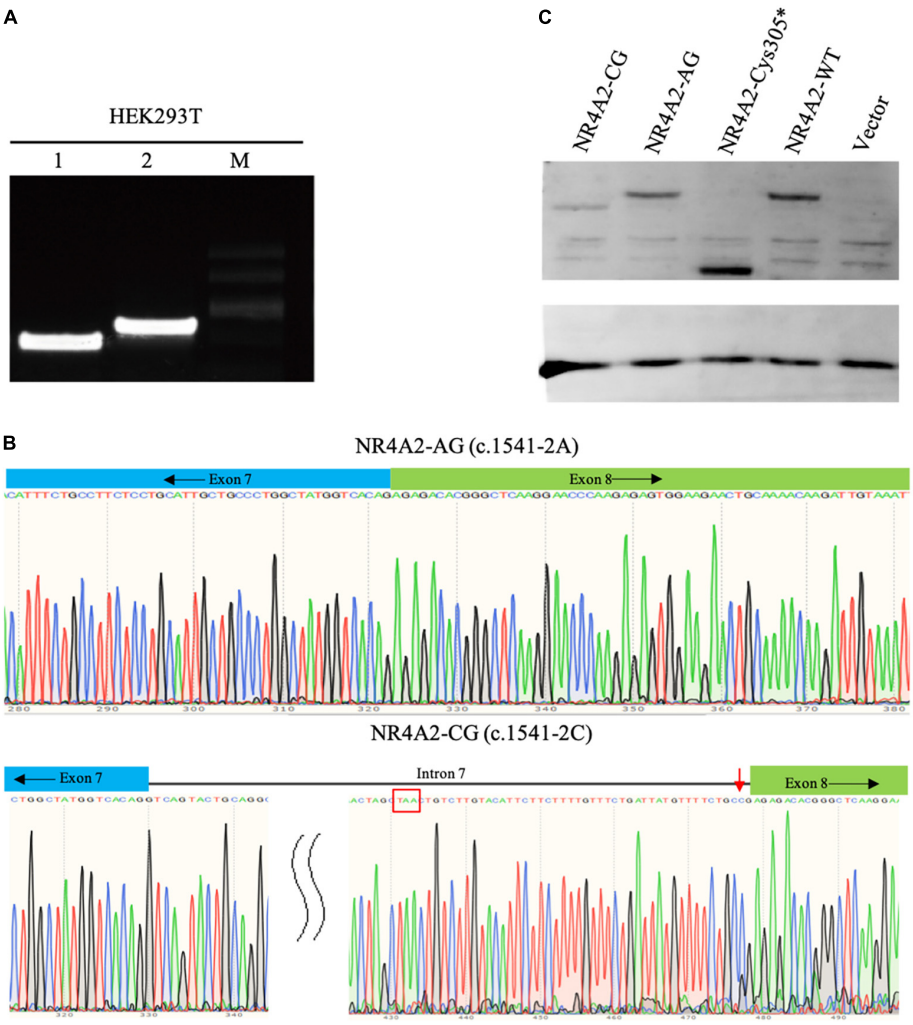


FIGURE 2
Functional analyses of two *NR4A2* variants: c.1541-2A > C and c.915C > A. **(A)** Agarose gel electrophoresis analysis of the RT-PCR amplification products: 1-NR4A2-AG construct; 2-NR4A2-CG construct; M-marker. **(B)** Sequence analyses of the PCR amplification products of the target fragment of NR4A2-AG and NR4A2-CG. NR4A2-CG resulted in the retention of intron 7 and an early termination codon TAA (red box). **(C)** Protein expression analyses of NR4A2 with c.915C > A or c.1541-2A > C variants. Both c.915C > A and c.1541-2A > C caused a truncated NR4A2 protein.

consisting of exons 1–7 and intron 7, followed by exon 8, were successfully constructed and were confirmed by plasmid sequencing (Supplementary Figure 1). The functional analyses demonstrated that the mutant NR4A2-CG resulted in a longer PCR product than the wild-type NR4A2-AG, which was confirmed by sequencing (Figure 2A). We found that the NR4A2-CG caused aberrant splicing and produced a longer transcript with the retention of intron 7 (Figure 2B). Additionally, an early termination codon was detected in intron 7 (Figure 2B), which may finally produce a premature truncated protein.

To verify the effect of two detected variants c.1541-2A > C and c.915C > A (p.Cys305*) on protein expression, western blot studies were performed. The protein analyses revealed different molecular weight patterns in the cells transfected with different plasmids. The c.1541-2A product (NR4A2-AG) and NR4A2-WT produced the protein with the same size, while both variants c.915C > A and c.1541-2A > C ended in smaller molecular weight pattern compared to that of the wild type (Figure 2C). This indicated that the two variants produced truncated NR4A2 proteins, resulting in the loss of key protein domains, which could potentially cause the loss of function (LoF) of NR4A2. The variant c.1541-2A > C caused decreased protein expression compared to the wild type, as expected. Unexpectedly, we found that the nonsense variant c.915C > A (p.Cys305*) led to higher protein expression than the wild type, which could be attributed to the increased stability and lower degradation of the truncated protein.

Brief literature review about NR4A2 variants and clinical manifestations

To date, sixteen patients carrying *NR4A2* LP/P single nucleotide variants (SNVs), or small insertion-deletion (InDel) have been reported (Vissers et al., 2017; Ramos et al., 2019; Singh et al., 2020; Wirth et al., 2020; Jesus et al., 2021; Winter et al., 2021). The main clinical features and genetic information of patients with *NR4A2* LP/P variants are summarized in **Table 1**. The main manifestations are characterized by ID, language impairment, seizures, psychobehavioral problems, and movement disorders. No significant difference was observed in age of onset and phenotypic severity between patients with LP/P missense variants and loss-of-function variants. The majority of the patients with *NR4A2* variants presented with ID and language impairment ranging from mild to severe. A schematic view of the distribution of LP/P variants in NR4A2 protein domain is shown in **Figure 3A**, including eight missense variants and seven loss-of-function variants. NR4A2 protein consists of an N-terminal, a DNA binding domain (DBD), a ligand binding domain (LBD), and a linker region between DBD and LBD. DBD domain contains two zinc fingers. All the pathogenic missense variants were found to occur in the zinc finger domain with the exception of p.Asp392Gly, which was located in the linker region adjacent to LBD. Loss-of-function variants have been observed to occur in any region of NR4A2 protein. We performed MetaDome analysis which using data from gnomAD and ClinVar for evaluating the tolerance of *NR4A2* missense variants[1] (Wiel et al., 2019). The *NR4A2* tolerance landscape has shown that NR4A2 protein contains two key homologous domains, Pkinase Pfam protein domain PF00105 and PF00104. The PF00105 (consistent with zinc finger domain in **Figure 3A**) is intolerant compared to other regions in NR4A2 protein, which helps to understand why missense variants occurring in this region are more likely damaging (**Figure 3B**).

Discussion

NR4A2 is ubiquitously expressed in subcellular regions of the human brain and is particularly prominent in the hippocampus, on which memory-inducing activities, such as language development and learning, are dependent (Abrahams et al., 2007; Hawk and Abel, 2011). Microdeletions encompassing *NR4A2* and point mutations resulting in haploinsufficiency in *NR4A2* have been previously reported in individuals with ID and language impairments, suggesting the role of *NR4A2* in human NDD (Reuter et al., 2017; Levy et al., 2018; Ramos et al., 2019). A recent genetic study with larger cohort identified eight *de novo* variants of *NR4A2* and a deletion containing *NR4A2* in nine patients with NDD and epilepsy, further strengthening the association of *NR4A2* heterozygous variants with NDD (Singh et al., 2020). Here, we reported two novel variants occurring *de novo* in *NR4A2*, c.1541-2A > C and c.915C > A (p.Cys305*), in two patients presenting with mild intellectual disability, language impairment, and attention deficit disorder. Our findings in the patients of Chinese origin provided additional evidence supporting *NR4A2* as a disease-causing gene of NDD. The first patient also presented with behavioral problems, including tantrums, aggression, and hyperactivity, which is consistent with the phenotypes related to *NR4A2* haploinsufficiency described previously (Singh et al., 2020). Epilepsy, a commonly phenotype described in *NR4A2*-associated patients, was not observed in our patients. In previous studies, variants in *NR4A2* have also been associated with early-onset dystonia parkinsonism (Wirth et al., 2020; Jesus et al., 2021; Winter et al., 2021). Although our pediatric patients have not yet presented with movement disorders, these phenotypes should be monitored carefully in future clinical follow-ups.

The variant c.1541-2A > C in *NR4A2* was predicted to lead to aberrant splicing. The mechanism by which it affected *NR4A2* pre-mRNA splicing was confirmed by *in vitro* analysis. The PCR and Sanger sequencing results showed that the c.1541-2A > C variant abolished the splice acceptor site, retaining intron 7 and leading to a premature termination at 104 bp of intron 7, resulting in a truncated protein with 547 amino acid residues compared to the normal protein with 598 amino acid residues. This variant was predicted to affect the expression of the NR4A2 protein, which was partially confirmed by our experiments.

Our experimental studies using HEK-293T cells demonstrated that the variant c.1541-2A > C resulted in a smaller NR4A2 protein with decreased expression which NMD might underlie. The variant c.915C > A produced a significantly truncated protein with increased expression level unexpectedly, suggesting a dominant-negative effect on wild-type protein beyond haploinsufficiency. However, the NR4A2 protein, as an orphan nuclear receptor, has two main domains: a DBD and an LBD (de Vera et al., 2019). The variant c.915C > A was located in the DBD and resulted in the loss of DBD and the downstream domain LBD, indicating that the truncating variant potentially causes loss of the ability of NR4A2 protein binding DNA elements and subsequent functions.

Thus far, a total of 636 variants of *NR4A2* have been recorded in the gnomAD database,[2] including nine heterozygous variants with loss of function mutations (two stop-gaining, four frameshift, and three splicing), presenting with extremely low frequencies in the assumed healthy population, ranging from 3.98×10^{-6} to 3.07×10^{-5} with no presence of homozygotes. The pLI (probability of

1 https://stuart.radboudumc.nl/metadome

2 https://gnomad.broadinstitute.org/

TABLE 1 Genetic and clinical features of patients with *NR4A2* LP/P variants.

Patient	References	Variant	Variant type	Sex/Age (Y)	Seizures	Intellectual disability	Language impairment	Psychobehavioral problems	Movement disorders
1	Vissers et al. (2017)	c.920T > G, p.Val307Gly	Missense	F/15	No	Psychomotor retardation	NA	NA	No
2	Ramos et al. (2019)	c.326dupA, p.Ser110Valfs*2	Frameshift	NA	Yes	Mild	Yes	NA	No
3	Wirth et al. (2020)	c.326dupA, p.Ser110Valfs*2	Frameshift	M/29	Yes	Mild	Yes	No	Adult-onset dystonia-parkinsonism
4	Wirth et al. (2020)	c.881dupA, p.Asn294fs	Frameshift	F/57	No	Mild	Yes	No	Adult-onset dystonia-parkinsonism
5	Singh et al. (2020)	c.839G > A, p.Cys280Tyr	Missense	F/15	Yes	Severe	NA	Autism	No
6	Singh et al. (2020)	c.865-1_865delGCinsAAAAAGGAGT, p.?	Splicing	M/12	Yes	Mild	Yes	Hyperactivity, anxiety	Joint hypermobility in the setting of hypotonia
7	Singh et al. (2020)	c.914G > A, p.Cys305Tyr	Missense	F/9	Yes	Mild to moderate	NA	NA	Dystonia, choreathetoid movements, ataxic gait
8	Singh et al. (2020)	c.1175A > G, p.Asp392Gly	Missense	F/3	Yes	Severe	NA	No	Dystonia
9	Singh et al. (2020)	c.1576G > T, p.Glu526*	Non-sense	M/5	No	Mild	Yes	Attachment disorder, hyposensitivity	No
10	Singh et al. (2020)	c.325dupC, p.Gln109Profs*3	Frameshift	M/2	Yes	NA	Yes	Sensory sensitivity	No
11	Singh et al. (2020)	c.857T > C, p.Phe286Ser	Missense	F/4	No	Moderate	Yes	No	No
12	Singh et al. (2020)	c.968G > T, p.Cys323Phe	Missense	F/19	No	Moderate to severe	Yes	No	No
13	Jesus et al. (2021)	c.956G > A, p.Arg319Gln	Missense	M/30	No	Mild	Yes	Attention deficit	Dystonia-parkinsonism
14	Winter et al. (2021)	c.863A > G, p.Lys288Arg	Missense	M/2.5	NA	Mild to moderate	Yes	No	Early-onset dystonia
15	Present study	c.915C > A, p.Cys305*	Non-sense	M/12	No	Mild	Yes	Attention deficit, hyperactivity, aggression	No
16	Present study	c.1541-2A > C, p.?	Splicing	M/11	No	Mild	Yes	Attention deficit	No
Total					7/16	15/16	12/16	7/16	7/16

F, female; M, male; Y, year; NA, not available.

10.3389/fnins.2022.956429

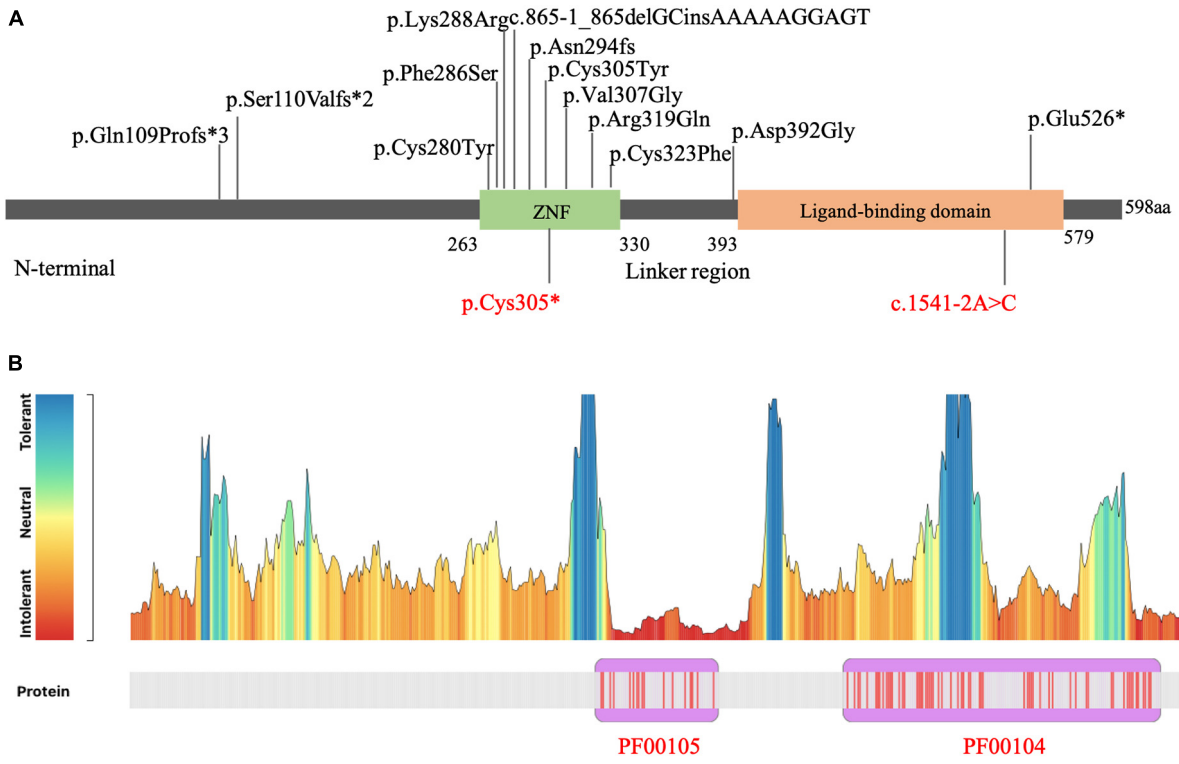


FIGURE 3
Distribution of *NR4A2* variants. **(A)** Schematic view of the domain distribution of variants in *NR4A2*. The variants identified in the present study (in red) and previous studies (in black). **(B)** *NR4A2* missense variation tolerance landscape. The tolerance landscape depicts a missense over synonymous ratio which are annotated from gnomAD dataset. Pkinase Pfam protein domains are shown in purple. PF00105 is clearly seen as intolerant to missense variation compared to other parts of NR4A2.

loss of function intolerance) and LOEUF (loss-of-function observed/expected upper bound fraction) scores reported for *NR4A2* were 1.0 and 0, respectively, suggestive of the intolerance of haploinsufficiency due to truncating variants and whole gene deletion, which have been shown to cause NDD. The majority of truncating mutations potentially cause loss of the encoded protein mainly by NMD. *NR4A2* deficiency was previously associated with impaired dopaminergic function and increased the vulnerability of midbrain dopaminergic neurons (Jankovic et al., 2005). The heterozygous deletion of *NR4A2* in mice led to a progressive loss of dopaminergic neurons in the substantia nigra (Jiang et al., 2005) and the occurrence of parkinsonian features (Zhang et al., 2012). The truncated NR4A2 protein perhaps causes NDD by damaging midbrain dopaminergic neurons.

Meanwhile, *NR4A2* also has a substantial Z score for missense variants (Z = 2.24), indicating a certain degree of intolerance to missense mutations. *NR4A2* missense variants considered to cause NDD occurred in DBD and LBD regions, while others occurring in N-terminal domain (c.289A > G, c.308A > G and c.374C > G) tend to be associated with susceptibility of Parkinson's disease or schizophrenia, in which a different mechanism might play a role. NR4A2 is known to function as a homodimer (Philips et al., 1997). Missense variants located at the functional domains might cause loss of NR4A2 via dominant negative effect. In addition, gain of function resulted from missense variants of *NR4A2* could not be excluded. Further studies to clarify various disease-causing mechanism contributing to different clinical phenotypes are needed.

Conclusion

In summary, we confirmed that both splicing site variant c.1541-2A > C and nonsense variant c.915C > A detected in our patients caused the truncation of the protein based on functional assays *in vitro*. To clarify the correlations between *NR4A2* variants and clinical presentations, which, thus far, are mainly relevant for NDD, additional patients of different ages and ethnic backgrounds need to be described. Our findings confirmed that *NR4A2* haploinsufficiency was responsible for certain clinical features in patients and proposed the potential appearance of dystonia and/or parkinsonism related to *NR4A2* which should be attached importance to in the follow-ups.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Shanghai Children's Hospital. 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 minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

Author contributions

XS and SW designed the study. XS and MX performed *in vitro* studies. WX and YL collected and evaluated clinical presentations. XL and XT performed NGS and Sanger sequencing. NX and GY provided critical feedback and helped conduct research. XS and WX drafted the manuscript. HZ and SW supervised the study and revised the manuscript. All authors discussed the final results, critically reviewed the manuscript, read, and agreed to the published version of the manuscript.

Funding

This study was funded by the Shanghai Municipal Science and Technology Committee (grant no. 19411965000), the Shanghai Jiao Tong University "Star of Jiao Tong University" Medical-Engineering Cross Research Fund (grant no. YG2019QNB01), Clinical Research Plan of SHDC (grant no. SHDC2020CR1047B), the Shanghai Municipal Key Clinical Specialty (grant no. shslczdzk06902), the Shanghai Children's Hospital Funding (grant no. 2021YGZQ07), Shanghai Jiao Tong University Genetic Development and Psychoneurotic Diseases Research Fund (grant no. 2020GDND02), and Joint Research Initiative of the Shanghai Jiao Tong University, School of Medicine (grant no. 18XD1403200).

Acknowledgments

We would like to thank the patients and their parents who participated in the study, as well as Chengkan Du for his technical assistance that contributed to the success of this project.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.956429/full#supplementary-material

References

Abrahams, B. S., Tentler, D., Perederiy, J. V., Oldham, M. C., Coppola, G., and Geschwind, D. H. (2007). Genome-wide analyses of human perisylvian cerebral cortical patterning. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17849–17854. doi: 10.1073/pnas.0706128104

Barge-Schaapveld, D. Q., Ofman, R., Knegt, A. C., Alders, M., Hohne, W., Kemp, S., et al. (2013). Intellectual disability and hemizygous GPD2 mutation. *Am. J. Med. Genet A* 161A, 1044–1050. doi: 10.1002/ajmg.a.35873

Buervenich, S., Carmine, A., Arvidsson, M., Xiang, F., Zhang, Z., Sydow, O., et al. (2000). NURR1 mutations in cases of schizophrenia and manic-depressive disorder. *Am. J. Med. Genet.* 96, 808–813. doi: 10.1002/1096-8628(20001204)96:63.0.co;2-e

Chen, C. M., Chen, I. C., Chang, K. H., Chen, Y. C., Lyu, R. K., Liu, Y. T., et al. (2007). Nuclear receptor NR4A2 IVS6 +18insG and brain derived neurotrophic factor (BDNF) V66M polymorphisms and risk of taiwanese Parkinson's disease. *Am. J. Med. Genet B Neuropsychiatr. Genet* 144B, 458–462. doi: 10.1002/ajmg.b.30476

de Vera, I. M. S., Munoz-Tello, P., Zheng, J., Dharmarajan, V., Marciano, D. P., Matta-Camacho, E., et al. (2019). Defining a canonical ligand-binding pocket in the orphan nuclear receptor Nurr1. *Structure* 27:e65. doi: 10.1016/j.str.2018.10.002

Grimes, D. A., Han, F., Panisset, M., Racacho, L., Xiao, F., Zou, R., et al. (2006). Translated mutation in the Nurr1 gene as a cause for Parkinson's disease. *Mov. Disord.* 21, 906–909. doi: 10.1002/mds.20820

Hawk, J. D., and Abel, T. (2011). The role of NR4A transcription factors in memory formation. *Brain Res. Bull.* 85, 21–29. doi: 10.1016/j.brainresbull.2011.02.001

Jankovic, J., Chen, S., and Le, W. D. (2005). The role of Nurr1 in the development of dopaminergic neurons and Parkinson's disease. *Prog. Neurobiol.* 77, 128–138. doi: 10.1016/j.pneurobio.2005.09.001

Jesus, S., Hinarejos, I., Carrillo, F., Martinez-Rubio, D., Macias-Garcia, D., Sanchez-Monteagudo, A., et al. (2021). NR4A2 mutations can cause intellectual disability and language impairment with persistent dystonia-parkinsonism. *Neurol. Genet.* 7:e543. doi: 10.1212/NXG.0000000000000543

Jiang, C., Wan, X., He, Y., Pan, T., Jankovic, J., and Le, W. (2005). Age-dependent dopaminergic dysfunction in Nurr1 knockout mice. *Exp. Neurol.* 191, 154–162. doi: 10.1016/j.expneurol.2004.08.035

Le, W. D., Xu, P., Jankovic, J., Jiang, H., Appel, S. H., Smith, R. G., et al. (2003). Mutations in NR4A2 associated with familial Parkinson disease. *Nat. Genet.* 33, 85–89. doi: 10.1038/ng1066

Leppa, V. M., Kravitz, S. N., Martin, C. L., Andrieux, J., Le Caignec, C., Martin-Coignard, D., et al. (2016). Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. *Am. J. Hum. Genet.* 99, 540–554. doi: 10.1016/j.ajhg.2016.06.036

Levy, J., Grotto, S., Mignot, C., Maruani, A., Delahaye-Duriez, A., Benzacken, B., et al. (2018). NR4A2 haploinsufficiency is associated with intellectual disability and autism spectrum disorder. *Clin. Genet* 94, 264–268. doi: 10.1111/cge.13383

Liu, H., Tao, Q., Deng, H., Ming, M., Ding, Y., Xu, P., et al. (2013). Genetic analysis of NR4A2 gene in a large population of han Chinese patients with Parkinson's disease. *Eur. J. Neurol.* 20, 584–587. doi: 10.1111/j.1468-1331.2012.03824.x

Philips, A., Lesage, S., Gingras, R., Maira, M. H., Gauthier, Y., Hugo, P., et al. (1997). Novel dimeric Nur77 signaling mechanism in endocrine and lymphoid cells. *Mol. Cell. Biol.* 17, 5946–5951. doi: 10.1128/MCB.17.10.5946

Ramos, L. L. P., Monteiro, F. P., Sampaio, L. P. B., Costa, L. A., Ribeiro, M. D. O., Freitas, E. L., et al. (2019). Heterozygous loss of function of NR4A2 is associated with intellectual deficiency, rolandic epilepsy, and language impairment. *Clin. Case Rep.* 7, 1582–1584. doi: 10.1002/ccr3.2260

Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., et al. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet Med.* 18, 696–704. doi: 10.1038/gim.2015.148

Reuter, M. S., Krumbiegel, M., Schluter, G., Ekici, A. B., Reis, A., and Zweier, C. (2017). Haploinsufficiency of NR4A2 is associated with a neurodevelopmental phenotype with prominent language impairment. *Am. J. Med. Genet A* 173, 2231–2234. doi: 10.1002/ajmg.a.38288

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

Safe, S., Jin, U. H., Morpurgo, B., Abudayyeh, A., Singh, M., and Tjalkens, R. B. (2016). Nuclear receptor 4A (NR4A) family - orphans no more. *J. Steroid Biochem. Mol. Biol.* 157, 48–60. doi: 10.1016/j.jsbmb.2015.04.016

Shimojima, K., Okamoto, N., and Yamamoto, T. (2017). Possible genes responsible for developmental delay observed in patients with rare 2q23q24 microdeletion syndrome: literature review and description of an additional patient. *Cong. Anom (Kyoto)* 57, 109–113. doi: 10.1111/cga.12205

Singh, S., Gupta, A., Zech, M., Sigafoos, A. N., Clark, K. J., Dincer, Y., et al. (2020). De novo variants of NR4A2 are associated with neurodevelopmental disorder and epilepsy. *Genet Med.* 22, 1413–1417. doi: 10.1038/s41436-020-0815-4

Vissers, L., van Nimwegen, K. J. M., Schieving, J. H., Kamsteeg, E. J., Kleefstra, T., Yntema, H. G., et al. (2017). A clinical utility study of exome sequencing versus conventional genetic testing in pediatric neurology. *Genet Med.* 19, 1055–1063. doi: 10.1038/gim.2017.1

Wang, Z., Benoit, G., Liu, J., Prasad, S., Aarnisalo, P., Liu, X., et al. (2003). Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature* 423, 555–560. doi: 10.1038/nature01645

Wiel, L., Baakman, C., Gilissen, D., Veltman, J. A., Vriend, G., and Gilissen, C. (2019). MetaDome: Pathogenicity analysis of genetic variants through aggregation of homologous human protein domains. *Hum. Mutat.* 40, 1030–1038. doi: 10.1002/humu.23798

Winter, B., Kramer, J., Meinhardt, T., Berner, D., Alt, K., Wenzel, M., et al. (2021). NR4A2 and dystonia with dopa responsiveness. *Mov. Dis.* 36, 2203–2204. doi: 10.1002/mds.28701

Wirth, T., Mariani, L. L., Bergant, G., Baulac, M., Habert, M. O., Drouot, N., et al. (2020). Loss-of-function mutations in NR4A2 cause dopa-responsive dystonia parkinsonism. *Mov. Dis.* 35, 880–885. doi: 10.1002/mds.27982

Xiaozhen, S., Fan, Y., Fang, Y., Xiaoping, L., Jia, J., Wuhen, X., et al. (2021). Novel truncating and missense variants in SEMA6B in patients with early-onset epilepsy. *Front .Cell Dev. Biol.* 9:633819. doi: 10.3389/fcell.2021.633819

Zetterstrom, R. H., Solomin, L., Jansson, L., Hoffer, B. J., Olson, L., and Perlmann, T. (1997). Dopamine neuron agenesis in Nurr1-deficient mice. *Science* 276, 248–250. doi: 10.1126/science.276.5310.248

Zhang, J., Yao, Y., He, H., and Shen, J. (2020). Clinical interpretation of sequence variants. *Curr. Protoc. Hum. Genet.* 106:e98. doi: 10.1002/cphg.98

Zhang, L., Le, W., Xie, W., and Dani, J. A. (2012). Age-related changes in dopamine signaling in Nurr1 deficient mice as a model of Parkinson's disease. *Neurobiol. Aging* 33, e1007–e1016. doi: 10.1016/j.neurobiolaging.2011.03.022

Zheng, K., Heydari, B., and Simon, D. K. (2003). A common NURR1 polymorphism associated with Parkinson disease and diffuse Lewy body disease. *Arch. Neurol.* 60, 722–725. doi: 10.1001/archneur.60.5.722

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 10 August 2022
DOI 10.3389/fnins.2022.963813

OPEN ACCESS

EDITED BY
Tianyun Wang,
Peking University, China

REVIEWED BY
Valerio Conti,
University of Florence, Italy
Madelyn Gillentine,
Seattle Children's Hospital,
United States

*CORRESPONDENCE
Huiping Li
lihuiping@fudan.edu.cn
Weijun Feng
fengweijun@fudan.edu.cn
Xiu Xu
xuxiu@fudan.edu.cn

†These authors have contributed equally to this work

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

RECEIVED 07 June 2022
ACCEPTED 19 July 2022
PUBLISHED 10 August 2022

CITATION
Hu M, Li H, Huang Z, Li D, Xu Y, Xu Q, Chen B, Wang Y, Deng J, Zhu M, Feng W and Xu X (2022) Novel compound heterozygous mutation in *STAMBP* causes a neurodevelopmental disorder by disrupting cortical proliferation. *Front. Neurosci.* 16:963813. doi: 10.3389/fnins.2022.963813

Novel compound heterozygous mutation in *STAMBP* causes a neurodevelopmental disorder by disrupting cortical proliferation

Meixin Hu[1†], Huiping Li[1*†], Zhuxi Huang[2,3], Dongyun Li[1], Ying Xu[2,3], Qiong Xu[1], Bo Chen[2,3], Yi Wang[1], Jingxin Deng[1], Ming Zhu[2,3], Weijun Feng[2,3*] and Xiu Xu[1*]

[1]Department of Child Health Care, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [2]Institute of Pediatrics, Children's Hospital of Fudan University, Shanghai, China, [3]Shanghai Key Laboratory of Medical Epigenetics, International Co-Laboratory of Medical Epigenetics and Metabolism, Institutes of Biomedical Sciences, Shanghai Medical College, Fudan University, Shanghai, China

Background: Mutations in the *STAMBP* gene, which encodes a deubiquitinating isopeptidase called STAM-binding protein, are related to global developmental delay, microcephaly, and capillary malformation. Owing to the limited number of reported cases, the functional and phenotypic characteristics of *STAMBP* variants require further elucidation.

Materials and methods: Whole exome sequencing was performed on a patient presenting with a neurodevelopmental disorder. Novel compound heterozygous mutations in *STAMBP* [c.843_844del (p.C282Wfs*11) and c.920G > A (p.G307E)] were identified and validated using Sanger sequencing. A 3D human cortical organoid model was used to investigate the function of *STAMBP* and the pathogenicity of the novel mutation (c.920G > A, p.G307E).

Results: The patient was presented with global developmental delay, autism spectrum disorder, microcephaly, epilepsy, and dysmorphic facial features but without apparent capillary malformation on the skin and organs. Cortical organoids with *STAMBP* knockout (KO) showed significantly lower proliferation of neural stem cells (NSCs), leading to smaller organoids that are characteristic of microcephaly. Furthermore, *STAMBP* disruption did not affect apoptosis in early cortical organoids. After re-expressing wild-type STAMBP, $STAMBP^{G307E}$, and $STAMBP^{T313I}$ (a known pathogenic mutation) within *STAMBP* KO organoids, only $STAMBP^{WT}$ rescued the impaired proliferation of *STAMBP* deficient organoids, but not $STAMBP^{G307E}$ and $STAMBP^{T313I}$.

Conclusion: Our findings demonstrate that the clinical phenotype of *STAMBP* mutations is highly variable, and patients with different *STAMBP* mutations

show differences in the severity of symptoms. The *STAMBP* missense mutation identified here is a novel pathogenic mutation that impairs the proliferation of NSCs in human brain development.

KEYWORDS

***STAMBP*, neurodevelopmental disorder, microcephaly, cortical organoids, novel mutation**

Introduction

The *STAMBP* gene (alias *AMSH*, associated molecule with the SH3 domain of STAM) on chromosome 2p13 encodes a deubiquitinating (DUB) isopeptidase called STAM-binding protein (Tanaka et al., 1999). Homozygous or compound heterozygous mutations in *STAMBP* have been reported to be pathogenic. Patients with microcephaly-capillary malformation syndrome (MICCAP: OMIM #614261) have biallelic mutations in the *STAMBP* gene (McDonell et al., 2013; Pavlović et al., 2014; Faqeih et al., 2015; Naseer et al., 2016; Demikova et al., 2018; Hori et al., 2018; Lm, 2019; Wu et al., 2019). The phenotype of MICCAP consists of global developmental delay, progressive microcephaly, intractable epilepsy, and generalized capillary malformations on the skin. However, the number of reported cases with *STAMBP* mutations is limited. The functional and phenotypic characteristics of mutations in *STAMBP* have not been fully elucidated.

The STAMBP protein contains a microtubule-interacting and transport (MIT) domain and a STAM-binding motif, both of which interact with endosomal sorting and trafficking machinery (Agromayor and Martin-Serrano, 2006; Wright et al., 2011; Davies et al., 2013). STAMBP is important for cell surface receptor-mediated endocytosis and sorting. Previously, it was reported that insensitive activation of the RAS-mitogen activated protein kinase (RAS-MAPK) and PI3K-AKT-mTOR pathways might contribute to vascular and capillary malformation and apoptosis induction by a defective DUB, which may be responsible for microcephaly (McDonell et al., 2013). Interestingly, one study failed to repeat the constitutive activation of the PI3K-AKT-mTOR pathway in patient-derived lymphoblastoid cell lines (LCLs) with a novel *STAMBP* mutation (Hori et al., 2018). This raises the question of whether different mutations in *STAMBP* affect different signaling pathways and lead to a variable clinical phenotype.

Studies involving the *Stambp*$^{-/-}$ mice found that homozygous knockout mice were morphologically indistinguishable from their littermates at birth, and histopathological examination revealed normal morphogenesis in all tissues tested (Ishii et al., 2001). Early postnatal mice display neurodegenerative apoptotic activation and ubiquitin-conjugated protein aggregation in the hippocampus and cerebral cortex (Suzuki et al., 2011). However, patients carrying *STAMBP* mutations exhibited clinical phenotypes congenitally, indicating that the function of *STAMBP* in humans is different from mice. To address the gap between mouse models and human diseases, human brain organoids generated from pluripotent stem cells have emerged as a promising approach for investigating the disease characteristics in relevant cellular and genetic contexts (Lancaster et al., 2013; Chen et al., 2014; Bershteyn et al., 2017; Quadrato et al., 2017).

Here, in this study, we present a Chinese patient diagnosed with a neurodevelopmental disorder carrying compound heterozygous *STAMBP* mutations, including a novel STAMBP^{G307E} mutation. Using the human cortical organoids (hCOs) model, we showed that the *STAMBP* deficiency disrupts the proliferation of neural stem cells (NSCs), leading to a dramatic size reduction in hCOs. We also found that the newly identified STAMBP^{G307E} mutation could not rescue phenotypes caused by *STAMBP* deficiency.

Patient and methods

Ethics statement

Ethical approval for the present study was obtained from the Ethics Committee of Children's Hospital affiliated with Fudan University (permit no. 2016-131). Informed consent was obtained from the patient's parents.

Patient

The patient who was a girl 3 years and 11 months old was the only child of healthy non-consanguineous Chinese parents (pregnant in their early 30s). She was delivered at term *via* normal spontaneous vaginal delivery, with a birth weight of 2800 g (15th percentile, according to the WHO Child Growth Standards). Birth length and head circumference have not been reported.

Physical examination revealed an apparent growth delay, especially in head circumference. At 18 months of age, her weight was 8.6 kg (5th–15th percentile), height was 75.1 cm

(1st–3rd percentile), and head circumference was 42.5 cm (<1st percentile). At 3.5 years, her weight was 12.1 kg (5th percentile), her height was 85.1 cm (<1st percentile), and her head circumference was 44.8 cm (<1st percentile). Dysmorphic features included hypertelorism, low-set ears, anteverted nares, and drooping corners of the mouth. In contrast to the previously reported cases, the girl did not show apparent capillary malformations on the skin. Ultrasound examination revealed no capillary malformation in the liver, spleen, or kidneys.

Her seizures began in infancy and manifested as apnea and eyes turning up during sleep (generally within 10–20 min of falling asleep) without limb shaking or stiffness and restored consciousness after several minutes. Even with a small head circumference, cerebral magnetic resonance imaging (MRI) with spectroscopy revealed no remarkable anomalies (**Figure 1E**). However, the electroencephalogram (EEG) examination was strikingly abnormal, with generalized epileptogenic activity.

She had a global developmental delay and did not walk until 20 months of age. Her speech and language acquisition were delayed and she only could say "mama" and "papa." She avoided eye contact and did not respond to the adults' simple instructions. She was hyperactive and exhibited a series of repetitive hand movements, such as putting her hands in front of her eyes, flicking her fingers, digging her fingers into a wall or book, and flipping through books.

She has suffered from feeding problems, refused to accept new foods, and rejected drinking milk and eating fruits and vegetables. Her major food sources were limited to rice and eggs. We performed a laboratory examination of liver and kidney function. Tests showed that she had hyperuricemia and hypercholesterolemia (blood uric acid was 522 μmol/L; total cholesterol was 6.39 mmol/L; high-density lipoprotein was 2.07 mmol/L; free fatty acid was 634 μmol/L). After 2 months of dietary adjustment, her blood tests of uric acid and lipid levels were recovered (blood uric acid was 348 μmol/L; total cholesterol was 5.02 mmol/L; high-density lipoprotein was 1.66 mmol/L; free fatty acid was 794 μmol/L).

Variation analysis

Peripheral blood samples (2 ml) taken from the patient and her parents were collected into tubes containing EDTA anticoagulant. Sample preparation and whole-exome sequencing were performed at the Molecular Genetic Diagnosis Center of Children's Hospital of Fudan University. The genomic DNA of all samples was sequenced using the HiSeq 2000 platform (Illumina). Sequencing data were generated and assembled using the Ensemble GRCh37/hg19 reference genome. Mutation validation was performed using Sanger sequencing with two primer pairs to amplify exon 6 (chr2:74076590-74076591, forward primer:5′-AGGGCTCAGTGGTCGCAGA-3′; reverse primer:5′-GAGAG TCACAGGATGCCAAGAG-3′) and exon 7 (chr2:74077555, forward primer:5′-GCTTACCTTTCCACTGTCGG-3′; reverse primer,5′-TAAAAGCCCTAAGTGTTCCCAGA-3′) of the *STAMBP* gene (NM_006463). Mutations were predicted using Addgene software. The pathogenicity of the genetic variants was predicted using bioinformatics tools, such as PolyPhen-2,[1] PROVEAN, and SIFT.[2] The variants were classified according to American College of Medical Genetics and Genomics (ACMG) guidelines.

Literature review

The PubMed,[3] CNKI,[4] Wanfang,[5] and ClinVar[6] databases were used to retrieve previous studies with the keywords of "STAMBP or AMSH" until April 2022.

Human embryonic stem cells culture

H9 human embryonic stem cells (hESCs) were cultured on Matrigel-coated tissue culture dishes containing mTeSR plus medium. Cells were passaged every 4 days using EDTA (0.5 mM).

CRISPR/Cas9-based genome editing

The CRISPR guide for *STAMBP* was designed using the Benchling CRISPR Guide Design Tool.[7] The guide was designed to maximize on-target efficiency and minimize off-target sites in the intragenic regions. *STAMBP* sgRNA:5′-TCCCAAAGCAGAAGAGCTGA -3′ (PAM: AGG). The corresponding coding sequence was cloned into the pSpCas9(BB)-2A-Puro (PX459) construct (Addgene, plasmid #62988). A total of 1×10^6 of single cells dissociated from H9 hESCs cultured were transfected with 3 μg PX459-sgRNA using Lipofectamine 3000. After transfection, the cells were plated onto Matrigel-coated 6-well plates with mTeSR plus medium containing 10 μM Y27632. Two days of puromycin selection were performed starting 24 h after transfection. After 5–7 days of recovery, the cells were dissociated into single cells using Accutase and cultured in Matrigel-coated 96-well plates.

1 http://genetics.bwh.harvard.edu/pph2/
2 http://provean.jcvi.org/index.php
3 https://pubmed.ncbi.nlm.nih.gov/
4 https://www.cnki.net/
5 https://new.wanfangdata.com.cn/index.html
6 https://www.ncbi.nlm.nih.gov/clinvar/
7 https://www.benchling.com/

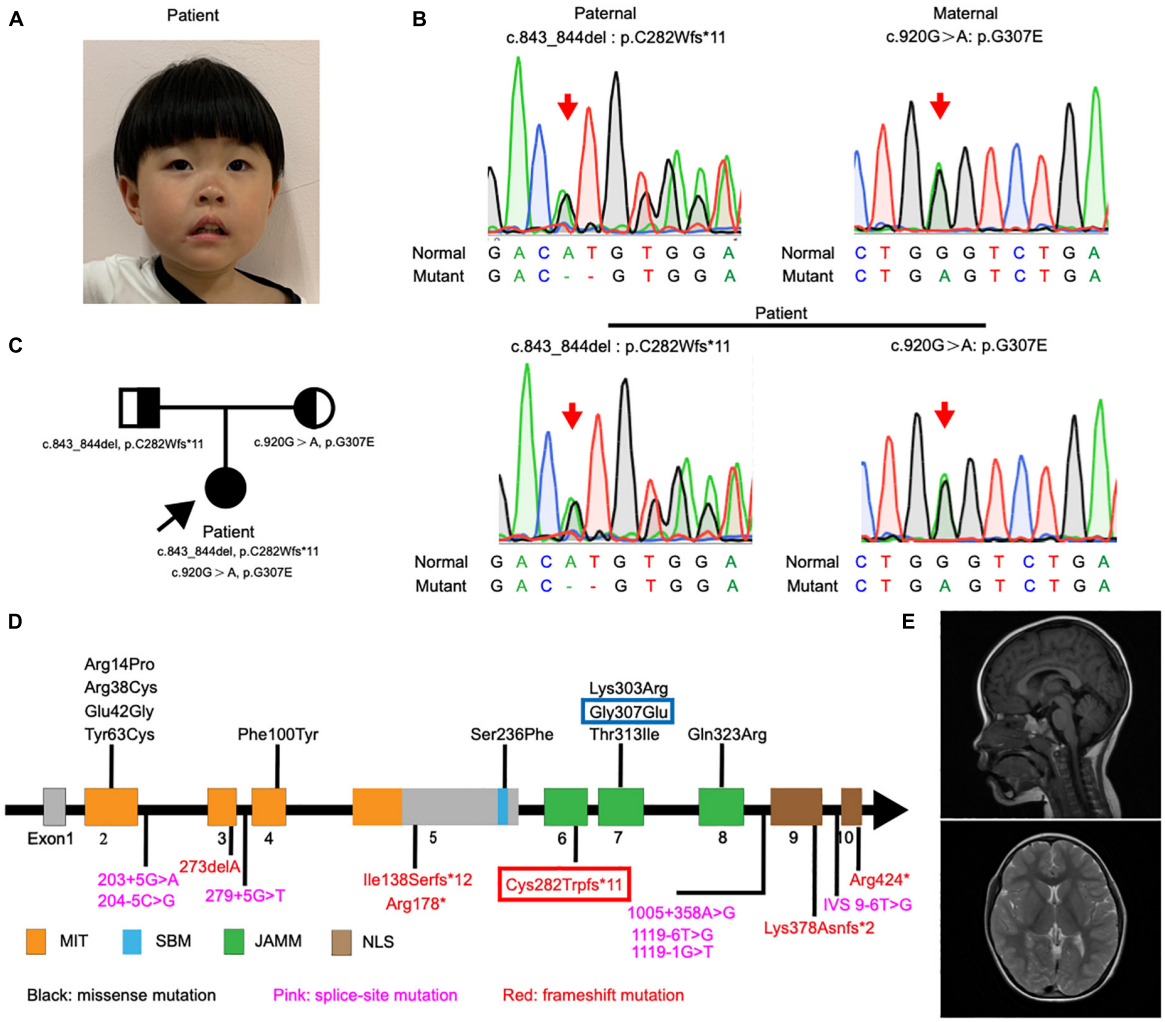


FIGURE 1
Detection and location of *STAMBP* mutation in case patient. **(A)** Clinical photograph showing craniofacial phenotypes (hypertelorism, wide nasal bridge, anteverted nares, and drooping corners of the mouth). **(B)** Sanger sequencing of the *STAMBP* mutations showed that the case patient has compound heterozygous mutations from her father (c.843_844del, p. C282Wfs*11) and mother (c.920G > A, p.G307E) (arrow). **(C)** Pedigree of the patient. **(D)** Diagram indicates the location of the patients' mutations of *STAMBP* described in previous reports and ours. Blue frame labels this case's mutation inherited from her mother. Red frame labels this case's mutation inherited from her father. STAMBP contains a microtubule-interacting and transport (MIT) domain, an SH3 binding motif (SBM), a JAMM (JAB/MPN/MOV34) motif, and a nuclear localization signal (NLS). **(E)** Representative images showing cerebral MRI scans of the patient.

After 10–14 days of culture, half of each clone was collected for genomic DNA isolation. Clones were screened using PCR and Sanger sequencing. The target clones were expanded for downstream use.

Cortical organoids generation

Human cortical organoids (hCOs) were generated using a protocol adapted from the study done by Xiang et al. (2017). H9 hESCs were dissociated using Accutase and cultured in neural induction medium (DMEM-F12, 15% (v/v) KSR, 1% (v/v) MEM-NEAA, 1% (v/v) GlutaMAXmax, 100 μM β-Mercaptoethanol, 100 nM LDN-193189, 10 nM SB-431542, and 2 μM XAV-939) supplemented with 50 μM Y27632 and 5% (v/v) FBS. A total of 9,000 cells were plated in each well of an ultra-low-attachment 96-well plate. Neural induction media were replenished every other day until day 10 (FBS was removed from day 2 and Y27632 was removed from day 4). On day 10, organoids were transferred into ultra-low-attachment 6-well plates for spinning culture (85 rpm) and cultured with neuronal differentiation media (1:1 mixture DMEM-F12 and Neurobasal media, 0.5% (v/v) N2 supplement, 1% (v/v) B27 supplement without vitamin A, 1% (v/v) Glutamax, 0.5% (v/v) MEM-NEAA, 0.025% (v/v) insulin, 50 μM β-Mercaptoethanol, and 1% (v/v) penicillin/streptomycin). Neural differential media

were replenished every other day until day 18. On day 18, a B27 supplement with vitamin A was used, and 20 ng/ml BDNF, 200 μM ascorbic acid, and 200 μM cAMP were added. The medium was replenished every 3 days thereafter.

Western blot analysis

Human embryonic stem cells were lysed in a buffer containing 20 mM Tris-HCI (pH 8.0), 1 mM EDTA, 1 mM EGTA, 1% Triton X-100, 450 mM NaCl, 1 × protease inhibitor cocktail (Roche). Protein samples were loaded onto a 10% PAGE gel (Epizyme) and transferred onto PVDF membranes. The membranes then were incubated with anti-SATMBP (mouse, SCBT sc-271641, 1:1,000), anti-Flag (mouse, Shanghai Genomics GNI4110-FG, 1:1000) at 4°C overnight and incubated with secondary antibodies at room temperature for 1 h. The blots were developed using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) and the ChemiDoc System (Bio-Rad).

RT–qPCR

All RNA samples were extracted using the TRIzol reagent (Sigma-Aldrich). The PrimeScript RT Master Mix (Takara) was used for cDNA synthesis. Quantitative PCR was performed using TB Green Premix Ex Taq II (Takara) on a CFX384 Touch Real-Time PCR Detection System (Bio-Rad). The primers used for RT-qPCR are listed in the **Supplementary Table 3**.

Immunostaining and fluorescence quantification

Organoids were fixed in 4% paraformaldehyde for 2 h at room temperature, followed by washing with PBS 3 times for 5 min each. The fixed organoids were then allowed to sink in 30% sucrose overnight at 4°C, then embedded in O.C.T. and cryosectioned at 20 μm. The organoid sections were blocked with 3% donkey serum, 0.2% TritonX-100, and 0.1% Tween in PBS for 1 h at RT. Then, the samples were incubated with primary antibodies overnight at 4°C and secondary antibodies for 1 h at RT. Images of these sections were taken using a confocal microscope (Leica TSC SP8). ImageJ software was used for image processing and quantification of fluorescence intensity. These methods have been previously described (Guo et al., 2021). To measure the surface area of the organoid, a circle was drawn surrounding it and quantified using ImageJ software. To measure the areas of SOX2-positive cells, a circle was drawn surrounding the ventricular zone (VZ)-like a rosette and quantified using the ImageJ software. The mean fluorescence signal in each section was detected and quantified using ImageJ to measure the fluorescence intensity of Ki67-positive, PH3-positive, and CC3-positive cells. The antibodies used for immunostaining are listed in the **Supplementary Table 2**.

STAMBP knockout rescue experiment

FLAG-HA-STAMBP plasmid was purchased from Addgene (#22560). The FLAG-tagged STAMBP^{WT}, FLAG-tagged STAMBP^{G307E}, and FLAG-tagged STAMBP^{T313I} sequences were cloned into the PPB-FH-IRES-PuroR vector. For STAMBP overexpression-mediated rescue experiments, *STAMBP* KO hESCs were overexpressing FLAG-tagged STAMBP^{WT}, FLAG-tagged STAMBP^{G307E}, or FLAG-tagged STAMBP^{T313I} by plasmid transfection using Lipofectamine 3000 followed by the generation of cortical organoids.

Statistical analysis

All the raw data are showed in the **Supplementary Data Sheet 1**. The data were analyzed and visualized using GraphPad PRISM 7.0a. The ShapiroWilk test and the Kolmogorov–Smirnov test were used to test for normality. Normally distributed datasets were compared using Student's *t*-test and one-way ANOVA with Dunnett's multiple comparison tests. Values are presented as mean ± SEM. Data that were not normally distributed were analyzed using the Mann–Whitney test and Kruskal–Wallis ANOVA with Dunn's multiple comparisons tests and presented as a median ± 95% confidence interval. All statistical tests were two-tailed, and statistical significance was defined as $p < 0.05$. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Results

The clinical features and molecular phenotype of the patient

The patient was diagnosed with global developmental delay and autism spectrum disorder at 3 years of age. Consistent with the clinical observations, cognitive/developmental evaluation using the Griffiths Mental Development Scales (GMDS) showed that her locomotor, personal-social, hearing-speech, coordination, and performance skills were well below the age range (<1st percentile). The Autism Diagnostic Observation Schedule, second edition (ADOS-2) Module 1 was used to assess autism spectrum disorders (ASD). The results of the ADOS-2 (score of social effect: 20; restricted and repetitive behavior: 5; overall total

10.3389/fnins.2022.963813

TABLE 1 Summary of *STAMBP* mutant patients' early development and evaluation results.

Subjects	This case	Reported cases
Gender	Female	15/20, male
Age	3 years 11 months	22 days–12 years
Short stature	Yes	15/20
Microcephaly	Yes	20/20
Epilepsy	Yes	20/20
Developmental delay	Yes	20/20
Autism-like behavior	Yes	2 (18 NA)/20
Age of walking (months)	20	—
Age of speaking (months)	Non-verbal at evaluation	Non-verbal at evaluation, 4 (16 NA)/20
Comorbidity		
Feeding problem	Yes	2 (18 NA)/20
Sleep disorder	No	2 (18 NA)/20
Clinical evaluations		
Age of GMDS (months)	35	—
Locomotor	<1st percentile	
Personal-social	<1st percentile	
Hearing-speech	<1st percentile	
Hand-eye coordination	<1st percentile	
Performance	<1st percentile	
Age of ADOS (months)	35	—
Social affect	20	
Restricted and repetitive behavior	5	
Total score	25 (ASD cut-off: 11)	

NA, not available; GMDS, Griffiths Mental Development Scales; ADOS, Autism Diagnostic Observation Schedule, second edition.

score: 25) led to the diagnosis of autism with a high level of symptomatology.

The patient showed a short stature and microcephaly. She also showed dysmorphic features including hypertelorism, low-set ears, and drooping corners of the mouth. Notably, we also observed a new feature of anteverted nares (**Figure 1A**).

Exome sequencing of genomic DNA from the patient revealed the presence of compound heterozygous mutations c.843_844del (p.C282Wfs*11) and c.920G > A (p.G307E) in the *STAMBP* gene. Her father and mother were heterozygous for c.843_844del and c.920G > A, respectively. Sanger sequencing was performed to validate the mutations (**Figure 1B**). The mode of inheritance in the proband is consistent with an autosomal recessive pattern, shown by pedigree analysis (**Figure 1C**). After reviewing *STAMBP* variants in patients reported previously, we summarized phenotypes and mutations in *STAMBP*: 10 missense mutations, six different frameshift mutations predicted to cause premature truncation of the STAMBP protein, and seven intronic mutations leading to alternative splicing of the *STAMBP* transcript (**Figure 1D**).

Most patients with *STAMBP* mutations have shown global developmental delay, microcephaly, capillary malformation, epilepsy, and dysmorphic features. Our patient met most of the characteristics of mutation in *STAMBP* (**Table 1** and **Supplementary Table 1**). Unlike previously reported cases, this girl did not show any obvious capillary malformations on the skin.

Protein consequences of *STAMBP* mutation

In our case, the patient carried compound heterozygous paternal (c. 843_844del, p. C282Wfs * 11) and maternal (c. 920G > A, p. G307E) mutations. After searching the PubMed, CNKI, Wanfang, and ClinVar databases as of April 2022, we found that p.C282Wfs*11 has been reported to be pathogenic in ClinVar (Variation ID:1034283), while p.G307E has not been reported. The protein structure of STAMBP (p.C282Wfs*11), predicted using SWISS-MODEL, showed a truncated protein caused by a frameshift mutation with a 2-basepair (AT) deletion (**Figure 2A**). For the protein consequences of STAMBP (p.G307E) mutation, multiple sequence alignment (MSA) showed that the missense mutation (c.920G > A) converts a conserved glycine residue to glutamic acid (p.G307E) (**Figure 2B**). The functional prediction programs PolyPhen-2, PROVEAN, and SIFT (Kumar et al., 2009; Adzhubei et al., 2010; Choi and Chan, 2015) were used to determine the effects of the p. G307E mutation. PolyPhen-2 predicted the change to be "possibly damaging" with a score of "0.763"; PROVEAN predicted the mutation to be "deleterious" with a score of "−3.25" and SIFT predicted the mutation to be "damaging" with a score of "0.010." In addition, the protein structure of the mutation p.G307E was predicted using SWISS-MODEL and PyMOL (Schwede et al., 2003; Johansson et al., 2012). The mutation p.G307E was located at the loop structure within the JAMM domain, which has the function of binding with ubiquitinated proteins. Altering hydrogen bonds at the side chain position was predicted to affect the function of the STAMBP protein (**Figure 2C**). Taken together, these predictions suggest that p. G307E likely has a negative impact on STAMBP function.

STAMBP is expressed in the developing human brain

First, we investigated the expression pattern of *STAMBP* in the human brain. Using the public gene-expression database, the Allen Brain Atlas, which characterizes the spatiotemporal expression pattern of the human brain, the data showed

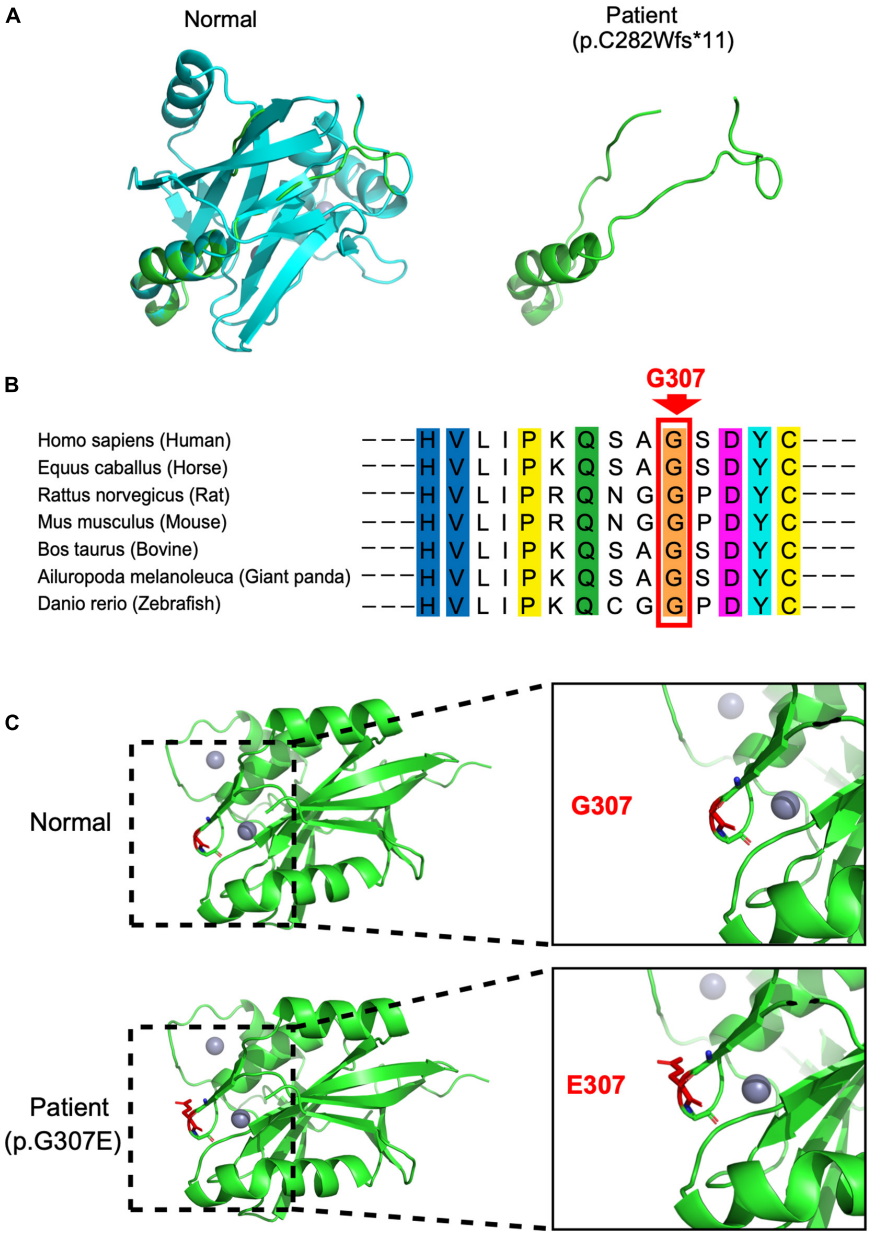


FIGURE 2
Protein consequences of *STAMBP* mutation. **(A)** Superimposition of the STAMBP (p.C283Wfs*11) protein was achieved using SWISS-MODEL and visualized using PyMOL. The protein structures of STAMBP (left) and STAMBP (p.C283Wfs*11) variation (right). The truncated protein is shown in green. **(B)** Conservation of mutated residue across evolution. **(C)** Superimposition of STAMBP (p.G307E) protein was performed using SWISS-MODEL and visualized using PyMOL. Expanded view of patient mutation (p.G307E) showing alternation of the Hydrogen bonds (red dashes) in a loop structure. Also, gray spheres show two bound Zn^{2+} atoms. α-helix and β-sheet are shown in green.

that *STAMBP* was expressed 8 weeks after conception and stably expressed into adulthood. *STAMBP* was widely expressed in multiple brain areas, including the cerebral cortex, hippocampus, and cerebellum (Figure 3A).

To test the expression of *STAMBP* during human brain development, we generated hESCs derived hCOs using H9 hESCs (Figure 3B). We found that radially organized cells resembling the VZ were enriched with $SOX2^+$ and $PAX6^+$ cells on day 21, confirming their identity as NSCs (Figure 3C, left column). The neuroblast-specific marker, doublecortin (DCX), was expressed on the outer side of $SOX2^+$ NSCs on day 28 (Figure 3C, middle column). On day 35, NeuN, indicative of differentiated neurons, was observed outside the VZ-like areas (Figure 3C, right column). These results demonstrate that hCOs differentiated to a similar extent and acquired morphological features of organized cortical structures. RT-PCR

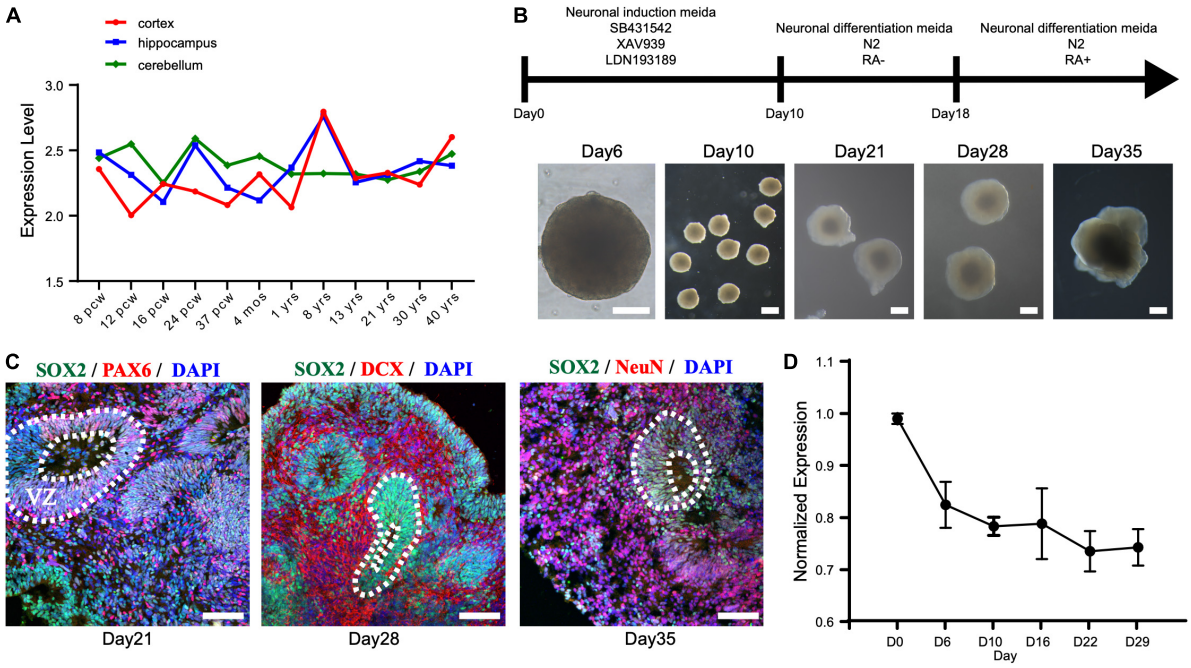


FIGURE 3
STAMBP expression during human cortical development. **(A)** STAMBP expression in the human cortex, hippocampus, and cerebellum. Data were downloaded from Allen Brain Atlas (available from: http://www.brain-map.org; Ensemble ID: ENSG00000124356). pcw, post-conceptional weeks; mos, months; yrs, years. **(B)** Schematic representation of the culture system described in detail in Section "Patient and methods." Examples of each stage are shown. Scale bars, 200 μm. **(C)** Immunofluorescent staining of hCOs of day 21 for neural stem cell markers Left column: SOX2, PAX6, and DAPI; middle column: SOX2, DCX, and DAPI; right column: SOX2, NeuN, and DAPI. Scale bars, 100 μm. **(D)** *STAMBP* mRNA expression in the different differentiated stages of hCOs. Mean ± SEM from three independent experiments.

was performed to analyze the expression patterns of *STAMBP* mRNA at different developmental stages. The results showed that *STAMBP* was expressed in hESCs and its expression was maintained during the induction and differentiation stages of cortical organoids (Figure 3D).

STAMBP deletion results in smaller cortical organoids and impairs neural stem cell proliferation

Next, we generated *STAMBP* mutant hESCs using the CRISPR/Cas9 approach (Ran et al., 2013). *STAMBP* KO hESCs occurred as one cytosine nucleotide insertion in exon 4, resulting in a frameshift, leading to the generation of a premature stop codon (Figure 4A). Western blotting confirmed the absence of STAMBP in the mutant hESCs (Figure 4B). To examine whether the deletion of *STAMBP* affects the pluripotency of hESCs, we examined the expression of pluripotent markers. Immunostaining and RT-PCR were performed to determine the expression of pluripotent markers, including SOX2, NANOG, and REX1. The results showed no significant changes in the expression of pluripotency markers, suggesting *STAMBP* deletion did not affect the pluripotency of hESCs (Supplementary Figures 1A,B).

The size of organoids generated from *STAMBP* KO hESCs and wild-type H9 cells was measured at different time points. *STAMBP* KO hCOs were drastically smaller in size and showed significantly reduced surface areas compared to wild-type H9 hCOs from day 10 (Figures 4C,D). Microcephaly may be caused by the abnormal proliferation and apoptosis of NSCs (Thornton and Woods, 2009; Manzini and Walsh, 2011). Therefore, we immunostained the NSC marker SOX2 to indicate VZ-like regions on day 28 organoids and measured the area of $SOX2^+$ VZ-like regions. Regardless of not reaching the threshold for statistical significance, *STAMBP* KO organoids showed a trend of reduction in $SOX2^+$ VZ-like regions (Figures 4E, first row,F). The proliferation of *STAMBP* KO organoids was monitored by immunostaining for the cell cycle marker Ki67, and the results showed a decrease in cell proliferation in *STAMBP* KO organoids compared to control organoids (Figures 4E second row,G). We also stained for the mitotic cell marker phospho-histone 3 (PH3) on day 28 organoids. The results showed a significantly reduced $PH3^+$/DAPI area in *STAMBP* KO organoids as compared to controls (Figures 4E third row,H), indicating impaired proliferation of NSC in *STAMBP*-deficient hCOs. Given that previous studies in a mouse model

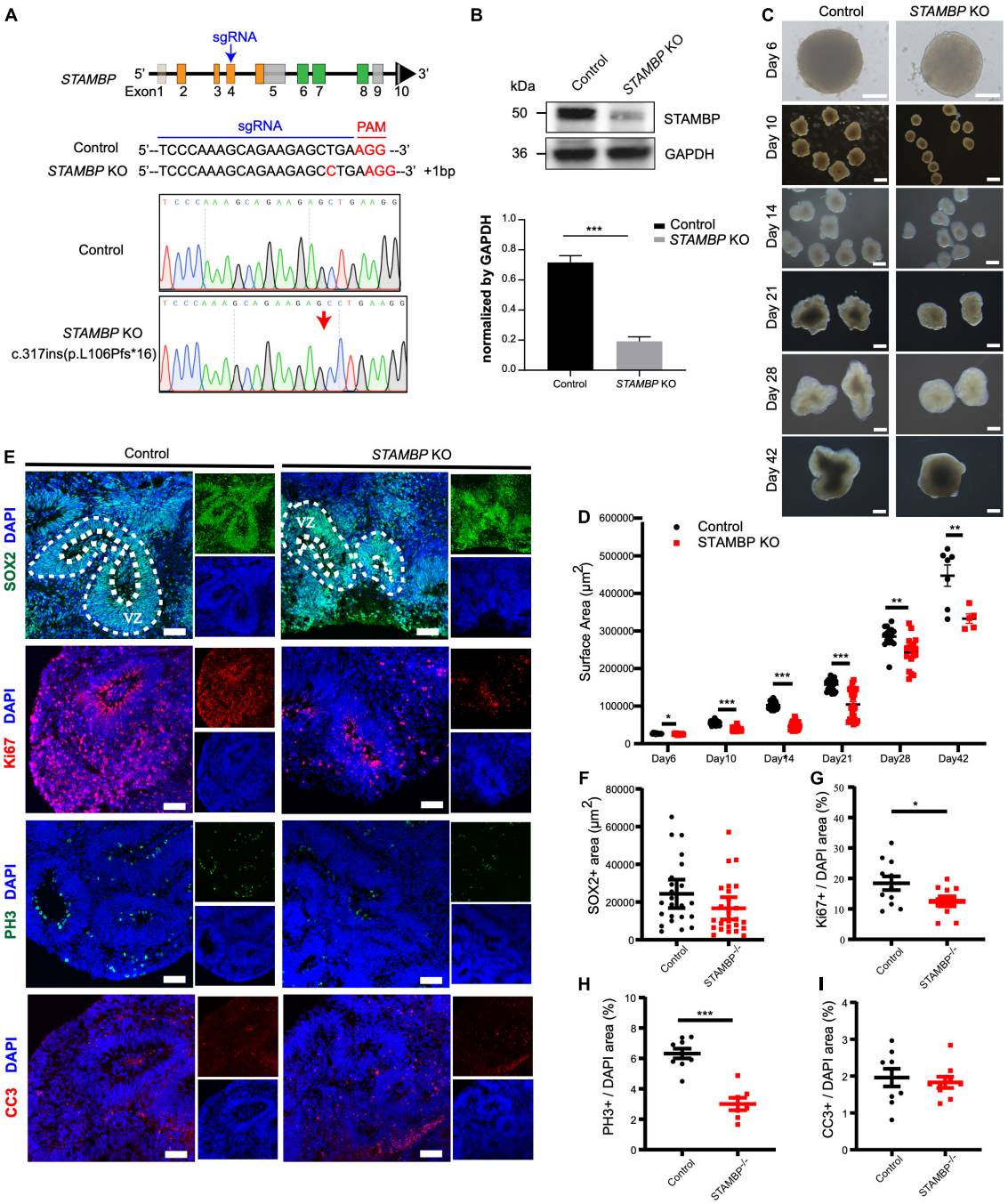


FIGURE 4
STAMBP deletion results in smaller cortical organoids sizes and impairs NSC proliferation. **(A)** CRISPR/Cas9-mediated gene editing of human *STAMBP* locus in hESCs resulted in a 1 bp insertion in exon 4. **(B)** Western blot showed ablation of STAMBP proteins due to premature mature stop codons in *STAMBP* KO hESCs. **(C)** Representative images of control and *STAMBP* KO cortical organoids at different differential stages. Scale bars, 200 μm. **(D)** Quantification of surface areas of cortical organoids in different developmental stages. Each plot represented an individual organoid. Experiments were repeated three times. Data of day 6, day 21, day 28, and day 42 organoids are presented as mean ± SEM with student's *t*-tests. Data of day 10 and day 14 organoids are presented as a median ± 95% confidence interval with the Mann–Whitney test. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. **(E)** Immunofluorescent staining was performed on sections of control and *STAMBP* KO cortical organoids on day 28. Sections were stained with SOX2 and DAPI (the first row), Ki67 and DAPI (the second row), PH3 and DAPI (the third row), and CC3 and DAPI (the fourth row). Scale bar: 100 μm. **(F–I)** Quantification of the VZ-like $SOX2^+$ rosette area **(F)**, the ratio of $Ki67^+$ vs. DAPI **(G)**, the ratio of $PH3^+$ vs. DAPI **(H)**, and the ratio of $CC3^+$ vs. DAPI **(I)**. Experiments were repeated three times. Data of $SOX2^+$ area are presented as median ± 95% confidence interval with the Mann–Whitney test. Data of $Ki67^+$/DAPI, $PH3^+$/DAPI, and $CC3^+$/DAPI are presented as mean ± SEM with student's *t*-tests. $*p < 0.05$, $***p < 0.001$.

showed that loss of STAMBP causes apoptosis in neuronal cells, we also stained for the apoptosis marker cleaved caspase 3 (CC3) (**Figure 4E**, fourth row). In contrast to the findings in the mouse model, our data showed no difference in apoptosis between *STAMBP* KO hCOs and wild-type H9 hCOs (**Figure 4I**). Collectively, these results suggest a disruption of NSC proliferation in *STAMBP* mutant organoids.

Overexpression of the patient's missense mutation, STAMBP^{G307E}, could not rescue the impaired proliferation of neural stem cells in *STAMBP* deficient organoids

To confirm the causal relationship between *STAMBP* deletion and phenotypes of cortical organoids, we used the PiggyBac transposon system to transpose the *STAMBP* gene into *STAMBP* KO hESCs (**Figure 5A**). Furthermore, it was unknown whether the novel missense variant STAMBP^{G307E} was pathogenic in this patient. Hence, we overexpressed FLAG-STAMBP^{WT}, FLAG-STAMBP^{G307E}, and FLAG-STAMBP^{T313I}, a previously reported *STAMBP* missense variant (c.938C > T, Thr313Ile) in *STAMBP* KO hESCs (**Supplementary Figure 2A**). The missense mutations were confirmed with Sanger sequencing (**Supplementary Figure 2B**). Western blotting confirmed the re-expression of STAMBP in *STAMBP* deficient hESCs (**Figure 5B**).

We subsequently repeated the experiments following the same organoid differentiation protocol as described above. Overexpressed STAMBP^{WT} in *STAMBP* KO hESCs increased the generated organoids' size compared to *STAMBP* KO organoids on day 28, whereas overexpressed STAMBP^{G307E} and STAMBP^{T313I} failed to rescue the size of *STAMBP* deficient organoids (**Figures 5C,D**).

In addition, we examined whether impaired cellular proliferation in *STAMBP* KO organoids could be rescued. Immunostaining analysis showed that the decreased Ki67^{+} and PH3^{+} NSCs pools were partially rescued in organoids generated from STAMBP^{WT} overexpression. However, organoids generated from STAMBP^{G307E} and STAMBP^{T313I} overexpression showed an indistinguishable difference in Ki67^{+} and PH3^{+} cells compared to *STAMBP* KO organoids on day 28 (**Figures 5E** first and second rows,**F,G**). To further analyze the possible impact on apoptosis, we also stained the apoptosis marker, CC3, in organoids generated from the STAMBP^{G307E} and STAMBP^{T313I} overexpressing hESCs. The results showed that there was no significant difference in the CC3^{+}/DAPI areas of hCOs generated from STAMBP^{G307E}, STAMBP^{T313I}, and STAMBP^{WT} compared to control or *STAMBP* KO hCOs on day 28 (**Figures 5E** third row,**H**). Together, these results demonstrate that the decreased size of *STAMBP* deficient cortical organoids is mainly due to the impaired proliferation of NSCs. The *STAMBP* missense variant, STAMBP^{G307E} is a novel pathogenic variant. The impaired proliferation of NSCs may be a major contributor to the pathogenic mechanism in our patients.

Discussion

In this study, we describe a girl with global developmental delay, autism spectrum disorder, microcephaly, and minor dysmorphic facial features. She had compound heterozygous mutations of *STAMBP* [c.843_844del (p.C282Wfs*11) and c.920G > A (p.G307E)] inherited from her parents. MSA, PolyPhen-2, PROVEAN, and SIFT software were used to predict the pathogenicity of the novel mutation (c.920G > A, p.G307E). These predictions suggested that p.G307E is likely to have a negative impact on STAMBP function. We also used human cortical organoids to investigate the effects of this novel mutation. The data showed that in contrast to STAMBP^{WT}, STAMBP^{G307E} overexpression could not rescue the impaired proliferation of NSCs in *STAMBP*-deficient organoids. We suggest that the missense mutation (c.920G > A) in the *STAMBP* gene is a novel pathogenic mutation.

At least 20 cases identified with *STAMBP* gene mutation have been reported (McDonell et al., 2013; Pavlović et al., 2014; Faqeih et al., 2015; Naseer et al., 2016; Demikova et al., 2018; Hori et al., 2018; Lm, 2019; Wu et al., 2019). Global developmental delay, progressive microcephaly, epilepsy, and capillary malformations are the typical clinical manifestations. The patient described here showed most of these symptoms. However, no capillary malformations were found on her skin, and it was difficult to find such malformations on the internal organs by ultrasound. Compared with those of previously reported patients, this patient's symptoms were relatively mild. Capillary malformations are cutaneous vascular abnormalities associated with RAS-MAPK pathway dysregulation (Tidyman and Rauen, 2009). The pathogenesis of capillary malformation in *STAMBP* mutations has not been completely addressed. McDonell et al. (2013) found that STAMBP interacts with the Grb2 adaptor, an important component of the RAS signal transduction pathway, and they proposed that the RAS pathway may be responsible for capillary abnormalities in patients with *STAMBP* mutation. Different mutation sites of *STAMBP* may have different downstream targets and cause various symptoms, which is supported by a study of the Ser236Phe mutation of *STAMBP* (McDonell et al., 2013; Hori et al., 2018). Therefore, we propose that capillary malformation is not a definite requirement for *STAMBP* mutation-related disease. However, we need to follow up with the patient and observe the changes in her clinical features.

Human brain size is determined through a tightly orchestrated and intricate process of neural stem cell proliferation, migration, organization, apoptosis, and cell

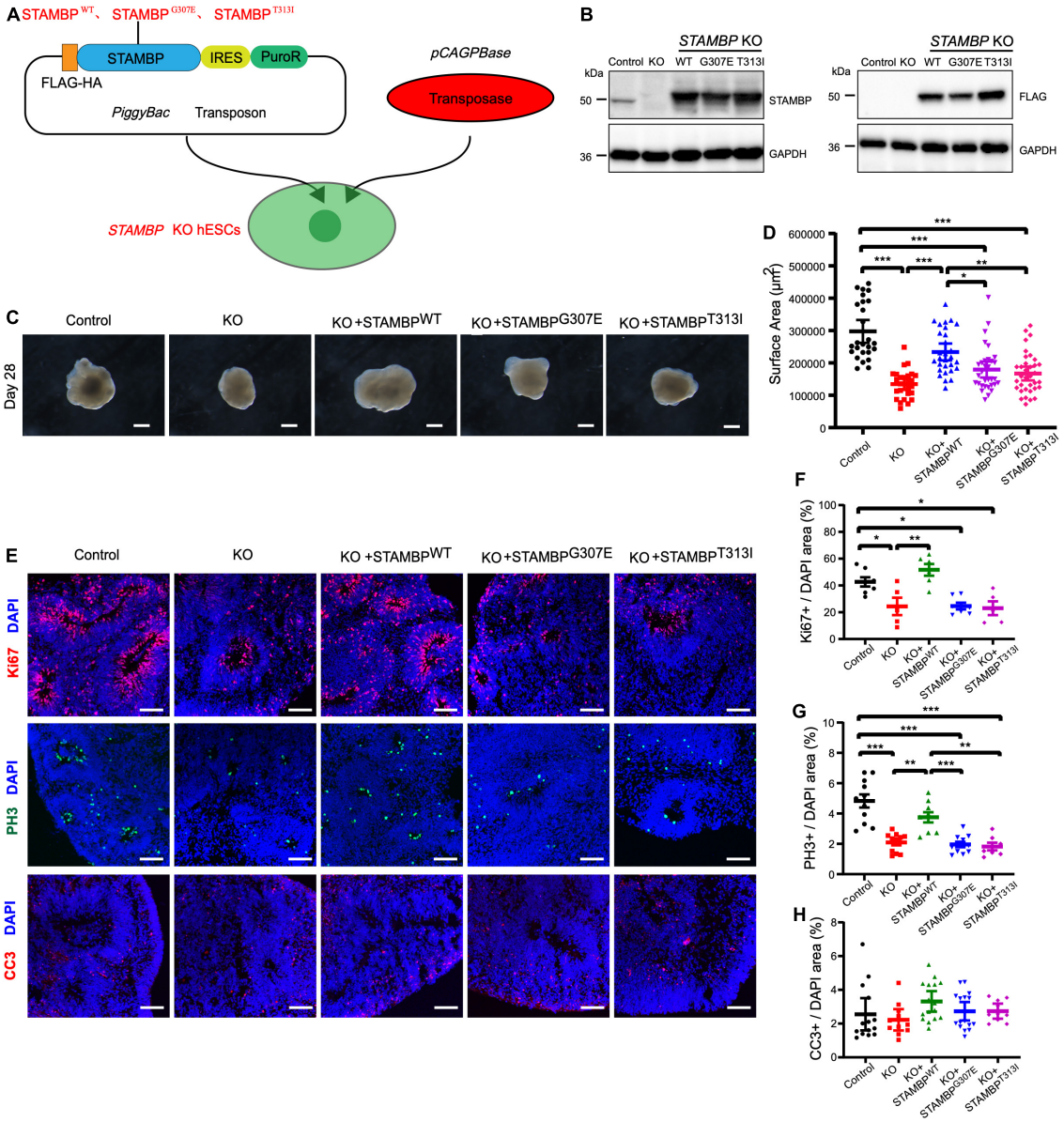


FIGURE 5
STAMBP^{WT} but not STAMBP^{G307E} and STAMBP^{T313I} overexpression rescues the phenotypes of reduced cortical size and impaired proliferation of NSCs. **(A)** Schematic representation of the PiggyBac-mediated STAMBP expression system. **(B)** Western blotting confirmed the presence of STAMBP^{WT}, STAMBP^{G307E}, and STAMBP^{T313I} protein generated from the *STAMBP* KO hESCs. **(C)** Representative images of control, *STAMBP* KO cortical organoids, and organoids generated from STAMBP^{WT}, STAMBP^{G307E}, and STAMBP^{T313I} overexpression on day 28. Scale bars, 200 μm. **(D)** Quantification of surface areas of cortical organoids. Each plot represented an individual organoid. Experiments were repeated three times. Groups were compared using a Kruskal–Wallis ANOVA with Dunn's multiple comparisons tests. Data are presented as median ± 95% confidence interval, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. **(E)** Immunofluorescent staining was performed on organoids on day 28. Sections were stained with Ki67 (the first row), PH3 (the second row), and CC3 (the third row). Scale bar: 100 μm. **(F–H)** Quantification of cell cycling marker Ki67-positive cells **(F)**, PH3-positive cells **(G)**, and cell apoptosis marker CC3 **(H)**. Each plot represented an individual section. Experiments were repeated three times. Data of Ki67$^+$/DAPI and PH3$^+$/DAPI were compared using a one-way ANOVA followed by a Dunnett's multiple comparisons test. Data are shown as mean ± SEM. Data of CC3$^+$/DAPI were using a Kruskal–Wallis ANOVA with Dunn's multiple comparisons tests. Data are presented as a median ± 95% confidence interval. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

growth (Pirozzi et al., 2018). The delicate study by McDonell et al. (2013) showed that reduced STAMBP expression was associated with the accumulation of ubiquitin-conjugated protein aggregates, elevated apoptosis, and insensitive activation of the RAS-MAPK and PI3K-AKT-mTOR pathways in lymphoblastoid cell lines (LCLs) from patients with *STAMBP* mutations. Our study is the first to provide evidence that *STAMBP* plays an important role in neocortical development

by maintaining neural stem cell proliferation. *STAMBP* deficient organoids showed a reduction in organoid size due to disruption of the proliferation of NSCs. There was no significant difference in apoptosis between *STAMBP* KO hCOs and controls during brain development. Notably, the microcephaly phenotype was not recapitulated in *Stambp* null mice. Ishii et al. (2001) reported that *Stambp*-deficient mice were morphologically indistinguishable from their littermates at birth and assumed that apoptotic activation was the major pathogenic mechanism of *STAMBP* mutation. Although mouse models can capture some clinical features of human conditions, human organoids provide unique functions in the study of human diseases and complement animal models. The pathophysiological mechanism that associates *STAMBP* mutations with microcephaly needs to be studied using multiple suitable models.

In conclusion, we present a Chinese patient with *STAMBP* mutation, and a novel pathogenic compound heterozygosity of the *STAMBP* gene broadening the spectrum of phenotypic presentations associated with *STAMBP* mutations. The results of the present study may improve our understanding of STAMBP function in human cortical development.

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 authors.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of Children's Hospital affiliated to Fudan University (permit no. 2016-131). Written informed consent to participate in this study was provided by the participants or their legal guardian/next of kin. Written informed consent was obtained from the minor(s)' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

Author contributions

HL, WF, and XX: conceptualization and supervision. WF, MH, ZH, YX, and BC: organoid methodology. HL, DL, QX, YW, and JD: clinical data collection and analysis. MH, HL, WF, and XX: investigation. MZ: data curation. MH and HL: writing – original draft. WF and XX: writing – review and editing. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (Grant No. 82171540), the Key Subject Construction Project of the Shanghai Municipal Health Commission (No. shslczdzk02903), the "Thousand Talents Program" to WF, and the Young Clinical Scientist Program of Children's Hospital of Fudan University (2022LCKXJ03).

Acknowledgments

We thank the patients who participated in this study and their families. We thank Dr. Lin Yang and Wei Lu for helping with the clinical data analysis. We also thank all members of the WF's lab for their valuable discussions on this work.

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.963813/full#supplementary-material

SUPPLEMENTARY FIGURE 1
STAMBP deletion did not affect the pluripotency of hESCs. Immunostaining and RT-PCR were done to confirm the pluripotency of H9 human embryonic stem cells (hESCs) before brain organoid generation. **(A)** Immunostaining analysis showed no significant changes in the expression of pluripotent markers NANOG and SOX2 between *STAMBP* KO hESCs and control. Scale bar, 100 μm. **(B)** RT-PCR showed that *STAMBP* KO hESCs have no significant changes in the expression of pluripotent markers SOX2 and REX1 compared to control. Error bars represent SEM of three independent experiments containing numbers of organoids as indicated; ns $p > 0.05$ (Student's *t*-test).

SUPPLEMENTARY FIGURE 2
The variants of patients were successfully inserted. **(A)** The protein structure shows the position of Gly307 and Thr313. **(B)** Sanger sequencing showed the variants c.920G>A (p.Gly307Glu) and c.938C>T (p.Thr313Ile) were inserted.

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

Agromayor, M., and Martin-Serrano, J. (2006). Interaction of AMSH with ESCRT-III and deubiquitination of endosomal cargo. *J. Biol. Chem.* 281, 23083–23091. doi: 10.1074/jbc.M513803200

Bershteyn, M., Nowakowski, T. J., Pollen, A. A., Di Lullo, E., Nene, A., Wynshaw-Boris, A., et al. (2017). Human iPSC-derived cerebral organoids model cellular features of lissencephaly and reveal prolonged mitosis of outer radial glia. *Cell Stem Cell* 20, 435–449.e4. doi: 10.1016/j.stem.2016.12.007

Chen, J. F., Zhang, Y., Wilde, J., Hansen, K. C., Lai, F., and Niswander, L. (2014). Microcephaly disease gene Wdr62 regulates mitotic progression of embryonic neural stem cells and brain size. *Nat. Commun.* 5:3885. doi: 10.1038/ncomms4885

Choi, Y., and Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31, 2745–2747. doi: 10.1093/bioinformatics/btv195

Davies, C. W., Paul, L. N., and Das, C. (2013). Mechanism of recruitment and activation of the endosome-associated deubiquitinase AMSH. *Biochemistry* 52, 7818–7829. doi: 10.1021/bi401106b

Demikova, N. S., Kakaulina, V. S., Pechatnikova, N. L., Polyakova, N. A., Zakharova, E. Y., Krylova, T. D., et al. (2018). First report of microcephaly-capillary malformations syndrome in Russia. *Egypt. J. Med. Hum. Genet.* 19, 147–150. doi: 10.1016/j.ejmhg.2017.08.011

Faqeih, E. A., Bastaki, L., Rosti, R. O., Spencer, E. G., Zada, A. P., Saleh, M. A., et al. (2015). Novel STAMBP mutation and additional findings in an Arabic family. *Am. J. Med. Genet. A.* 167A, 805–809. doi: 10.1002/ajmg.a.36782

Guo, Z., Chen, M., Chao, Y., Cai, C., Liu, L., Zhao, L., et al. (2021). RGCC balances self-renewal and neuronal differentiation of neural stem cells in the developing mammalian neocortex. *EMBO Rep.* 22:e51781. doi: 10.15252/embr.202051781

Hori, I., Miya, F., Negishi, Y., Hattori, A., Ando, N., Boroevich, K. A., et al. (2018). A novel homozygous missense mutation in the SH3-binding motif of STAMBP causing microcephaly-capillary malformation syndrome. *J. Hum. Genet.* 63, 957–963. doi: 10.1038/s10038-018-0482-3

Ishii, N., Owada, Y., Yamada, M., Miura, S., Murata, K., Asao, H., et al. (2001). Loss of neurons in the hippocampus and cerebral cortex of AMSH-deficient mice. *Mol. Cell. Biol.* 21, 8626–8637. doi: 10.1128/MCB.21.24.8626-8637.2001

Johansson, M. U., Zoete, V., Michielin, O., and Guex, N. (2012). Defining and searching for structural motifs using DeepView/Swiss-PdbViewer. *BMC Bioinform.* 13:173. doi: 10.1186/1471-2105-13-173

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

Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., et al. (2013). Cerebral organoids model human brain development and microcephaly. *Nature* 501, 373–379. doi: 10.1038/nature12517

Lm, S. (2019). Case of resistant epileptic encephalopathy a child with microcephalic capillary malformation syndrome. *Res. Pediatr. Neonatol.* 3, 243–248. doi: 10.31031/RPN.2019.03.000564

Manzini, M. C., and Walsh, C. A. (2011). What disorders of cortical development tell us about the cortex: One plus one does not always make two. *Curr. Opin. Genet. Dev.* 21, 333–339. doi: 10.1016/j.gde.2011.01.006

McDonell, L. M., Mirzaa, G. M., Alcantara, D., Schwartzentruber, J., Carter, M. T., Lee, L. J., et al. (2013). Mutations in STAMBP, encoding a deubiquitinating enzyme, cause microcephaly-capillary malformation syndrome. *Nat. Genet.* 45, 556–562. doi: 10.1038/ng.2602

Naseer, M. I., Sogaty, S., Rasool, M., Chaudhary, A. G., Abutalib, Y. A., Walker, S., et al. (2016). Microcephaly-capillary malformation syndrome: Brothers with a homozygous STAMBP mutation, uncovered by exome sequencing. *Am. J. Med. Genet. A.* 170, 3018–3022. doi: 10.1002/ajmg.a.37845

Pavlović, M., Neubauer, D., Al Tawari, A., and Heberle, L. C. (2014). The microcephaly-capillary malformation syndrome in two brothers with novel clinical features. *Pediatr. Neurol.* 51, 560–565. doi: 10.1016/j.pediatrneurol.2014.07.006

Pirozzi, F., Nelson, B., and Mirzaa, G. (2018). From microcephaly to megalencephaly: Determinants of brain size. *Dialogues Clin. Neurosci.* 20, 267–282. doi: 10.31887/DCNS.2018.20.4/gmirzaa

Quadrato, G., Nguyen, T., Macosko, E. Z., Sherwood, J. L., Min Yang, S., Berger, D. R., et al. (2017). Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* 545, 48–53. doi: 10.1038/nature22047

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res.* 31, 3381–3385. doi: 10.1093/nar/gkg520

Suzuki, S., Tamai, K., Watanabe, M., Kyuuma, M., Ono, M., Sugamura, K., et al. (2011). AMSH is required to degrade ubiquitinated proteins in the central nervous system. *Biochem. Biophys. Res. Commun.* 408, 582–588. doi: 10.1016/j.bbrc.2011.04.065

Tanaka, N., Kaneko, K., Asao, H., Kasai, H., Endo, Y., Fujita, T., et al. (1999). Possible involvement of a novel STAM-associated molecule "AMSH" in intracellular signal transduction mediated by cytokines. *J. Biol. Chem.* 274, 19129–19135. doi: 10.1074/jbc.274.27.19129

Thornton, G. K., and Woods, C. G. (2009). Primary microcephaly: Do all roads lead to Rome? *Trends Genet.* 25, 501–510. doi: 10.1016/j.tig.2009.09.011

Tidyman, W. E., and Rauen, K. A. (2009). The RASopathies: Developmental syndromes of Ras/MAPK pathway dysregulation. *Curr. Opin. Genet. Dev.* 19, 230–236. doi: 10.1016/j.gde.2009.04.001

Wright, M. H., Berlin, I., and Nash, P. D. (2011). Regulation of endocytic sorting by ESCRT-DUB-mediated deubiquitination. *Cell Biochem. Biophys.* 60, 39–46. doi: 10.1007/s12013-011-9181-9

Wu, F., Dai, Y., Wang, J., Cheng, M., Wang, Y., Li, X., et al. (2019). Early-onset epilepsy and microcephaly-capillary malformation syndrome caused by a novel STAMBP mutation in a Chinese boy. *Mol. Med. Rep.* 20, 5145–5151. doi: 10.3892/mmr.2019.10757

Xiang, Y., Tanaka, Y., Patterson, B., Kang, Y. J., Govindaiah, G., Roselaar, N., et al. (2017). Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. *Cell Stem Cell* 21, 383–398.e7. doi: 10.1016/j.stem.2017.07.007

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 12 August 2022
DOI 10.3389/fgene.2022.901228

OPEN ACCESS

EDITED BY
Davide Vecchio,
Bambino Gesù Children's Hospital (IRCCS), Italy

REVIEWED BY
Davide Gabellini,
San Raffaele Hospital (IRCCS), Italy
Viviana Caputo,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Holly A. F. Stessman,
hollystessman@creighton.edu

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 21 March 2022
ACCEPTED 11 July 2022
PUBLISHED 12 August 2022

CITATION
Hulen J, Kenny D, Black R, Hallgren J, Hammond KG, Bredahl EC, Wickramasekara RN, Abel PW and Stessman HAF (2022), KMT5B is required for early motor development. *Front. Genet.* 13:901228. doi: 10.3389/fgene.2022.901228

KMT5B is required for early motor development

Jason Hulen[1], Dorothy Kenny[1], Rebecca Black[1], Jodi Hallgren[1], Kelley G. Hammond[2], Eric C. Bredahl[2], Rochelle N. Wickramasekara[1,3], Peter W. Abel[1] and Holly A. F. Stessman[1]*

[1]Department of Pharmacology and Neuroscience, School of Medicine, Creighton University, Omaha, NE, United States, [2]Department of Exercise Science, College of Arts and Sciences, Creighton University, Omaha, NE, United States, [3]Molecular Diagnostic Laboratory, Boys Town National Research Hospital, Omaha, NE, United States

Disruptive variants in lysine methyl transferase 5B (KMT5B/SUV4-20H1) have been identified as likely-pathogenic among humans with neurodevelopmental phenotypes including motor deficits (i.e., hypotonia and motor delay). However, the role that this enzyme plays in early motor development is largely unknown. Using a *Kmt5b* gene trap mouse model, we assessed neuromuscular strength, skeletal muscle weight (i.e., muscle mass), neuromuscular junction (NMJ) structure, and myofiber type, size, and distribution. Tests were performed over developmental time (postnatal days 17 and 44) to represent postnatal versus adult structures in slow- and fast-twitch muscle types. Prior to the onset of puberty, slow-twitch muscle weight was significantly reduced in heterozygous compared to wild-type males but not females. At the young adult stage, we identified decreased neuromuscular strength, decreased skeletal muscle weights (both slow- and fast-twitch), increased NMJ fragmentation (in slow-twitch muscle), and smaller myofibers in both sexes. We conclude that *Kmt5b* haploinsufficiency results in a skeletal muscle developmental deficit causing reduced muscle mass and body weight.

KEYWORDS

H4K20, KMT5B, SUV420H, neuromuscular junction, neuromuscular development, histone methylation, hypotonia, hypertrophy

Introduction

High-throughput sequencing has identified lysine methyl transferase 5B, *KMT5B* (also known as *SUV4-20H1*), as a high impact neurodevelopmental disorder risk gene in humans (Stessman et al., 2017; Trinh et al., 2019). Primary *KMT5B* patient phenotypes include language delay, intellectual disability, autism spectrum disorder, and motor disorders (i.e., hypotonia and motor delay) (Faundes et al., 2017; Stessman et al.,

Abbreviations: EDL, extensor digitorum longus; HET, heterozygous Kmt5b+/GT; NMJ, neuromuscular junction; SOL, soleus; TA, tibialis anterior; WT, wild-type Kmt5b+/+

2017; Trinh et al., 2019). The KMT5B protein catalyzes the dimethylation (me2) of lysine 20 on histone 4 tails (H4K20), which is thought to be the primary substrate for H4K20 trimethylation (me3) by KMT5C (Schotta et al., 2004; Wu et al., 2013). H4K20 methylation affects chromatin conformation, altering chromatin accessibility (Jenuwein and Allis, 2001; Shoaib et al., 2021). H4K20me3 marks are co-localized to pericentric heterochromatin in the genome and are associated with chromatin compaction and gene silencing. Pericentric chromatin aggregation has been positively correlated with myogenic differentiation (Biron et al., 2004; Terranova et al., 2005; Tsang et al., 2010), and previous research has identified a role for KMT5B in the maintenance of the adult muscle satellite cell pool (Boonsanay et al., 2016). These studies strongly support the hypothesis that KMT5B activity contributes to muscle formation and health (Neguembor et al., 2013; Boonsanay et al., 2016). Early skeletal muscle development has not been previously explored in the context of KMT5B deficiencies and was, therefore, the focus of this study. We show that constitutive Kmt*5b* haploinsufficiency in a mouse model results in decreased neuromuscular strength, reduced muscle weight (i.e., muscle mass), and changes in myofiber pathology at both young adolescent and young adult stages of development resembling the motor symptoms noted among the *KMT5B* patient population. These deficits are in addition to the motor reflex and body weight deficits we have previously reported in this model (Wickramasekara et al., 2021). Our results represent data collected from two distinct muscles, the soleus (SOL) and extensor digitorum longus (EDL). Both muscles are located on the lower hind limb but are characterized as slow-twitch vs. fast-twitch, respectively, due to differences in myofiber type composition between the muscles. Given sex differences noted previously (Wickramasekara et al., 2021), we compared our results between both sexes and genotypes. We found that *Kmt5b* heterozygous males were affected earlier and more significantly than females. We propose a model in which motor delays present in *KMT5B* patients are due to changes in skeletal muscle maturation.

Materials and methods

Ethical statement

All mouse work was approved and monitored by the Creighton University Institutional Animal Care and Use Committee (IACUC) under protocol numbers 1039 and 1040.

Mouse model

Kmt5b gene-trap mice (Kmt5b$^{tm1a(KOMP)Wtsi\ flox-/+}$), referred to here as "*Kmt5b*$^{+/GT}$," are commercially available (KOMP, University of California, Davis, CA, United States). This mouse model contains a LacZ gene trap inserted between exon 4 and exon 5. Mice carrying one copy of the gene trap are considered germline haploinsufficient (Wickramasekara et al., 2021). All experiments were performed blind to mouse genotype.

RT-qPCR

SOL and EDL skeletal muscles were dissected from mice at 17 days old (P17) and stored in RNA*later* (ThermoFisher Scientific, Waltham, MA, United States). RNA extraction was performed using the RNeasy Fibrous Tissue Mini Kit (Qiagen, Hilden, Germany). Synthesis of cDNA was performed with the iScript cDNA Synthesis Kit (BioRad, Hercules, CA, United States) on a BioRad C1000 Touch Thermocycler followed by quantitative PCR using SsoAdvanced Universal SYBR Green Supermix (BioRad) and a BioRad CFX Connect Real Time System. PrimePCR SYBR Green Assay primers specific for β-Actin (*Actb*; qMmuCED0027505), *Kmt5b* (*Suv420h1*; qMmuCID0014487), and *Kmt5c* (*Suv420h2*; qMmuCID0020579) were sourced from BioRad. Primers for mouse *Gm16066* were designed by and purchased from IDT (Coralville, IA, United States); F-5′-GGCAATACCAGAGGAGAAAGAC-3′and R-5′-CACAGAGAACCCTTGTCTCAAA-3'. The *Gm16066* synthetic control sequence was also purchased from IDT; 5′-TGTAGGCAA TACCAGAGGAGAAAGACAGCATCGTTGTCTTGGTTTGTT TGATTTTAATTAATTAATTAATTAATTAATTATTTTGAG ACAAGGGTTCTCTGTGCCCT-3′.

Neuromuscular strength

Mice were held by the tail and allowed to grasp a triangular bar attached to a grip strength meter (Columbus Instruments, Columbus, OH, United States) with their forearm paws. The mouse was pulled *via* the tail in a parallel, opposite direction away from the force meter until the bar was released, and the peak force generated was recorded. Data presented represent an average of three trials per mouse with at least 5 min of rest between trials.

Neuromuscular junction immunofluorescence

Bilateral SOL and EDL muscles were dissected following perfusion with 4% paraformaldehyde. Muscles were stored in 4% paraformaldehyde at 4°C until further use. Muscles were rinsed in wash solution (2% Triton X-100 in phosphate buffered saline) for 30 min followed by incubation in a blocking solution consisting of 4% bovine serum albumin and 1% Triton X-100 in phosphate

buffered saline. Primary incubation was performed overnight in blocking solution at 4°C using mouse IgG antibodies specific for synaptic vesicle glycoprotein 2A (SV2) and neurofilament medium (2H3; Developmental Studies Hybridoma Bank, Iowa City, IA, United States) at concentrations of 0.29 µg/ml and 0.42 µg/ml, respectively. Following primary incubation, muscles were washed three times for 15 min each in wash solution. Secondary incubation was performed in blocking solution for 2 h at room temperature using a goat anti-mouse IgG Alexa Fluor 488 antibody (ThermoFisher Scientific) and tetramethylrhodamine α-bungarotoxin at concentrations of 8.0 µg/ml and 1.5 µg/ml, respectively. Following three final washes for 15 min each, the muscles were placed on slides and cover slipped after adding Vectashield mounting medium (Vector Laboratories, Burlingame, CA, United States). Maximum intensity projection of a Z-stacked image was achieved using a Nikon Ti-E inverted microscope with a Yokagawa Spinning Disc at ×40 magnification. Images were processed using ImageJ 1.52a software (Schneider et al., 2012) with the binaryconnectivity. class plugin (Landini, 2020). Neuromuscular junction images were processed according to previously published methods to obtain total area, discontinuity (1 adjacent pixel), and branch points (3 + adjacent pixels) for both muscle and nerve/synaptic vesicle areas (Pratt et al., 2018).

Myosin heavy chain immunofluorescence

Mice were sacrificed using CO_2 asphyxiation and the left SOL and EDL were dissected, frozen in liquid nitrogen, and stored at −80°C. Frozen muscles were embedded in OCT compound (ThermoFisher Scientific), cryosectioned near the mid-belly at 5 µm thickness, and collected onto slides. Frozen sections were thawed to room temperature for 15 min immediately prior to the immunofluorescence procedure. Sections were washed in 0.5% Triton X-100 in phosphate buffered saline for 30 min. Slides were then incubated in a humidified chamber at room temperature with a blocking solution consisting of 20% normal goat serum and 0.5% Triton X-100 in phosphate buffered saline for 1 h. Primary antibodies specific for myosin heavy chain type I (mouse IgG2b), myosin heavy chain IIa (mouse IgG1), myosin heavy chain IIb (mouse IgM) (BA-D5, SC-71, and BF-F3, respectively; Developmental Studies Hybridoma Bank), and laminin (rabbit IgG; Sigma-Aldrich, St. Louis, MO, United States) were prepared in blocking solution at concentrations of 7.625 µg/ml BA-D5, 1.935 µg/ml SC-71, 2.015 µg/ml BF-F3, and 1 µg/ml laminin, and incubated for 1 h at room temperature in a humidified chamber. After a 30-min wash, secondary antibody incubation was performed with 20 µg/ml goat anti-mouse IgG2b Alexa Fluor 647, 2.6 µg/ml goat anti-mouse IgG1 Alexa Fluor 546, 5.7 µg/ml goat anti-mouse IgM Alexa Fluor 488, and 4.0 µg/ml goat anti-rabbit IgG Alexa Fluor 405 (ThermoFisher Scientific) in blocking solution for 2 hours at room temperature in a humidified chamber. Following a final 30-min wash, sections had Vectashield medium added and were cover slipped. Imaging was achieved using a Nikon Ti-E inverted microscope with a Yokagawa Spinning Disc at ×10 magnification. Images were processed and analyzed using ImageJ 1.52a software (Schneider et al., 2012) for total area and immunofluorescence area. The number of myofiber cells for each antibody was counted manually and blinded.

Data analyses

All statistical calculations were performed using GraphPad Prism version 8.3.0 for Windows (GraphPad Software, San Diego, CA, United States). All primary data were analyzed using two-way ANOVAs with independent variables of genotype and sex. Mouse body weight was analyzed using a two-way (genotype x age) ANOVA with repeated measures. Where indicated, a post hoc test, Šídák's test, Tukey's multiple comparisons test, or Fisher's LSD, was used to further identify specific relationships. Pearson's correlation tests were used to test specific pairs of continuous variables where noted. Genotype effects of the mother (i.e., dam) were analyzed using three-way ANOVAs with independent variables of dam genotype, mouse genotype, and sex.

Results

Kmt5b HET mice exhibit neuromuscular strength deficits consistent with human hypotonia

Motor delays and hypotonia have been frequently reported in humans carrying germline heterozygous disruptive *KMT5B* variation (Trinh et al., 2019; Wickramasekara and Stessman, 2019). Previously, we reported early motor reflex deficits using a LacZ gene trapping allele (Wickramasekara et al., 2021) to mimic germline heterozygous *Kmt5b* loss in mice (*Kmt5b*$^{+/GT}$; referred to as "*HET*"). To better understand the contribution of KMT5B to early muscle development and function, we have further utilized this mouse model to specifically study the effects of *Kmt5b* loss on hindlimb skeletal muscle development. Homozygous *Kmt5b* knockout is perinatally lethal (Schotta et al., 2008; Wickramasekara et al., 2021); therefore, only wild-type (*WT*) and heterozygous (*HET*) genotypes were studied. Quantitative RT-PCR on mRNA extracted from P17 soleus (SOL) and extensor digitorum longus (EDL) hindlimb muscles confirmed that *Kmt5b* is expressed in skeletal muscle during development and that gene trapping on one allele (i.e., *HET*) reduced this expression by approximately

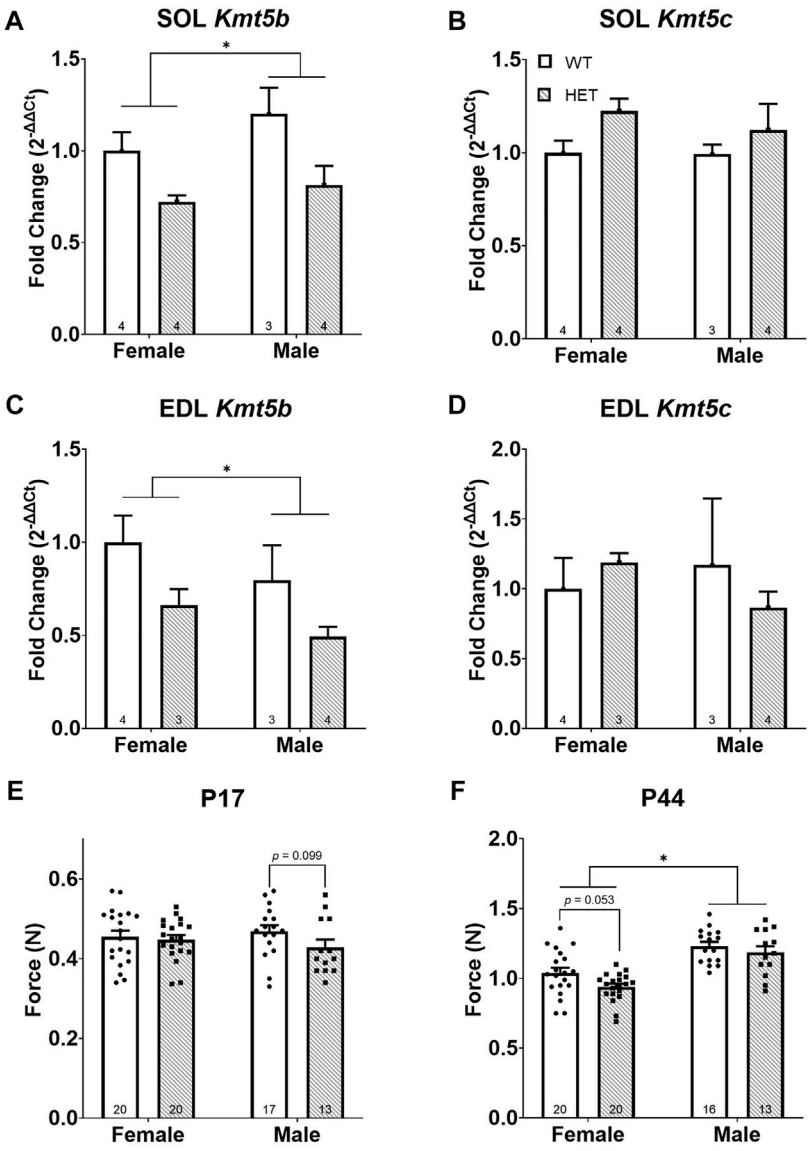


FIGURE 1
Kmt5b HET mice show a progressive strength deficit. Gene expression of *Kmt5b* **(A,C)** and *Kmt5c* **(B,D)** measured by qRT-PCR shown for soleus [SOL; **(A,B)**] and extensor digitorum longus [EDL; **(C,D)**] muscles at postnatal day (P) 17. Results of pull bar strength testing are shown collected at P17 **(E)** and P44 **(F)**. *WT*: dots and open bars; *HET*: squares and grey hatched bars. *$p < 0.05$; (2-way ANOVA). *p*-values shown are the results of post hoc testing. Error bars show ±SEM; number of biological replicates are shown within each bar.

half in both sexes, as expected (Figures 1A,C). *Kmt5c* expression showed no changes in expression by genotype or sex (Figures 1B,D).

To compare strength between *WT* and *HET* animals, we performed a neuromuscular "pull bar" strength test (Smith et al., 1995) on postnatal (P) days 17 and P44. These time points were chosen to most closely mimic the neonatal human infant state (Draeger et al., 1987; Butler-Browne et al., 1990) and early sexual maturity, respectively. We observed no significant differences in neuromuscular strength in *HET* mice at age P17 (Figure 1E). At P44, as expected, sex had a significant effect on neuromuscular strength [$F(1, 65) = 41.14$; $p < 0.0001$; Figure 1F] where it was increased in males compared to females. Genotype significantly impacted strength where *HET* mice showed less neuromuscular strength than *WT* mice [$F(1, 65) = 4.550$; $p = 0.0367$]; however, neither sex survived post hoc testing (Figure 1F). Body weight was significantly decreased in both *HET* males [$F(1, 27) = 36.58$; $p < 0.0001$] and females [$F(1, 38) = 10.80$; $p = 0.0022$] compared to their *WT* counterparts (data not shown), consistent with our previous work

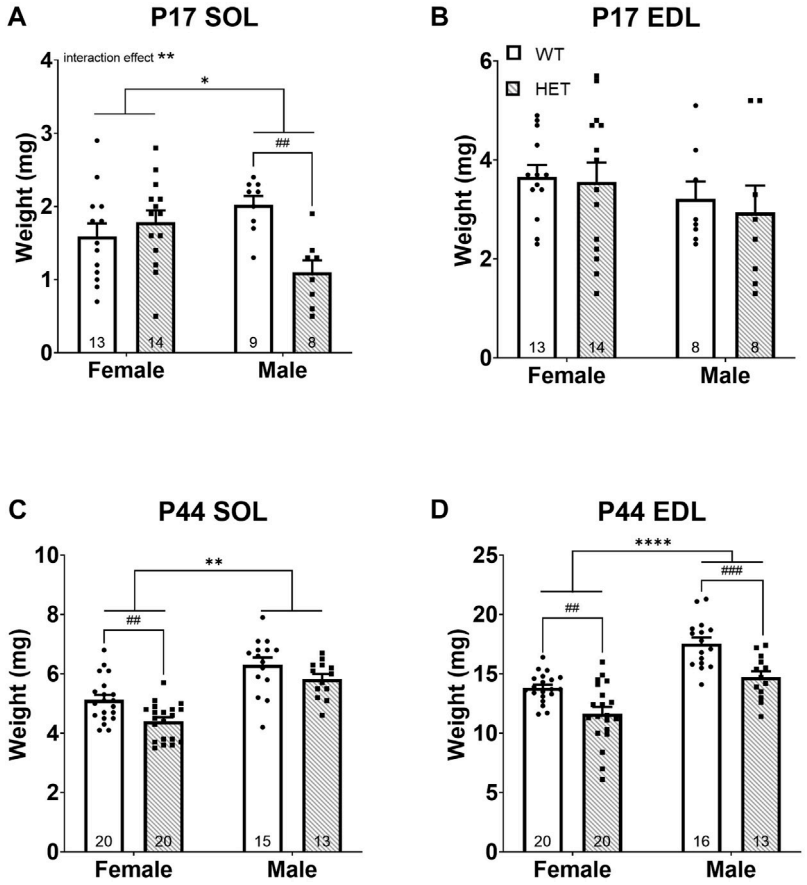


FIGURE 2
HET mice have decreased skeletal muscle weights (i.e., muscle mass). Skeletal muscles weight at ages P17 and P44 for SOL **(A,C)** and EDL **(B,D)** are shown. *$p < 0.05$; **$p < 0.01$; ****$p < 0.0001$ (2-way ANOVA). ##$p < 0.01$; ###$p < 0.001$ (post hoc test; effect of genotype only). Error bars show ±SEM; number of biological replicates are shown within each bar.

(Wickramasekara et al., 2021). Neuromuscular strength and body weight were only correlated consistently among *WT* females (at both P17 and P44) but not among males (Supplementary Figure S1). These results supported the hypothesis that *Kmt5b* loss is directly associated with hypotonia (i.e., reduced muscle tone).

Progressive muscle weakness is accompanied by decreasing muscle weights

Given the known role of KMT5B in adult muscle regeneration (Boonsanay et al., 2016), we hypothesized that the neuromuscular strength deficits observed were due to a muscle-specific requirement for KMT5B in early development. The SOL and EDL muscles were chosen for further study as they are both located on the lower hind limb but represent two distinct skeletal muscle populations; the SOL represents a slow-twitch (type I) muscle and the EDL represents a fast-twitch (type II) muscle.

At P17, SOL weights were significantly impacted by both genotype [F (1, 40) = 4.544; p = 0.0392] and an interaction between sex and genotype [F (1, 40) = 10.65; p = 0.0023]. Post hoc testing showed that this effect was driven by a significant decrease in SOL weight in male *HET* compared to *WT* animals (p = 0.0072; Tukey's multiple comparisons test; Figure 2A). There were no differences in EDL weight at P17 (Figure 2B). At P44, both SOL [F (1, 64) = 10.99; p = 0.0015; Figure 2C] and EDL [F (1, 65) = 26.65; $p < 0.0001$; Figure 2D] weights were significantly impacted by genotype. In both muscle types, *HET* muscles weighed less than in *WT* animals. This effect survived post hoc testing in females for both the SOL (p = 0.0057; Šídák's test) and EDL (p = 0.0018; Šídák's test; Figures 2C,D) and was even more significant in males for the EDL (p = 0.0006; Šídák's test; Figure 2D). As expected, there was also a significant

difference between sexes after puberty (i.e., P44); males had increased muscle weights overall [SOL: F (1, 64) = 51.06; p < 0.0001; EDL: F (1, 65) = 48.89; p < 0.0001; Figures 2C,D].

Muscle weight (i.e., lean muscle mass) is a known significant contributor to body weight and was also significantly correlated with body weight in this study (Supplementary Figure S2). Given the significantly reduced body weights in *HET* compared to *WT* mice, we compared muscle weights after normalization to body weight at the time of collection. Normalized P17 SOL muscle weights showed a significant interaction between sex and genotype [F (1, 40) = 12.55; p = 0.0010; Supplementary Figure S3A driven by strikingly lower muscle weight in *HET* males compared to *WT* males (p = 0.0037; Šídák's test) but no differences in normalized P17 EDL (Supplementary Figure S3B). Therefore, male *HET* SOL muscles were even smaller than would be predicted for their body weight. To determine whether this effect was SOL specific, additional organs (brain, heart, kidney, and liver) were dissected and weighed at P17 (Supplementary Figure S4). While liver (Supplementary Figure S4A), heart (Supplementary Figure S4C), and kidney (Supplementary Figure S4E) weights trended smaller in *HET* males at P17, the liver was most significantly impacted (p = 0.0493; Šídák's test). However, none of these other organs were significantly smaller in *HET* animals after normalization to body weight (Supplementary Figures S4B,D,F,H) suggesting a specific effect on SOL muscles at this age. Interestingly, male *HET* brains were significantly heavier than *WT* animals at P17 after normalization to body weight (Supplementary Figure S4H). There were no significant differences in normalized P44 SOL weights (Supplementary Figure S3C). A significant effect of genotype was noted among normalized P44 EDL muscle weights [F (1, 65) = 5.747; p = 0.0194; Supplementary Figure S3D]. Female *HET* normalized EDL weights remained modestly lower after post hoc testing (p = 0.0445; Šídák's test). Taken together, these data highlighted decreased neuromuscular strength and muscle mass among *HET* animals affecting both sexes that worsened with age. Loss of slow-twitch (SOL) muscle mass was only apparent in *HET* males at P17 but in all *HET* animals in both muscle types at P44. Fast-twitch (EDL) muscle was more significantly affected as time progressed; this effect was more pronounced among *HET* females.

KMT5B haploinsufficiency increases neuromuscular junction fragmentation by young adulthood

The major functional unit of skeletal muscle activity is the neuromuscular junction (NMJ). NMJ pathology has been shown to underlie other strength deficits (Souayah et al., 2009; Chai et al., 2011; Klooster et al., 2012; Rudolf et al., 2013; Haddix et al., 2018). Given the strength deficit identified at P44 in our model (Figure 1F), we hypothesized that this was caused by NMJ dysfunction. Both NMJ size and continuity were scored for the nerve/vesicular and muscle portions of the NMJ using immunofluorescence on P44 SOL and EDL muscles (Supplementary Figures S5A,B). NMJ occupancy (defined as the ratio of nerve/vesicular area vs. muscular NMJ area) did not significantly differ in the SOL (Supplementary Figure S5C) or the EDL (Supplementary Figure S5F). No significant differences were identified in nerve/vesicular NMJ area or continuity in either muscle (SOL: Supplementary Figures S5D,E; EDL: Supplementary Figures S5G,H). In the muscle portion of the NMJ (Figure 3), SOL NMJ areas did not differ (Figure 3A), but genotype contributed significantly to discontinuity [F (1, 35) = 6.447; p = 0.0157; Figure 3B] where *HET* NMJs had increased discontinuity compared to *WT* NMJs. Sex, but not genotype, contributed significantly to NMJ areas in the EDL [F (1, 31) = 5.684; p = 0.0234; Figure 3C]; area was increased in males compared to females. However, neither sex survived post hoc testing. A trend toward increased discontinuity among *HETs* was noted in the EDL (Figure 3D) but was not statistically significant. These data showed that increased NMJ fragmentation accompanies decreased neuromuscular strength and reduced muscle mass at young adult stages in *HET* mice.

Fast-twitch myofibers are increased in number but decreased in size in Kmt5b HET mice

Proper motor neuron innervation at the NMJ is required for skeletal muscle myofiber maturation (Jin et al., 2008) and maintenance (Rowan et al., 2012). Myofibers are the predominant structural unit comprising skeletal muscle (Schiaffino and Reggiani, 1985). Changes in myofiber composition have also been linked to strength deficits (Pedemonte et al., 1999; Tanner et al., 2002; Celegato et al., 2006; Chai et al., 2011; Nilwik et al., 2013). Given the increased muscular NMJ fragmentation, reduced neuromuscular strength, and decreased muscle mass among *HET* mice, we analyzed myofiber abundance and size. Both slow- and fast-twitch muscles are composed of type I, IIa, IIx, and IIb fibers at differing ratios (Schiaffino and Reggiani, 1985). Myofiber types can be distinguished through their myosin heavy chain expression (Termin et al., 1989).

We first quantified the relative myofiber number for myofiber types I, IIa, and IIb in SOL and EDL muscles (Figures 4A–L). For the slow-twitch SOL muscle (Figures 4A–F), there were no differences in myofiber number at P17 (Figures 4A–C). In young adult animals (P44; Figures 4D–F), there was a significant effect of sex on both type I [F (1, 43) = 13.85; p = 0.0006; Figure 4D] and type IIa [F (1, 58) = 17.16; p = 0.0001; Figure 4E] myofibers; males had fewer of these myofibers than females. There was also a significant effect of genotype on type IIa myofiber number [F (1, 58) = 5.622, p = 0.021; Figure 4E].

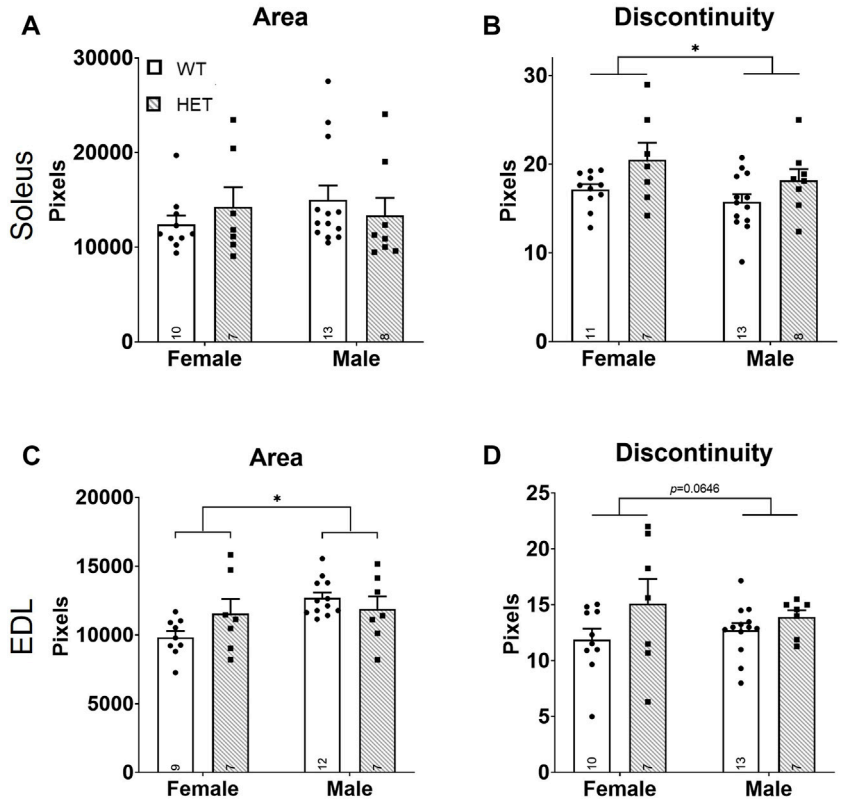


FIGURE 3
Increased NMJ fragmentation in *HET* skeletal muscles. Whole mount immunofluorescence staining for both nerve/vesicle and skeletal muscle portions of NMJs was performed at P44. Only data for skeletal muscle NMJ portions are shown. Graphs show total area **(A,C)** and calculated NMJ discontinuity **(B,D)** for SOL **(A,B)** and EDL **(C,D)** skeletal muscles. Y-axes of graphs are numbers of pixels. *WT*: dots and open bars; *HET*: squares and grey hatched bars. *$p < 0.05$ (2-way ANOVA). Error bars show ±SEM; number of biological replicates are shown within each bar.

Post hoc testing showed that this effect was driven by an increase in type IIa fibers in *HET* compared to *WT* females ($p = 0.0051$; Šídák's test). For the fast-twitch EDL muscle (Figures 4G–L), there was a significant effect of sex on type I [$F(1, 49) = 4.247$; $p = 0.0447$; Figure 4G] and type IIa [$F(1, 48) = 5.118$; $p = 0.0282$; Figure 4H] myofiber numbers at P17 and type I fibers at P44 [$F(1, 48) = 8.712$; $p = 0.0049$; Figure 4J] where, again, male counts were decreased compared to females. Genotype contributed significantly to type IIb myofiber number at P44 [$F(1, 53) = 17.17$, $p = 0.0001$; Figure 4L]. Šídák's post-hoc analysis identified significantly increased IIb myofiber numbers in both female and male *HET* mice compared to *WT* ($p = 0.018$ and $p = 0.0054$, respectively).

The total fluorescent signal for each myofiber marker was collected and normalized to the total pixel area of the section to identify the percent area for each myofiber type (Supplementary Figures S6A–L). Significant effects of sex were identified for P44 SOL type I [$F(1, 43) = 4.475$; $p = 0.0402$; Supplementary Figure S6D] and IIb [$F(1, 59) = 9.624$; $p = 0.0029$; Supplementary Figure S6F] myofibers and P44 EDL type I [$F(1, 47) = 11.61$; $p = 0.0014$; Supplementary Figure S6J] and IIa [$F(1, 55) = 5.525$; $p = 0.0224$; Supplementary Figure S6K] myofiber areas. A significant interaction between sex and genotype was noted in P17 SOL type IIb myofibers [$F(1, 42) = 4.107$, $p = 0.0491$; Supplementary Figure S6C) that was driven by an increase in type IIb myofibers in female *HET* compared to *WT* ($p = 0.0439$; post hoc uncorrected Fisher's LSD). Genotype significantly affected the percent area of P44 EDL type IIb [$F(1, 53) = 6.323$; $p = 0.0150$; Supplementary Figure S6L] myofibers; however, this result did not survive post hoc testing. Both cell count and percent area by myofiber type were used to calculate a myofiber relative ratio for each by sex and genotype (Supplementary Figure S7). While myofiber compositions were as expected (e.g., EDL with a majority of fast-twitch and SOL with few fast-twitch myofibers (Agbulut et al., 2003); Supplementary Figures S7B,C), no significant differences among relative proportions [either by count (Supplementary Figure S7B) or area (Supplementary Figure S7C)] were identified between genotype or sex.

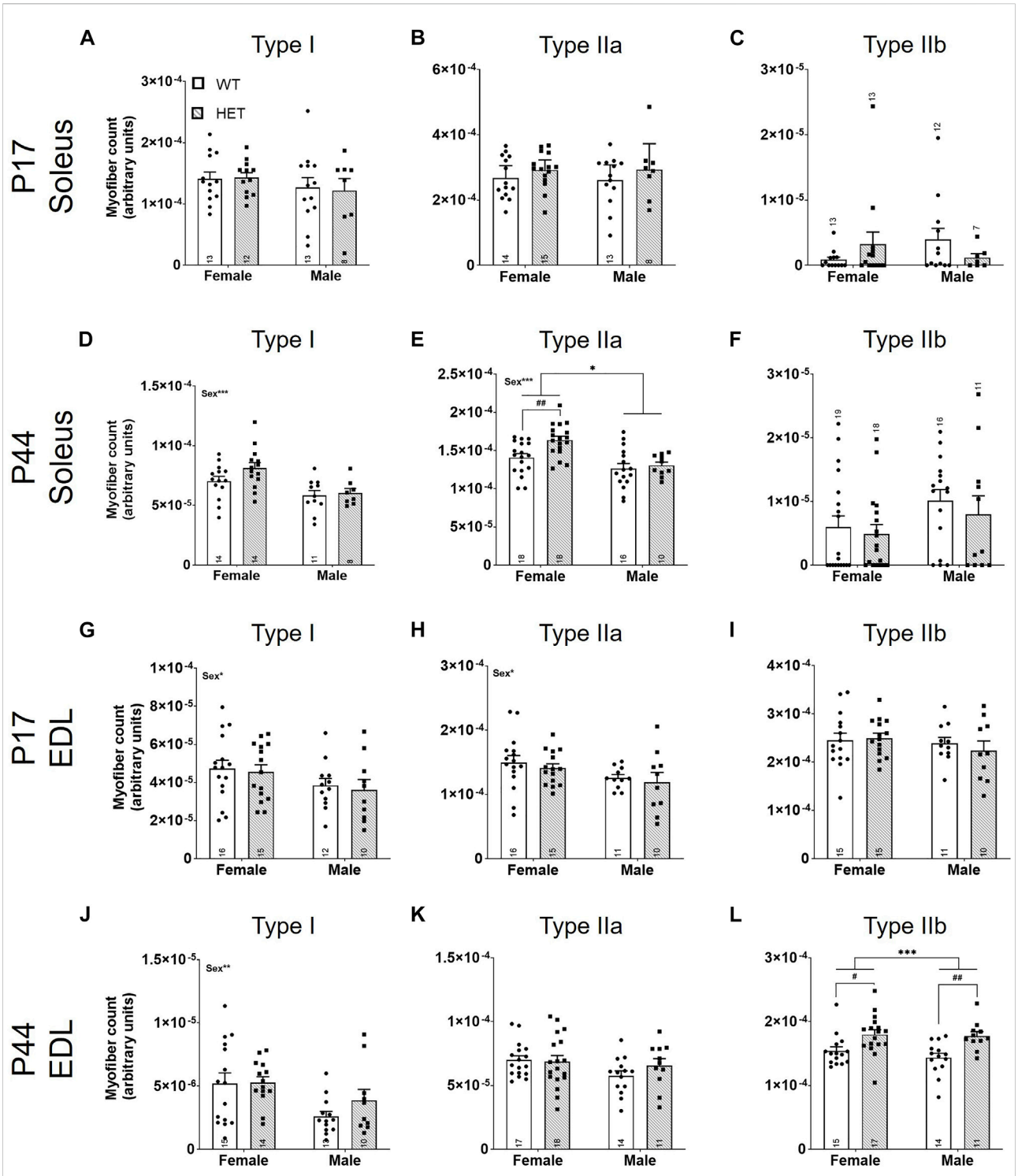


FIGURE 4
Fast-twitch myofiber counts are increased in *HET* muscles at P44. Myofiber types (I, IIa, and IIb) were scored using immunofluorescence. Individual myofibers per type were counted and normalized to the total area of the section (represented as myofiber count; y-axes). Graphs show each myofiber type in males and females from P17 SOL **(A–C)**, P44 SOL **(D–F)**, P17 EDL **(G–I)**, and P44 EDL **(J–L)**. *WT*: dots and open bars; *HET*: squares and grey hatched bars. $*p < 0.05$; $**p < 0.01$; $***p < 0.001$ (2-way ANOVA). $\#p < 0.05$; $\#\#p < 0.01$ (post hoc test; effect of genotype only). Effects of sex by 2-way ANOVA testing are shown as text in the upper left of the panel. Error bars show ±SEM; number of biological replicates are shown within each bar.

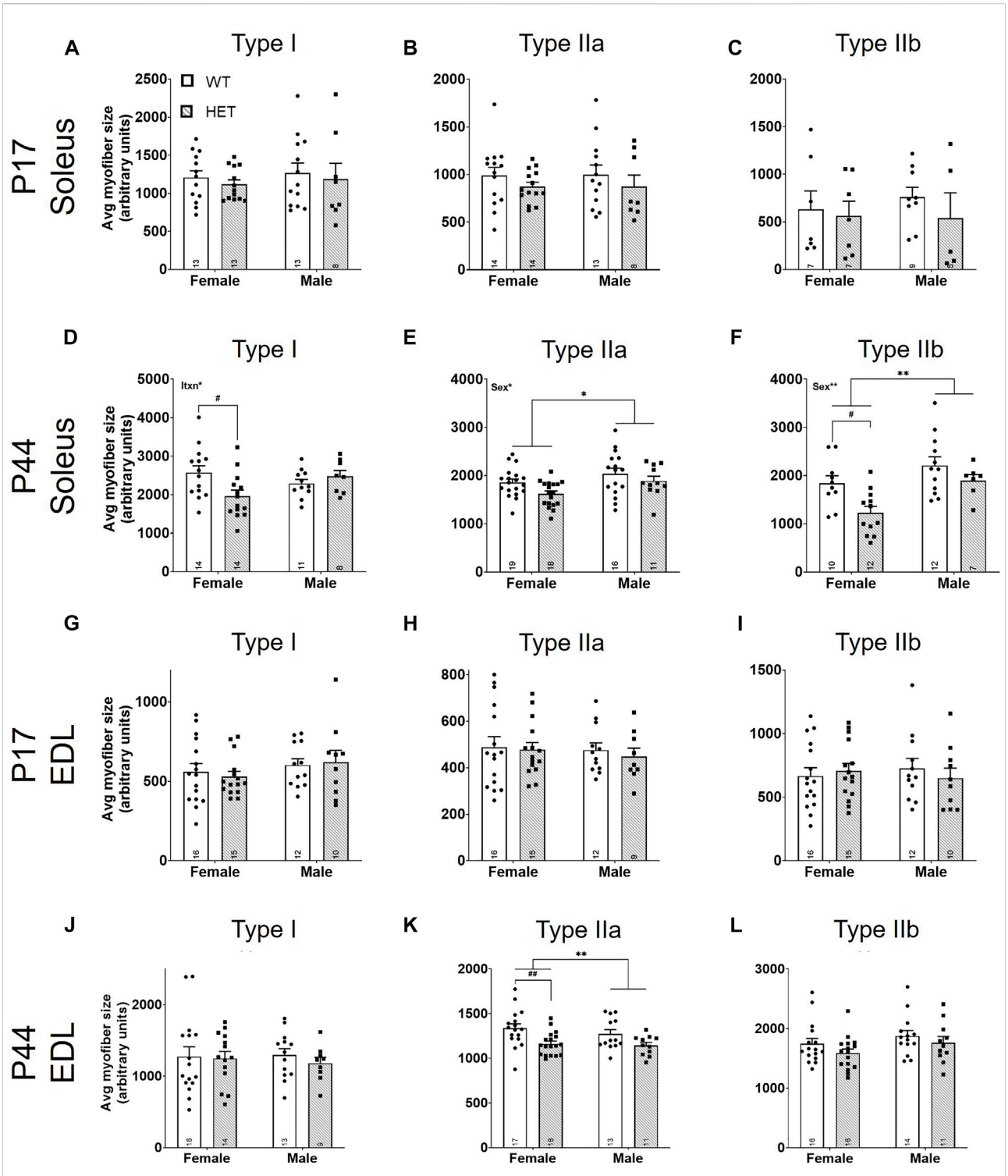


FIGURE 5
Average fast-twitch myofiber size is decreased at P44. For myofiber types (I, IIa, and IIb), average size was scored using the total myofiber type area by immunofluorescence/myofiber type count (Figure 4; y-axes). Graphs show results of each myofiber type in males and females from P17 SOL **(A–C)**, P44 SOL **(D–F)**, P17 EDL **(G–I)**, and P44 EDL **(J–L)**. *WT*: dots and open bars; *HET*: squares and grey hatched bars. *$p < 0.05$; **$p < 0.01$ (2-way ANOVA). #$p < 0.05$; ##$p < 0.01$ (post hoc test; effect of genotype only). Effects of sex or an interaction (itxn) between sex and genotype by 2-way ANOVA testing are shown as text in the upper left of the panel. Error bars show ±SEM; number of biological replicates are shown within each bar.

Using the percent area for each myofiber type and the myofiber count for each section, we calculated an average size for each myofiber type (Figures 5A–L). There were no significant differences between myofiber sizes in the SOL (Figures 5A–C) or the EDL (Figures 5G–I) at P17. At P44, there were significant effects of sex on the size of SOL type IIa myofibers [F (1, 60) = 6.569; p = 0.0129; Figure 5E] and IIb myofibers [F (1, 37) = 10.30; p = 0.0027; Figure 5F]; male myofibers were larger. SOL type I myofibers had a significant interaction between sex and genotype [F (1, 43) = 6.089; p = 0.0177; Figure 5D]. By post hoc testing, females had significantly smaller average SOL type I myofiber sizes in *HET* mice compared to *WT* mice (p = 0.0266; Šídák's test). Genotype contributed significantly to average SOL type IIa [F (1, 60) = 4.995; p = 0.0292; Figure 5E], SOL type IIb [F (1, 37) = 8.260; p = 0.0067; Figure 5F], and EDL type IIa myofiber sizes [F (1, 55) = 12.00, p = 0.0010; Figure 5K]. Šídák's post-hoc analyses identified that female *HETs* had significantly decreased SOL type IIb (p = 0.0144) and EDL type IIa (p = 0.0057) myofiber sizes compared to *WT* mice.

Discussion

In this study, we explored the pathological basis for motor deficits that have been identified in human patients carrying heterozygous disruptive *KMT5B* variants (Stessman et al., 2017; Trinh et al., 2019) and in a mouse model of germline *Kmt5b* haploinsufficiency (Wickramasekara et al., 2021). *Kmt5b HET* mice showed a progressive decline in neuromuscular strength from adolescence (P17) to young adulthood (P44). This was underscored by reduced skeletal muscle mass that affected slow-twitch (SOL) at a younger age than fast-twitch (EDL) muscles. This decreased muscle mass likely contributes to the decreased body weights that we have reported previously (Wickramasekara et al., 2021) and also confirmed in this study. Analysis of the NMJ revealed increased fragmentation in *HET* animals, particularly in the slow-twitch SOL. A detailed analysis of skeletal muscle myofiber composition identified increased myofiber counts by young adulthood; however, these myofibers were overall smaller in *HET* compared to *WT* animals. These data highlight an effect of germline *Kmt5b* genotype on the size, structure, and function of skeletal muscle.

Skeletal muscle growth occurs through three primary mechanisms, 1) formation of new cells (myogenesis), 2) merging of myonuclei to growing myofibers (accretion), and 3) increasing the size of cells through increased protein synthesis and decreased degradation (hypertrophy). Myogenesis happens exclusively before birth in mice. Myofiber numbers are considered fixed during postnatal life except in cases of degradation without regeneration (White et al., 2010). The scaffold for adult skeletal muscle is formed during primary myogenesis (embryonic days E10.5–E12.5) from the fusion of $Pax3^+$ myocytes expressing exclusively slow-twitch markers. Secondary myogenesis, occurring from E14.5–17.5, is achieved through the fusion of $Pax7^+$ progenitor cells to the primary scaffold (Lyons et al., 1990; Horst et al., 2006). This is when fast-twitch markers begin to be expressed, and motor neurons innervate the developing muscles forming NMJs (Slater, 1982); innervation dictates myofiber type moving forward. There is currently no evidence that loss of one germline copy of *Kmt5b* significantly disrupts primary or secondary myogenesis as all muscles analyzed at P17 expressed both the myofiber types and ratios expected. Further, when not corrected for tissue area, myofiber counts were not different between *WT* and *HET* animals at P17 or P44 (not shown) suggesting that the total initial number of myofibers required are present in the *HET* condition.

Accretion occurs postnatally and is often linked with hypertrophy to grow muscle size; all muscle weights in our model were significantly reduced in *HET* animals at P44. Accretion involves the fusion of muscle satellite cells (i.e., myonuclei) with myofibers to support the increased capacity of protein synthesis required for hypertrophy. KMT5B has been previously linked to increased mouse muscle turnover in adulthood due to changes in the satellite cell population (Neguembor et al., 2013; Boonsanay et al., 2016), a specific cell population which we did not directly analyze in this study. In our model, there was no evidence of increased degenerating/regenerating myofibers in *HET* compared to *WT* tibialis anterior muscle sections collected at P44 (not shown). However, newly regenerated myofibers are known to initially express type II (fast-twitch) markers, and specific slow motor neuron input is required to further switch to the type I (slow-twitch) myofiber program (Jerkovic et al., 1997). In other muscle pathologies where myofiber turnover is increased, increased degeneration of myofibers drives fragmentation of the NMJ (Li and Thompson, 2011; Pratt SJ et al., 2015; Haddix et al., 2018). We observe increased yet smaller type II myofibers and increased NMJ fragmentation in SOL muscles of *HET* mice at P44 (young adulthood). While these small type II myofibers could represent newly regenerated myofibers, we believe it more likely that proper developmental accretion/hypertrophy requires KMT5B activity. However, it has been shown that specific *Kmt5b* knockout in $Pax7^+$ skeletal muscle progenitor cells produces no obvious growth or clinical abnormalities (Boonsanay et al., 2016) suggesting that KMT5B's effects on skeletal muscle development could be imparted from outside the skeletal muscles themselves.

Two opposing pathways are known to regulate muscle hypertrophy where growth is stimulated by the insulin-like growth factor 1— phosphoinositide-3-kinase–Akt/protein kinase B—mammalian target of rapamycin (IGF1—PI3K—Akt/PKB—mTOR) pathway and opposed by myostatin (TGFβ superfamily)—Smad3 signaling (Egerman and Glass, 2014). Growth hormone (GH), secreted from the anterior pituitary gland, stimulates the production of IGF1 (Oberbauer, 2013), highlighting a strong force external to the

skeletal muscle that is required for growth and maturation. While most IGF1 in the body is expressed from the liver and secreted into circulation (endocrine), target tissues expressing GH receptors including skeletal muscle also express IGF1 locally (autocrine/paracrine) (Wang et al., 2013). Germline knockout of IGF1 or GH receptor in mice results in significant growth deficiencies (Coschigano et al., 2000; Liu et al., 2000). Interestingly, very similar phenotypes to those presented here–reduced muscle and body size and increased yet smaller type II fibers—have been identified in the GH receptor knockout mouse model (Sotiropoulos et al., 2006; Mavalli et al., 2010) supporting our new hypothesis that KMT5B is involved in the anabolic regulation of muscle hypertrophy.

It is interesting that both the sexes and different muscle types appear to respond differently to *Kmt5b* haploinsufficiency. Prior to the onset of puberty (i.e., P17), we noted specific effects on slow-twitch SOL muscles in *HET* males. We have previously shown at this age that *HET* males are also more impacted by body weight and motor reflex changes than *HET* females (Wickramasekara et al., 2021). In this study, the SOL weight deficit was diminished by P44 even though male *HET* body weight remained significantly lower than *WT* males, suggesting a SOL developmental delay in males. Testosterone is a potent anabolic stimulant of muscle hypertrophy, and increases in testosterone are required for sexual maturity. In male mice, testosterone increases in two waves, the first from P25-P30 and the second from P45-P50 (Barkley and Goldman, 1977). Indeed, many motor delays/deficits in human children with neurodevelopmental disorders improve with age (Ming et al., 2007). Whether this is related to the expression of androgens in certain cases or the result of interventional physical and occupational therapies remains unknown (Ming et al., 2007). Adult muscle effects were often more pronounced in females than males in our model. *HET* female mice in this study appeared to develop normally until puberty after which they presented muscle phenotypes (P44) reinforcing that KMT5B is also involved in maintaining adult muscle health, which has also been proposed by others (Neguembor et al., 2013; Boonsanay et al., 2016).

Still other mechanisms may contribute to the phenotypes observed in our model. For example, a recent publication reported decreased body fat in tissue-specific *Kmt5b*/*Kmt5c* double knockout mice (Pedrotti et al., 2019). Decreased intramuscular fat stores in skeletal muscle may contribute to the decreased skeletal muscle weights observed in male *HET* mice. We have previously reported an effect of litter on motor performance in neonatal mice (Wickramasekara et al., 2021). Neonatal nutrition is known to contribute to total body weight and muscle maturation (Glore and Layman, 1983; Miller et al., 1984; Nemeth et al., 1992; O'Connor et al., 2003). Particularly prior to weaning, litter/litter size can contribute to resource acquisition. There were no significant interactions between litter and genotype in this study. However, we did not directly measure food intake at any age in this study. Effects of litter can also be caused by the genotype of the dam (not controlled for in this study). There was no significant effect of dam genotype on neuromuscular strength (Supplementary Figures S8A,B), muscle weight (Supplementary Figures S8C–F), or mouse body weight (Supplementary Figures S8G,H). We conclude that while variability between litters existed in each experiment, the relationship between *Kmt5b* genotype and neuromuscular strength, body weight, and muscle weights remains fixed. Finally, it is possible that the LacZ gene trap affects the expression of neighboring genes compounding this phenotype, such as the predicted lncRNA, *Gm16066*, transcribed from the opposite strand. The expression pattern and function of this lncRNA are largely unknown. We found that *Gm16066* was not expressed in SOL, EDL, or liver (data not shown) by qRT-PCR in these mice; therefore, this is not likely a confounding variable.

Muscle development in humans and mice is similar; however, there are some notable differences. In humans, embryonic muscle development occurs over three myogenesis stages in which primary, secondary, and tertiary myotubes form (Draeger et al., 1987). Around birth, the majority of myofibers in human skeletal muscle are adult myofiber types (Butler-Browne et al., 1990; Schiaffino et al., 2015), as opposed to mice which will gradually transition to adult stages through P17-P21 (Agbulut et al., 2003). Additionally, humans do not express a "true" type IIb myofiber, unlike mice (Smerdu et al., 1994). Therefore, the P17-21 mouse neuromuscular development resembles an early postnatal infant and the P44 mouse a sexually mature human adolescent. We have recently reviewed elsewhere (Wickramasekara and Stessman, 2019) the phenotypes of 34 unique individuals carrying putative disruptive (nonsense, frameshift, splice donor, or missense) variants in *KMT5B*. This included all publicly available individuals from the denovoDB database (Turner et al., 2017) (denovo-db, Seattle, WA), DECIPHER (Firth et al., 2009), and ClinVar (Landrum et al., 2018), as well as large-scale genetic screening publications (Yuen et al., 2000; De Rubeis et al., 2014; Iossifov et al., 2014; Homsy et al., 2015; Lelieveld et al., 2016; Turner et al., 2016; Jonsson et al., 2017; Yuen et al., 2017; Guo et al., 2018). There was not a highly penetrant undergrowth phenotype among *KMT5B* patients like has been observed in the *Kmt5b* mouse models (Schotta et al., 2008; Wickramasekara et al., 2021). Further, some have even suggested a *KMT5B*-associated overgrowth syndrome including tall stature and macrocephaly in these patients (Faundes et al., 2017; Wang et al., 2021). However, height data were not provided for the parents of *KMT5B* patients meaning that we cannot calculate the contribution of genetic background to this phenotype. Also, head circumference is often not normalized to height such that we cannot distinguish relative macrocephaly from an overgrowth syndrome in these patients. Indeed, we have identified increased relative brain weight in males in our mouse model (Supplementary Figure S4H). The resolution of these subtle developmental phenotypes will require a larger cohort of deeply-phenotyped *KMT5B* patients. Finally, we cannot rule out differences between mouse and man, both genetic

and societal. Particularly in industrialized countries, failure to thrive in children is detected and treated early such that the natural progression of disease is never actualized. As we have reported previously, a conservative 38% of reported *KMT5B* patients had confirmed motor phenotypes (e.g., hypotonia) and/or global developmental delay, which often includes motor delays. For many individuals, data regarding motor development are never published. Indeed, motor deficits are thought to be vastly underreported among autists (Bhat, 2020). Given the rate of motor deficits and delays among the *KMT5B* patient population, the muscle phenotypes reported here, and work published by others (Neguembor et al., 2013; Boonsanay et al., 2016) we believe the mouse to be an important model for better understanding this patient population moving forward, including whether the same patients are affected by both motor and behavioral phenotypes.

In summary, this is the first study to evaluate the effects of *Kmt5b* loss on early muscle development within a genetic state and timeframe that are most relevant to humans carrying disruptive variation in *KMT5B*. We found that *Kmt5b* haploinsufficiency in mice results in significantly reduced skeletal muscle mass and neuromuscular strength underscored by increased NMJ fragmentation and an increased number of smaller type II myofibers. We conclude that KMT5B likely contributes to the process of muscle hypertrophy during early postnatal development, a mechanism that is perhaps independent from the adult quiescent muscle stem cell depletion phenotype associated with *Kmt5b* loss by others (Neguembor et al., 2013; Boonsanay et al., 2016). This is an exciting model as it suggests that some cases of neurodevelopmental motor delays may be the result of changes in skeletal muscle structure and maturation rather than the motor circuits in the brain.

Data availability statement

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

Ethics statement

The animal study was reviewed and approved by Creighton University Institutional Animal Care and Use Committee (IACUC), Omaha, NE 68178.

Author contributions

JH and HS conceived the study and designed the analyses. JH, DK, RB, and JH collected the data. KH, RW, EB, and PA contributed additional physical and/or intellectual tools and resources. JH, KH, and HS performed the analyses. JH and HS wrote the paper with contributions from all authors.

Funding

This work was funded by the LB692 (HS) and LB595 (PWA) Nebraska Tobacco Settlement Biomedical Research Development Program and the Simons Foundation Autism Research Initiative-Bridge to Independence Award (SFARI 381192) to HS. Additional funding was provided by the NIGMS (5P30GM110768-05; PD: Shelley Smith) to HS.

Acknowledgments

We thank A. Keefe for helpful discussions related to this work. This research was partially conducted using the Histology Core and Integrated Biomedical Imaging Facility (IBIF) at Creighton University, Omaha, NE. The IBIF facility is supported by the Creighton University School of Medicine and grants GM103427 and GM139762 from NIGMS. The facility was constructed with support from grants from the National Center for Research Resources (RR016469) and the NIGMS (GM103427).

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

Supplementary Material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.901228/full#supplementary-material

References

Agbulut, O., Noirez, P., Beaumont, F., and Butler-Browne, G. (2003). Myosin heavy chain isoforms in postnatal muscle development of mice. *Biol. Cell.* 95 (6), 399–406. doi:10.1016/s0248-4900(03)00087-x

Barkley, M. S., and Goldman, B. D. (1977). A quantitative study of serum testosterone, sex accessory organ growth, and the development of intermale aggression in the mouse. *Horm. Behav.* 8 (2), 208–218. doi:10.1016/0018-506x(77)90038-1

Bhat, A. N. (2020). Is motor impairment in autism spectrum disorder distinct from developmental coordination disorder? A report from the spark study. *Phys. Ther.* 100 (4), 633–644. doi:10.1093/ptj/pzz190

Biron, V. L., McManus, K. J., Hu, N., Hendzel, M. J., and Underhill, D. A. (2004). Distinct dynamics and distribution of histone methyl-lysine derivatives in mouse development. *Dev. Biol.* 276 (2), 337–351. doi:10.1016/j.ydbio.2004.08.038

Boonsanay, V., Zhang, T., Georgieva, A., Kostin, S., Qi, H., Yuan, X., et al. (2016). Regulation of skeletal muscle stem cell quiescence by suv4-20h1-dependent facultative heterochromatin formation. *Cell. Stem Cell.* 18 (2), 229–242. doi:10.1016/j.stem.2015.11.002

Butler-Browne, G. S., Barbet, J. P., and Thornell, L. E. (1990). Myosin heavy and light chain expression during human skeletal muscle development and precocious muscle maturation induced by thyroid hormone. *Anat. Embryol.* 181 (6), 513–522. doi:10.1007/BF00174624

Celegato, B., Capitanio, D., Pescatori, M., Romualdi, C., Pacchioni, B., Cagnin, S., et al. (2006). Parallel protein and transcript profiles of FSHD patient muscles correlate to the D4Z4 arrangement and reveal a common impairment of slow to fast fibre differentiation and a general deregulation of MyoD-dependent genes. *Proteomics* 6 (19), 5303–5321. doi:10.1002/pmic.200600056

Chai, R. J., Vukovic, J., Dunlop, S., Grounds, M. D., and Shavlakadze, T. (2011). Striking denervation of neuromuscular junctions without lumbar motoneuron loss in geriatric mouse muscle. *PLoS One* 6 (12), e28090. doi:10.1371/journal.pone.0028090

Coschigano, K. T., Clemmons, D., Bellush, L. L., and Kopchick, J. J. (2000). Assessment of growth parameters and life span of GHR/BP gene-disrupted mice. *Endocrinology* 141 (7), 2608–2613. doi:10.1210/endo.141.7.7586

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 (7526), 209–215. doi:10.1038/nature13772

Draeger, A., Weeds, A. G., and Fitzsimons, R. B. (1987). Primary, secondary and tertiary myotubes in developing skeletal muscle: A new approach to the analysis of human myogenesis. *J. Neurol. Sci.* 81 (1), 19–43. doi:10.1016/0022-510x(87)90181-x

Egerman, M. A., and Glass, D. J. (2014). Signaling pathways controlling skeletal muscle mass. *Crit. Rev. Biochem. Mol. Biol.* 49 (1), 59–68. doi:10.3109/10409238.2013.857291

Faundes, V., Newman, W. G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., et al. (2017). Histone lysine methylases and demethylases in the landscape of human developmental disorders. *Am. J. Hum. Genet.* 102, 175–187. doi:10.1016/j.ajhg.2017.11.013

Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., et al. (2009). Decipher: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.* 84 (4), 524–533. doi:10.1016/j.ajhg.2009.03.010

Glore, S. R., and Layman, D. K. (1983). Cellular development of skeletal muscle during early periods of nutritional restriction and subsequent rehabilitation. *Pediatr. Res.* 17 (7), 602–605. doi:10.1203/00006450-198307000-00017

Guo, H., Wang, T., Wu, H., Long, M., Coe, B. P., Li, H., et al. (2018). Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Mol. Autism* 9, 64. doi:10.1186/s13229-018-0247-z

Haddix, S. G., Lee, Y. I., Kornegay, J. N., and Thompson, W. J. (2018). Cycles of myofiber degeneration and regeneration lead to remodeling of the neuromuscular junction in two mammalian models of Duchenne muscular dystrophy. *PLoS One* 13 (10), e0205926. doi:10.1371/journal.pone.0205926

Homsy, J., Zaidi, S., Shen, Y., Ware, J. S., Samocha, K. E., Karczewski, K. J., et al. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* 350 (6265), 1262–1266. doi:10.1126/science.aac9396

Horst, D., Ustanina, S., Sergi, C., Mikuz, G., Juergens, H., Braun, T., et al. (2006). Comparative expression analysis of Pax3 and Pax7 during mouse myogenesis. *Int. J. Dev. Biol.* 50 (1), 47–54. doi:10.1387/ijdb.052111dh

Iossifov, I., O'Roak, B. J., Sanders, S. J., Ronemus, M., Krumm, N., Levy, D., et al. (2014). The contribution of de novo coding mutations to autism spectrum disorder. *Nature* 515 (7526), 216–221. doi:10.1038/nature13908

Jenuwein, T., and Allis, C. D. (2001). Translating the histone code. *Science* 293 (5532), 1074–1080. doi:10.1126/science.1063127

Jerkovic, R., Argentini, C., Serrano-Sanchez, A., Cordonnier, C., and Schiaffino, S. (1997). Early myosin switching induced by nerve activity in regenerating slow skeletal muscle. *Cell. Struct. Funct.* 22 (1), 147–153. doi:10.1247/csf.22.147

Jin, T. E., Wernig, A., and Witzemann, V. (2008). Changes in acetylcholine receptor function induce shifts in muscle fiber type composition. *FEBS J.* 275 (9), 2042–2054. doi:10.1111/j.1742-4658.2008.06359.x

Jonsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., et al. (2017). Parental influence on human germline de novo mutations in 1, 548 trios from Iceland. *Nature* 549 (7673), 519–522. doi:10.1038/nature24018

Klooster, R., Plomp, J. J., Huijbers, M. G., Niks, E. H., Straasheijm, K. R., Detmers, F. J., et al. (2012). Muscle-specific kinase myasthenia gravis IgG4 autoantibodies cause severe neuromuscular junction dysfunction in mice. *Brain* 135 (4), 1081–1101. doi:10.1093/brain/aws025

Landini, G. (2020). Novel context-based segmentation algorithms for intelligent microscopy. AvaliableAt: https://blog.bham.ac.uk/intellimic/g-landini-software/2020.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., et al. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46 (D1), D1062–D1067. doi:10.1093/nar/gkx1153

Lelieveld, S. H., Reijnders, M. R., Pfundt, R., Yntema, H. G., Kamsteeg, E. J., de Vries, P., et al. (2016). Meta-analysis of 2, 104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* 19 (9), 1194–1196. doi:10.1038/nn.4352

Li, Y., and Thompson, W. J. (2011). Nerve terminal growth remodels neuromuscular synapses in mice following regeneration of the postsynaptic muscle fiber. *J. Neurosci.* 31 (37), 13191–13203. doi:10.1523/JNEUROSCI.2953-11.2011

Liu, J. L., Yakar, S., and LeRoith, D. (2000). Conditional knockout of mouse insulin-like growth factor-1 gene using the Cre/loxP system. *Proc. Soc. Exp. Biol. Med.* 223 (4), 344–351. doi:10.1046/j.1525-1373.2000.22349.x

Lyons, G. E., Ontell, M., Cox, R., Sassoon, D., and Buckingham, M. (1990). The expression of myosin genes in developing skeletal muscle in the mouse embryo. *J. Cell. Biol.* 111 (4), 1465–1476. doi:10.1083/jcb.111.4.1465

Mavalli, M. D., DiGirolamo, D. J., Fan, Y., Riddle, R. C., Campbell, K. S., van Groen, T., et al. (2010). Distinct growth hormone receptor signaling modes regulate skeletal muscle development and insulin sensitivity in mice. *J. Clin. Invest.* 120 (11), 4007–4020. doi:10.1172/JCI42447

Miller, W. C., Bryce, G. R., and Conlee, R. K. (1984). Adaptations to a high-fat diet that increase exercise endurance in male rats. *J. Appl. Physiol. Respir. Environ. Exerc. Physiol.* 56 (1), 78–83. doi:10.1152/jappl.1984.56.1.78

Ming, X., Brimacombe, M., and Wagner, G. C. (2007). Prevalence of motor impairment in autism spectrum disorders. *Brain Dev.* 29 (9), 565–570. doi:10.1016/j.braindev.2007.03.002

Neguembor, M. V., Xynos, A., Onorati, M. C., Caccia, R., Bortolanza, S., Godio, C., et al. (2013). FSHD muscular dystrophy region gene 1 binds Suv4-20h1 histone methyltransferase and impairs myogenesis. *J. Mol. Cell. Biol.* 5 (5), 294–307. doi:10.1093/jmcb/mjt018

Nemeth, P. M., Rosser, B. W., Choksi, R. M., Norris, B. J., and Baker, K. M. (1992). Metabolic response to a high-fat diet in neonatal and adult rat muscle. *Am. J. Physiol.* 262 (1), C282–C286. doi:10.1152/ajpcell.1992.262.2.C282

Nilwik, R., Snijders, T., Leenders, M., Groen, B. B., van Kranenburg, J., Verdijk, L. B., et al. (2013). The decline in skeletal muscle mass with aging is mainly attributed to a reduction in type II muscle fiber size. *Exp. Gerontol.* 48 (5), 492–498. doi:10.1016/j.exger.2013.02.012

O'Connor, P. M., Bush, J. A., Suryawan, A., Nguyen, H. V., and Davis, T. A. (2003). Insulin and amino acids independently stimulate skeletal muscle protein synthesis in neonatal pigs. *Am. J. Physiol. Endocrinol. Metab.* 284 (1), E110–E119. doi:10.1152/ajpendo.00326.2002

Oberbauer, A. M. (2013). The regulation of IGF-1 gene transcription and splicing during development and aging. *Front. Endocrinol.* 4, 39. doi:10.3389/fendo.2013.00039

Pedemonte, M., Sandri, C., Schiaffino, S., and Minetti, C. (1999). Early decrease of IIx myosin heavy chain transcripts in Duchenne muscular dystrophy. *Biochem. Biophys. Res. Commun.* 255 (2), 466–469. doi:10.1006/bbrc.1999.0213

Pedrotti, S., Caccia, R., Neguembor, M. V., Garcia-Manteiga, J. M., Ferri, G., de Palma, C., et al. (2019). The Suv420h histone methyltransferases regulate PPAR-

gamma and energy expenditure in response to environmental stimuli. *Sci. Adv.* 5 (4), eaav1472. doi:10.1126/sciadv.aav1472

Pratt Sj, Valencia A. P., Le, G. K., Shah, S. B., and Lovering, R. M. (2015). Pre- and postsynaptic changes in the neuromuscular junction in dystrophic mice. *Front. Physiol.* 6, 252. doi:10.3389/fphys.2015.00252

Pratt, S. J. P., Iyer, S. R., Shah, S. B., and Lovering, R. M. (2018). Imaging analysis of the neuromuscular junction in dystrophic muscle. *Methods Mol. Biol.* 1687, 57–72. doi:10.1007/978-1-4939-7374-3_5

Rowan, S. L., Rygiel, K., Purves-Smith, F. M., Solbak, N. M., Turnbull, D. M., Hepple, R. T., et al. (2012). Denervation causes fiber atrophy and myosin heavy chain co-expression in senescent skeletal muscle. *PLoS One* 7 (1), e29082. doi:10.1371/journal.pone.0029082

Rudolf, R., Bogomolovas, J., Strack, S., Choi, K. R., Khan, M. M., Wagner, A., et al. (2013). Regulation of nicotinic acetylcholine receptor turnover by MuRF1 connects muscle activity to endo/lysosomal and atrophy pathways. *Age* 35 (5), 1663–1674. doi:10.1007/s11357-012-9468-9

Schiaffino, S., and Reggiani, C. (1985). Myosin isoforms in mammalian skeletal muscle. *J. Appl. Physiol.* 77 (2), 493–501. doi:10.1152/jappl.1994.77.2.493

Schiaffino, S., Rossi, A. C., Smerdu, V., Leinwand, L. A., and Reggiani, C. (2015). Developmental myosins: Expression patterns and functional significance. *Skelet. Muscle* 5, 22. doi:10.1186/s13395-015-0046-6

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH image to ImageJ: 25 years of image analysis. *Nat. Methods* 9 (7), 671–675. doi:10.1038/nmeth.2089

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., et al. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes. Dev.* 18 (11), 1251–1262. doi:10.1101/gad.300704

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., et al. (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes. Dev.* 22 (15), 2048–2061. doi:10.1101/gad.476008

Shoaib, M., Chen, Q., Shi, X., Nair, N., Prasanna, C., Yang, R., et al. (2021). Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes. *Nat. Commun.* 12 (1), 4800. doi:10.1038/s41467-021-25051-2

Slater, C. R. (1982). Postnatal maturation of nerve-muscle junctions in hindlimb muscles of the mouse. *Dev. Biol.* 94 (1), 11–22. doi:10.1016/0012-1606(82)90063-x

Smerdu, V., Karsch-Mizrachi, I., Campione, M., Leinwand, L., and Schiaffino, S. (1994). Type IIx myosin heavy chain transcripts are expressed in type IIb fibers of human skeletal muscle. *Am. J. Physiol.* 267 (1), C1723–C1728. doi:10.1152/ajpcell.1994.267.6.C1723

Smith, J. P., Hicks, P. S., Ortiz, L. R., Martinez, M. J., and Mandler, R. N. (1995). Quantitative measurement of muscle strength in the mouse. *J. Neurosci. Methods* 62 (1-2), 15–19. doi:10.1016/0165-0270(95)00049-6

Sotiropoulos, A., Ohanna, M., Kedzia, C., Menon, R. K., Kopchick, J. J., Kelly, P. A., et al. (2006). Growth hormone promotes skeletal muscle cell fusion independent of insulin-like growth factor 1 up-regulation. *Proc. Natl. Acad. Sci. U. S. A.* 103 (19), 7315–7320. doi:10.1073/pnas.0510033103

Souayah, N., Potian, J. G., Garcia, C. C., Krivitskaya, N., Boone, C., Routh, V. H., et al. (2009). Motor unit number estimate as a predictor of motor dysfunction in an animal model of type 1 diabetes. *Am. J. Physiol. Endocrinol. Metab.* 297 (3), E602–E608. doi:10.1152/ajpendo.00245.2009

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 (4), 515–526. doi:10.1038/ng.3792

Tanner, C. J., Barakat, H. A., Dohm, G. L., Pories, W. J., MacDonald, K. G., Cunningham, P. R., et al. (2002). Muscle fiber type is associated with obesity and weight loss. *Am. J. Physiol. Endocrinol. Metab.* 282 (6), E1191–E1196. doi:10.1152/ajpendo.00416.2001

Termin, A., Staron, R. S., and Pette, D. (1989). Myosin heavy chain isoforms in histochemically defined fiber types of rat muscle. *Histochemistry* 92 (6), 453–457. doi:10.1007/BF00524756

Terranova, R., Sauer, S., Merkenschlager, M., and Fisher, A. G. (2005). The reorganisation of constitutive heterochromatin in differentiating muscle requires HDAC activity. *Exp. Cell. Res.* 310 (2), 344–356. doi:10.1016/j.yexcr.2005.07.031

Trinh, J., Kandaswamy, K. K., Werber, M., Weiss, M. E. R., Oprea, G., Kishore, S., et al. (2019). Novel pathogenic variants and multiple molecular diagnoses in neurodevelopmental disorders. *J. Neurodev. Disord.* 11 (1), 11. doi:10.1186/s11689-019-9270-4

Tsang, L. W., Hu, N., and Underhill, D. A. (2010). Comparative analyses of SUV420H1 isoforms and SUV420H2 reveal differences in their cellular localization and effects on myogenic differentiation. *PLoS One* 5 (12), e14447. doi:10.1371/journal.pone.0014447

Turner, T. N., Hormozdiari, F., Duyzend, M. H., McClymont, S. A., Hook, P. W., Iossifov, I., et al. (2016). Genome sequencing of autism-affected families reveals disruption of putative noncoding regulatory DNA. *Am. J. Hum. Genet.* 98 (1), 58–74. doi:10.1016/j.ajhg.2015.11.023

Turner, T. N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., Ha, F. S., et al. (2017). denovo-db: a compendium of human de novo variants. *Nucleic Acids Res.* 45 (D1), D804–D811. doi:10.1093/nar/gkw865

Wang, Y., Bikle, D. D., and Chang, W. (2013). Autocrine and paracrine actions of IGF-I signaling in skeletal development. *Bone Res.* 1 (3), 249–259. doi:10.4248/BR201303003

Wang, Z. J., Rein, B., Zhong, P., Williams, J., Cao, Q., Yang, F., et al. (2021). Autism risk gene KMT5B deficiency in prefrontal cortex induces synaptic dysfunction and social deficits via alterations of DNA repair and gene transcription. *Neuropsychopharmacology* 46, 1617–1626. doi:10.1038/s41386-021-01029-y

White, R. B., Bierinx, A. S., Gnocchi, V. F., and Zammit, P. S. (2010). Dynamics of muscle fibre growth during postnatal mouse development. *BMC Dev. Biol.* 10, 21. doi:10.1186/1471-213X-10-21

Wickramasekara, R. N., Robertson, B., Hulen, J., Hallgren, J., and Stessman, H. A. F. (2021). Differential effects by sex with Kmt5b loss. *Autism Res.* 14, 1554–1571. doi:10.1002/aur.2516

Wickramasekara, R. N., and Stessman, H. A. F. (2019). Histone 4 lysine 20 methylation: A case for neurodevelopmental disease. *Biol. (Basel)* 8 (1), E11. doi:10.3390/biology8010011

Wu, H., Siarheyeva, A., Zeng, H., Lam, R., Dong, A., Wu, X. H., et al. (2013). Crystal structures of the human histone H4K20 methyltransferases SUV420H1 and SUV420H2. *FEBS Lett.* 587 (23), 3859–3868. doi:10.1016/j.febslet.2013.10.020

Yuen, K. C. J. (2000). "Growth hormone stimulation tests in assessing adult growth hormone deficiency," in *Endotext*. K. R. Feingold, B. Anawalt, A. Boyce, G. Chrousos, W. W. de Herder, K. Dhatariya, et al. (South Dartmouth (MA: MDText.com, Inc).

Yuen, R. K. C., Merico, D., Bookman, M., Howe, J. L., Thiruvahindrapuram, B., Patel, R. V., et al. (2017). Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat. Neurosci.* 20 (4), 602–611. doi:10.1038/nn.4524

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fnins.2022.980000

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

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

Li Liu[1†], Xiaoshu Feng[2†], Sihan Liu[2†], Yanqiu Zhou[1], Xiaojing Dong[1], Hong Yao[1*] and Bo Tan[1*]

[1]Department of Gynecology and Obstetrics, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2]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-βTrCP1 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; Yepez 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[1] to exclude previously reported polymorphisms. The non-polymorphic CNVs were compared with the entries in the DECIPHER,[2] ISCA,[3] ClinGen[4], or ClinVar[5] 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.

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

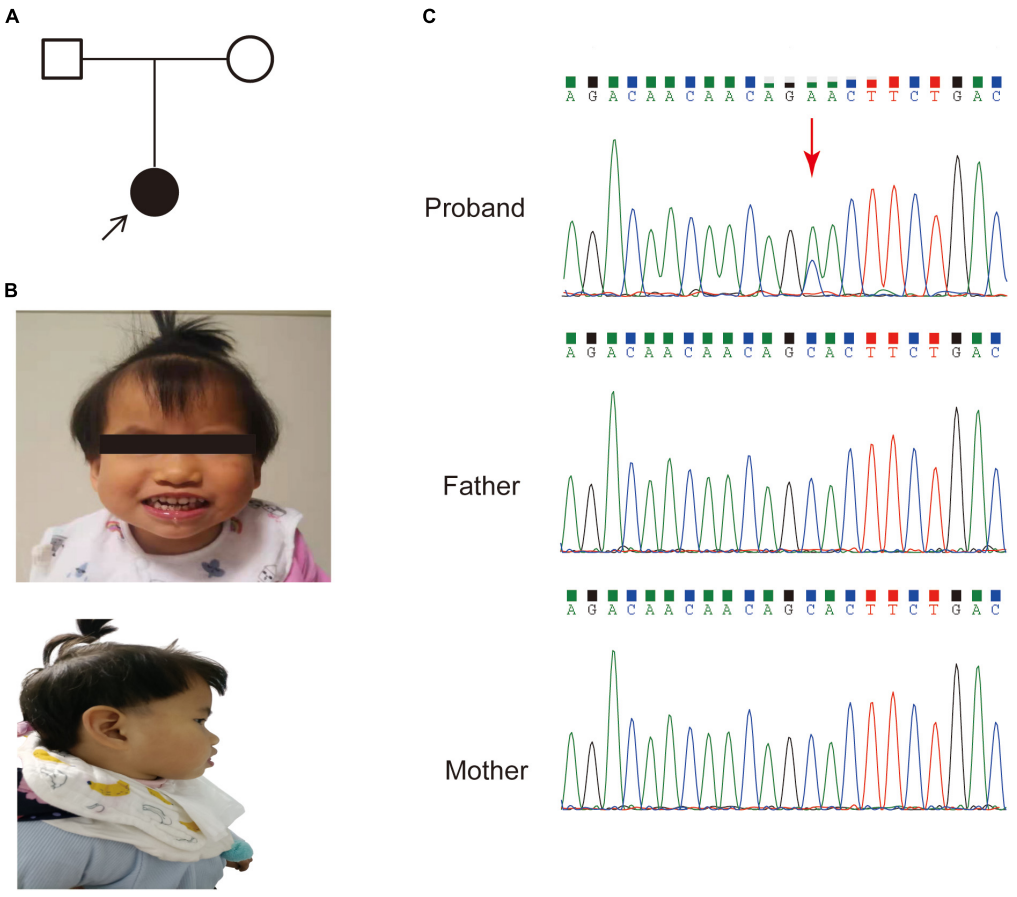


FIGURE 1
The pedigree and clinical features of the patient. **(A)** Family pedigree of the patient in this study. **(B)** Dysmorphic features, including a prominent forehead, midface hypoplasia and protruding tongue. **(C)** Sanger sequencing of the *SETBP1* gene (c.2631C > A; p. S877R) variant in genomic DNA from the family, confirming that the variant of the SETBP1 gene identified in the patient is *de novo*.

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.[6] 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.

6 https://zenodo.org/record/5638707#.Yt4B63ZByUk

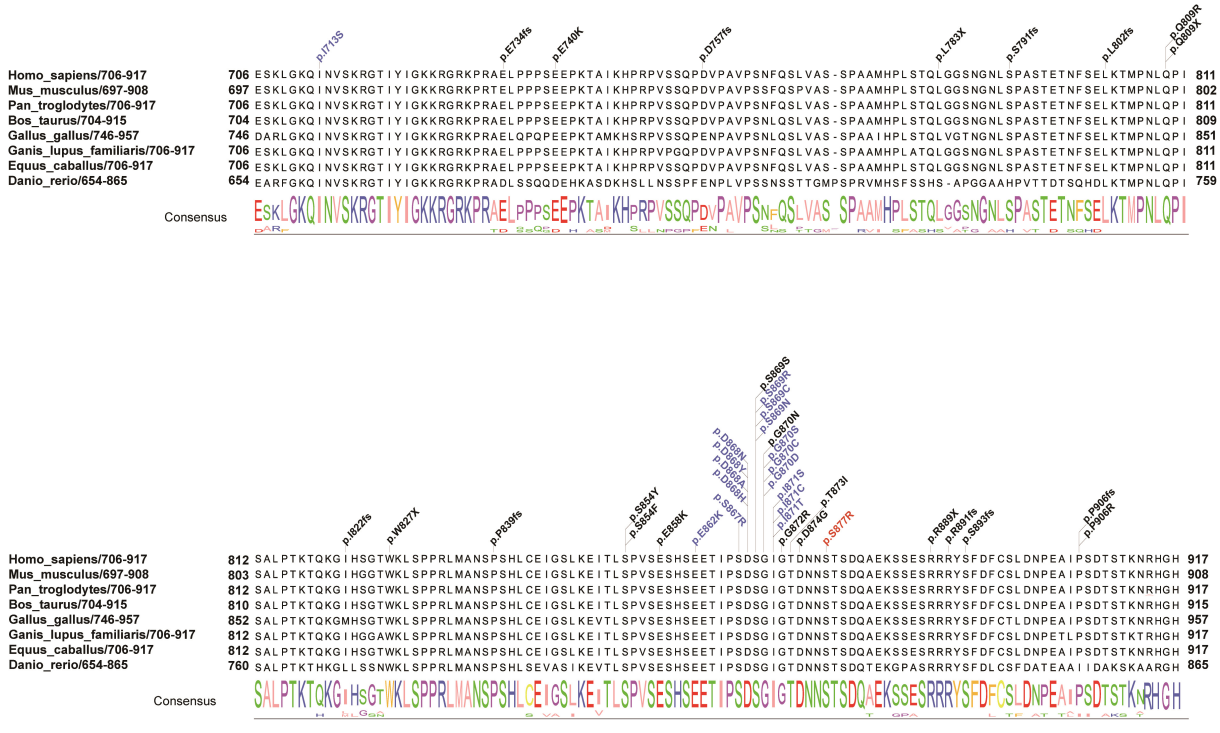


FIGURE 2
Schematic presentation of the linear SETBP1 protein of the SKI domain with all pathogenic variants. Red indicates the reported variant in the present study that caused an amino acid change (p. S877R) in a highly conserved region. Blue and black indicate variants found in SGS patients and other intellectual disability patients from public data and the literature. The consensus sequence shows the conserved score for each amino acid across eight species. Different colors of the consensus sequence represent different amino acids, and the larger the character, the more conserved the amino acid is.

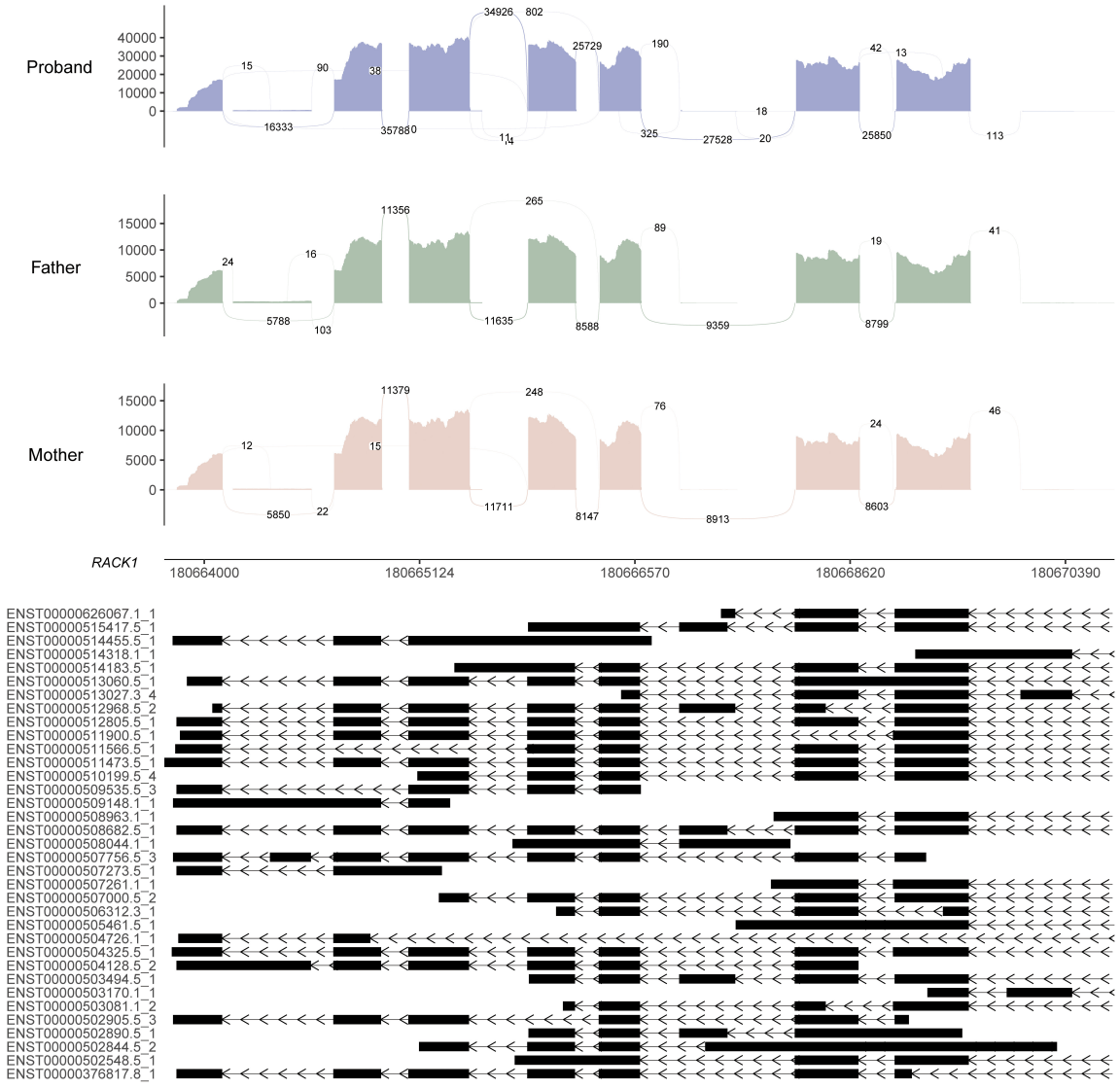


FIGURE 3
Sashimi plot of the *RACK1* gene. The coverage for each alignment track is plotted as a bar graph. Arcs representing splice junctions connect exons. Arcs display the number of reads split across the junction (junction depth). Genomic coordinates and the gene annotation track are shown below 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**).

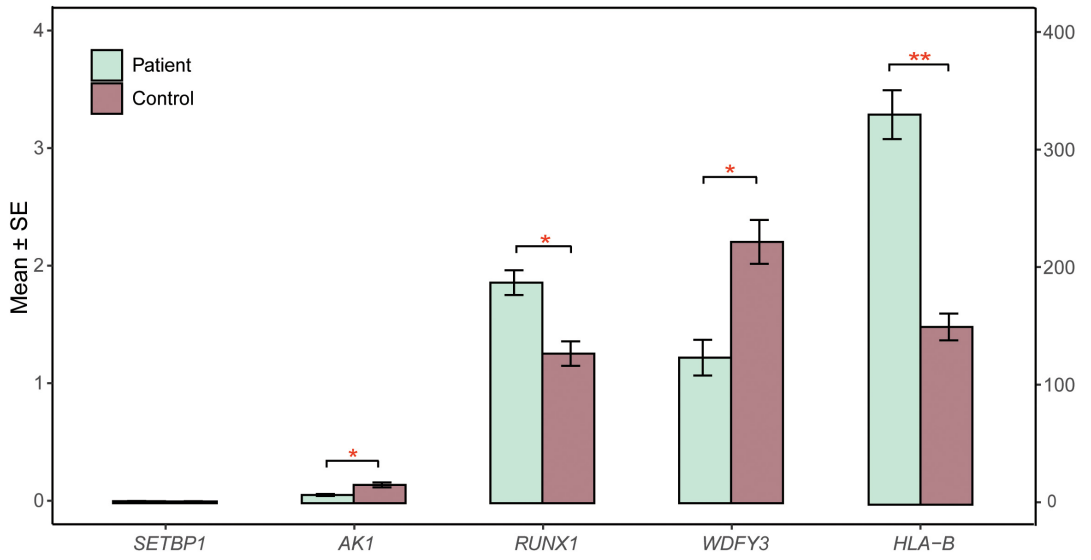


FIGURE 4
RT-qPCR validation. The relative expression of candidate genes (*SETBP1, WDFY3, AK1, RUNX1, HLA-B*) in patient and control blood cells was quantified by RT-qPCR. β-Actin mRNA was used as an internal control. The experiments were performed in triplicate. Data are shown as the mean ± SE. Two sample *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

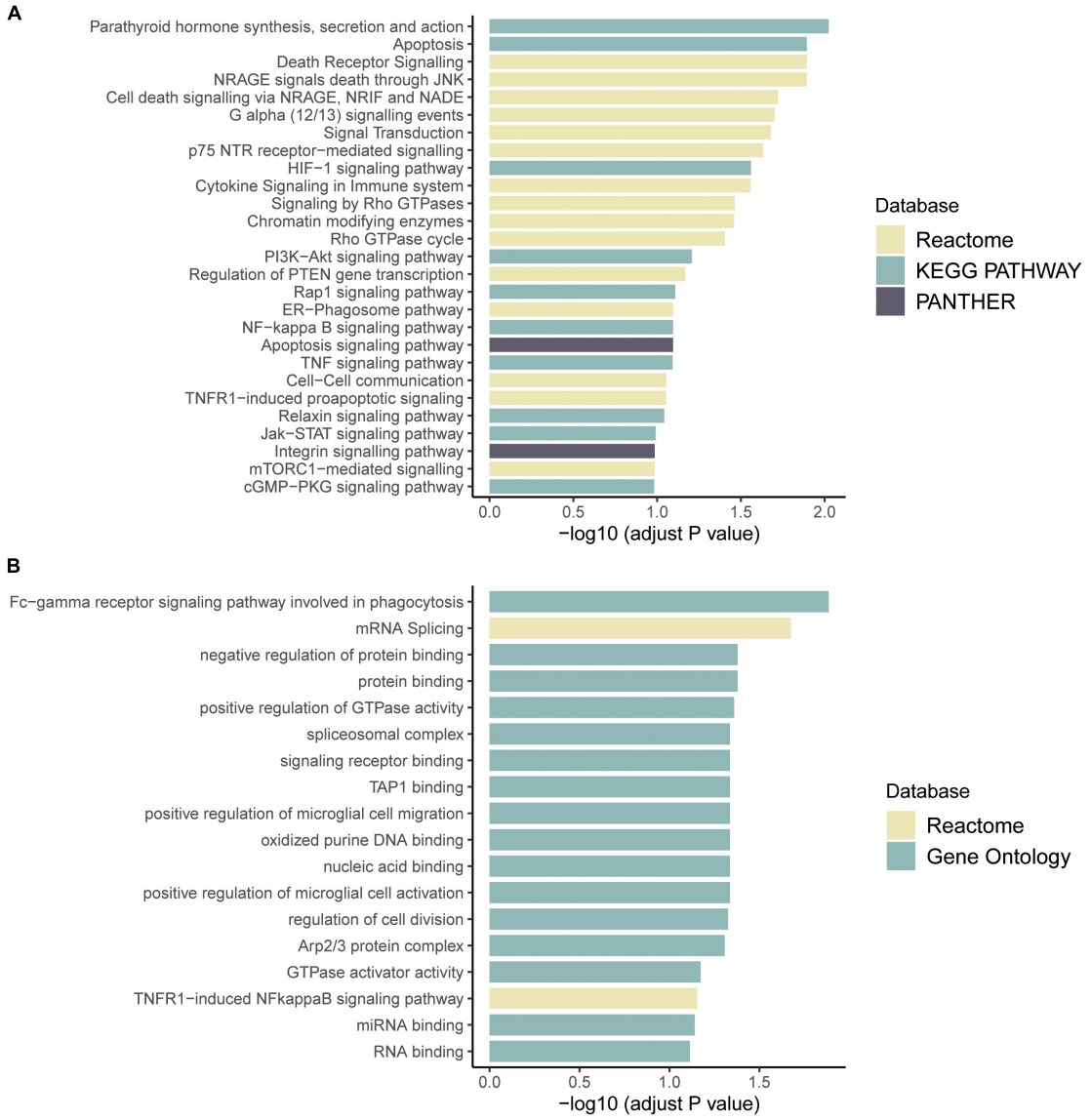


FIGURE 5
Pathway enrichment results of genes with aberrant expression **(A)** and with aberrant splicing **(B)**. Colors represent pathway items from different databases. The X-axis shows the log-transformed adjusted *p*-value. Pathways with an adjusted $p < 0.1$ were selected as significant pathways 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., Matuseviciute, A., Yepez, 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.0000000000009441

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., Kopajtich, 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., Gerdts, 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., Yepez, 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.bi1110s43

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

Yepez, 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

Yepez, 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

frontiers | Frontiers in Genetics

TYPE Original Research
PUBLISHED 22 November 2022
DOI 10.3389/fgene.2022.974357

OPEN ACCESS

EDITED BY
Zirui Dong,
The Chinese University of Hong Kong, China

REVIEWED BY
Lin Zhang,
China University of Mining and Technology, China
Jiawei Shi,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Min Chen,
edchen99@gmail.com

SPECIALTY SECTION
This article was submitted to Human and Medical Genomics, a section of the journal Frontiers in Genetics

RECEIVED 21 June 2022
ACCEPTED 04 November 2022
PUBLISHED 22 November 2022

CITATION
Yang J, Zhang L, Li Y and Chen M (2022), Identifying key m^{6}A-methylated lncRNAs and genes associated with neural tube defects *via* integrative MeRIP and RNA sequencing analyses. *Front. Genet.* 13:974357. doi: 10.3389/fgene.2022.974357

Identifying key m^{6}A-methylated lncRNAs and genes associated with neural tube defects *via* integrative MeRIP and RNA sequencing analyses

Jing Yang[1], Luting Zhang[2], Yingting Li[2] and Min Chen[2]*

[1]Department of Obstetrics, Affiliated Xiaoshan Hospital, Hangzhou Normal University, Hangzhou, Zhejiang, China, [2]Department of Obstetrics and Gynecology, Department of Fetal Medicine and Prenatal Diagnosis, Key Laboratory for Major Obstetric Diseases of Guangdong Province, The Third Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China

Objective: N^6-methyladenosine (m^{6}A) is a common post-transcriptional modification of messenger RNAs (mRNAs) and long non-coding RNAs (lncRNAs). However, m^{6}A-modified lncRNAs are still largely unexplored. This study aimed to investigate differentially m^{6}A-modified lncRNAs and genes involved in neural tube defect (NTD) development.

Methods: Pregnant Kunming mice (9–10 weeks of age) were treated with retinoic acid to construct NTD models. m^{6}A levels and methyltransferase-like 3 (*METTL3*) expression were evaluated in brain tissues of the NTD models. Methylated RNA immunoprecipitation sequencing (MeRIP-seq) and RNA sequencing (RNA-seq) were performed on the NovaSeq platform and Illumina HiSeq 2,500 platform, respectively. Differentially m^{6}A-methylated differentially expressed lncRNAs (DElncRNAs) and differentially expressed genes (DEGs) were identified, followed by GO biological process and KEGG pathway functional enrichment analyses. Expression levels of several DElncRNAs and DEGs were evaluated by quantitative reverse transcription-polymerase chain reaction (qRT-PCR) for validation.

Results: m^{6}A levels and *METTL3* expression levels were significantly lower in the brain tissues of the NTD mouse model than in controls. By integrating MeRIP-seq and RNA-seq data, 13 differentially m^{6}A-methylated DElncRNAs and 170 differentially m^{6}A-methylated DEGs were identified. They were significantly enriched in the Hippo signaling pathway and mannose-type *O*-glycan biosynthesis. The qRT-PCR results confirmed the decreased expression levels of lncRNAs, such as Mir100hg, Gm19265, Gm10544, and Malat1, and genes, such as *Zfp236*, *Erc2*, and *Hmg20a*, in the NTD group.

Conclusion: *METTL3*-mediated m^{6}A modifications may be involved in NTD development. In particular, decreased expression levels of Mir100hg, Gm19265, Gm10544, Malat1, *Zfp236*, *Erc2*, and *Hmg20a* may contribute to the development of NTD.

KEYWORDS

neural tube defects, N6-methyladenosine modification, long non-coding RNA, functional enrichment analysis, methylated RNA immunoprecipitation sequencing

Introduction

Neural tube defects (NTDs) are common congenital abnormalities caused by the failure of the neural tube to close during embryogenesis (Yadav et al., 2021). The prevalence of NTDs is estimated to be 18.6 per 10,000 live births (Finnell et al., 2021). Babies with NTDs are more likely to be stillborn, die shortly after birth, or develop different degrees of disability (Huang et al., 2021). The etiology of NTDs is complex and is associated with interactions between genetic factors and diverse environmental factors (Avagliano et al., 2019; Kakebeen and Niswander, 2021). However, the molecular mechanisms underlying NTDs have not yet been fully elucidated.

An N^6-methyladenosine (m^6A) modification is a dynamic and reversible process modulated by methyltransferase "writers" (such as methyltransferase-like 3 (*METTL3*)) and demethylase "erasers" (such as alkB homolog 5 (*ALKBH5*)) (Zhang W. et al., 2021). m^6A modifications are crucial for the regulation of RNA metabolism, including RNA stability, translation, alternative splicing, and translocation (Jiang et al., 2021). Moreover, m^6A modifications have functions in embryonic development and neurodevelopmental diseases (Yen and Chen, 2021). m^6A is the most prevalent messenger RNA (mRNA) and long non-coding RNA (lncRNA) modification (Tang et al., 2021). lncRNAs are a group of RNA transcripts longer than 200 nucleotides without open reading frames (Iyer et al., 2015). They have been implicated in the development of neurodevelopmental and neuropsychiatric disorders (Aliperti et al., 2021). Furthermore, m^6A-modified lncRNAs are involved in various diseases. For instance, the lncRNA DNA methylation-deregulated and RNA m6A reader-cooperating lncRNA (*DMDRMR*) interacts with the m^6A reader insulin-like growth factor 2 mRNA-binding protein 3 (IGF2BP3) to stabilize target genes, like cyclin-dependent kinase 4 (*CDK4*), in an m^6A-dependent manner, thus exerting an oncogenic effect in clear cell renal cell carcinoma (Gu et al., 2021). m^6A modifications of the lncRNA ZNFX1 Antisense RNA 1 (*ZFAS1*) and *RAB22A*, member RAS oncogene family (*RAB22A*) *via* METTL14 contributes to the development of atherosclerosis (Gong et al., 2021). Additionally, m^6A-modified lncRNAs play pivotal roles in obstructive nephropathy (Liu P. et al., 2020), intervertebral disc degeneration (Li et al., 2022), and muscle development (Xie et al., 2021). However, the key m^6A-modified lncRNAs and their regulatory mechanisms in NTD development have not yet been thoroughly investigated.

In the present study, we constructed a mouse model of NTD and investigated the overall m^6A levels and *METTL3* expression. We then performed m^6A-modified RNA immunoprecipitation sequencing (MeRIP-seq) and RNA sequencing (RNA-seq) to compare brain tissues of NTD embryos and control embryos and conducted comprehensive bioinformatics analyses to identify key m^6A-modified lncRNAs and genes associated with NTDs. Moreover, the expression levels of m^6A-modified lncRNAs and genes were experimentally validated. These results are expected to improve our understanding of the molecular mechanisms underlying NTDs.

Materials and methods

Animal models and samples

The Ethics Committee of the Third Affiliated Hospital of Guangzhou Medical University approved this study (2022-041, date of approval: 1 June 2022). Equal numbers of male and female Kunming mice (9–10 weeks) (Caven Biogle (Suzhou) Model Animal Research Co. Ltd., Suzhou, China) were mated overnight. The vaginal plug was examined the following day, and 25 pregnant mice were used for subsequent experiments. The day (08:00) a vaginal plug was observed was regarded as the embryonic day 0 (E0d), and 16:00 was considered E0.5d. NTD models were established as described previously (Yu et al., 2017). On E7.0d–7.25d, the mice in the NTD group (N = 18) were administered corn oil-dissolved retinoic acid (50 mg/kg of body weight) (R2625; Sigma, St. Louis, MO, USA) by one-time gavage. Mice in the control group (N = 7) were administered an equal amount of corn oil. On E16.5d, the mice were sacrificed by cervical dislocation, and embryos were taken from the uteri. NTDs were confirmed using a dissecting microscope. Brain tissues (anterior end of the neural tube) of NTD and control embryos were collected and frozen for storage.

m^6A quantification

Total RNA was isolated from the brain tissues of the NTD and control groups using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). Using an m^6A RNA Methylation Quantification Kit (Abcam, Cambridge, MA, USA), m^6A levels were colorimetrically quantified by determining absorbance at 450 nm.

Quantitative reverse transcription-polymerase chain reaction (qRT-PCR)

m^6A methyltransferase *METTL3* expression in brain tissues of the NTD and control groups was detected by real-time qRT-PCR. Total RNA was extracted from brain tissues using TRIzol reagent (Invitrogen). Reverse transcription for cDNA synthesis was performed using the PrimeScript™ first strand cDNA Synthesis

Kit (Takara, Beijing, China). Real-time qRT-PCR was conducted using Power SYBR Green PCR Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) and the 7900HT Fast qPCR System (Applied Biosystems, Foster City, CA, USA). Cycling conditions were as follows: initial denaturation at 95°C for 10 min and 40 cycles of 95°C for 15 s and 60°C for 60 s, followed by a melt curve analysis from 60°C to 95.0°C in increments of 0.5°C per 10 s. The internal control was glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and the relative expression levels of lncRNAs and genes were calculated using the $2^{-\Delta\Delta CT}$ method.

Methylated RNA immunoprecipitation (IP) sequencing (MeRIP-seq) and differential methylation analysis

Total RNA was extracted from the brain tissues of three NTD embryos and three control embryos using TRIzol reagent (Invitrogen) and was treated with DNase I (Roche Diagnostics, Mannheim, Germany) to remove residual DNA. RNA was fragmented and immunoprecipitated with a mixture of beads and an anti-m^{6}A antibody (Abcam) for 6 h at 4°C. The mixture was then immunoprecipitated by incubation with beads resuspended in IP reaction buffer (fragmented RNA, 5′ IP buffer, and RNasin Plus RNase Inhibitor) at 4°C for another 2 h. Then, the immunoprecipitated RNA was eluted and used for m^{6}A MeRIP library construction using the SMARTer Stranded Total RNA-Seq Kit v2 (Pico Input Mammalian, Takara/Clontech), following the manufacturer's protocols. Sequencing was performed on the NovaSeq platform. The raw data have been deposited in the NCBI Sequence Read Archive (SRA) database under accession number PRJNA879256.

Raw reads were filtered using Trimmomatic (v0.36) (Bolger et al., 2014), followed by a quality assessment to ensure the reliability of subsequent analyses. The data were aligned to the reference genome (mm10_gencode) using STAR (v2.5.2a) (Dobin et al., 2013). Uniquely mapped reads were used for subsequent analyses. Peak calling was used to detect regions significantly enriched in RNA (peaks), which were candidate m^{6}A-methylated sites. Peak calling was performed using the MetPeak package (Cui et al., 2016) in R (v4.1.0).

Differentially methylated sites on RNAs (m^{6}A peaks) between NTD and control samples were analyzed using MeTDiff (Cui et al., 2018) in R based on MeRIP-seq data. The cutoff values were |fold change| > 1 and $p < 0.05$. RNAs with differentially methylated m^{6}A sites were classified as mRNAs, lncRNAs, miRNAs, pseudogenes, or others to better understand the function of m^{6}A methylation in various pathological processes. Moreover, the R package ChIPseeker (v1.24.0) (Yu et al., 2015) was used to annotate differentially methylated m^{6}A sites and to analyze their distribution in functional regions according to their locations in RNA transcripts (i.e., 5′ untranslated region (UTR), 3′UTR, first exon, other exon, first intron, other intron, and distal intergenic regions). mRNAs with differentially methylated m^{6}A sites in the 3′UTR region were analyzed.

RNA sequencing (RNA-seq)

Total RNA was isolated from the brain tissues of five NTD embryos and five control embryos using TRIzol reagent (Invitrogen), following RNA concentration detection using the NanoDrop 2000. Ribosomal RNA (rRNA) was removed, and the RNA was fragmented. The RNA library was established using the Illumina TruSeq RNA Sample Prep Kit and then loaded on an Illumina HiSeq 2,500 platform for 150 bp paired-end sequencing. The raw data have been deposited in the NCBI Sequence Read Archive (SRA) database under accession number PRJNA879256.

Raw reads were subjected to data filtration to remove low-quality reads. Read alignment to the reference genome (Mus_musculus.GRCm39) was conducted using HISAT2 (v2.1.0) (Kim et al., 2015). lncRNA and mRNA read counts were generated using RSEM (Li and Dewey, 2011), and mRNA or lncRNA expression levels were quantified.

To explore the key molecules in the pathogenesis of NTD, differentially expressed genes (DEGs) and differentially expressed lncRNAs (DElncRNAs) between NTD and control samples were identified using the DESeq2 package (v1.22.2) in R based on RNA-seq data. The *p*-value was adjusted using the Benjamini–Hochberg (BH) method (Benjamini and Hochberg, 1995). The cutoff values were adjusted $p < 0.05$ and |fold change| > 1.

Integrative lncRNA/mRNA analysis with differentially methylated m^{6}A sites and DElncRNAs/DEGs

Data from MeRIP-seq and RNA-seq analyses were integrated. Then, m^{6}A-methylated DElncRNAs and DEGs in which m^{6}A levels were negatively correlated with expression levels were identified by analyzing the m^{6}A levels of lncRNAs/mRNAs with differentially methylated m^{6}A sites and DElncRNA/DEG expression levels.

Construction of a lncRNA-mRNA co-expression network

The "cor" function in R was used to calculate the Pearson correlation coefficients (PCCs) between m^{6}A-methylated DElncRNAs and DEGs. The false discovery rate (FDR)-adjusted *p*-value was used for multiple testing correction. Pairs with a PCC >0.95 and FDR <0.05 were selected. The lncRNA-mRNA co-expression network was established using Cytoscape (version 3.6.1) (Shannon et al., 2003). The topological properties of this co-expression network, including node degree,

betweenness, and closeness, were analyzed using the CytoNCA plugin (version 2.1.6) (Tang et al., 2015).

Functional enrichment analysis

Gene Ontology (GO) biological process (BP) (The Gene Ontology Consortium, 2019) and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) pathway enrichment analyses were performed using the clusterProfiler package (version 3.16.0) to elucidate the functions of m^{6}A-methylated DElncRNAs and DEGs in the co-expression network (Yu et al., 2012). A value of $p < 0.05$ (adjusted by the BH method) was considered significant.

Construction of a protein–protein interaction (PPI) network

Based on the STRING (version 11.0) database (Szklarczyk et al., 2015), the interactions between DEGs co-expressed with DElncRNAs were predicted, and a PPI network was constructed. The species was set as mice, and the PPI score was 0.15 (Yang et al., 2020; Zhao et al., 2022).

Validation of key DElncRNAs and DEGs using qRT-PCR

qRT-PCR was performed to detect the expression of several identified DElncRNAs and DEGs in the NTD and control groups to validate the reliability of the bioinformatics analysis. The DElncRNAs were Gm5165, Mir100hg, Gm19265, Gm10544, A730017L22Rik, and Malat1. The DEGs included *Zfp236*, *Erc2*, *Nudcd3*, and *Hmg20a*. qRT-PCR was conducted as described in 2.3.

Statistical analysis

All data are presented as the mean ± standard deviation (SD). The differences between the NTD and control groups were analyzed using Student's *t*-tests in GraphPad Prism 5.0 (GraphPad Software, San Diego, CA, USA). $p < 0.05$ was considered statistically significant.

Results

Overall m^{6}A levels and *METTL3* expression were decreased in NTDs

We first established NTD mouse models by the intragastric administration of corn oil-dissolved retinoic acid. In the control group, the brains of mouse embryos were completely closed, the appearance was full and smooth, the optic vesicle and otic vesicle were visible, and the spinal surface was intact without laceration. In the NTD group, the rate of stillbirth increased. Some typical morphological malformations were observed, including cracks in the top of the skull, abnormalities of the hindbrain and face, and spina bifida (Figure 1A). These data suggested that the retinoic acid-induced NTD mouse model was successfully established. We then analyzed overall m^{6}A levels in NTD and control brain samples. Compared with levels in the control group, the NTD group had significantly lower m^{6}A levels, indicating that methylation reactions could be compromised in NTD embryos (Figure 1B). Furthermore, *METTL3* expression in the NTD group was remarkably lower than that in the control group (Figure 1C), indicating that *METTL3* might be essential for m^{6}A modification in NTDs.

Identification of differentially methylated m^{6}A sites based on MeRIP-seq data

The NTD and control groups (N = 3 per group) were subjected to MeRIP-seq. Supplementary Table S1 provides a statistical summary of the raw and clean reads obtained by MeRIP-seq. After sequence alignment to the reference genome, the rates of uniquely mapped reads were higher than 70% (Supplementary Table S2), indicating that the data quality was sufficiently high for subsequent analyses. Based on MeRIP-seq data, 468 hypomethylated m^{6}A sites and 7,487 hypermethylated m^{6}A sites were found in the NTD group compared to the control group (Figure 2A). According to RNA categories, we found that 50, 402, 3, and 13 hypomethylated m^{6}A sites were located in lncRNAs, mRNAs, pseudogenes, and other regions, respectively (Figure 2B), and 310, 7,128, 50, and 32 hypermethylated m^{6}A sites were located in lncRNAs, mRNAs, pseudogenes, and other regions, respectively (Figure 2C). Figure 2D shows the distribution of differentially methylated m^{6}A sites in functional regions after peak annotation. When the mRNA 3′UTR had an m^{6}A modification, 81 and 3,502 mRNAs had hypomethylated and hypermethylated m^{6}A sites, respectively.

Screening DElncRNAs and DEGs based on RNA-seq data

The NTD and control groups (N = 5 per group) were subjected to RNA-seq. In general, 945, 603, 356 raw reads were detected using RNA-seq. Supplementary Table S3 shows a statistical summary of the raw and clean reads obtained from RNA-seq after quality control. After sequence alignment to the reference genome, the rates of uniquely mapped reads were higher than 84.97% (Supplementary Table S4), indicating that the data quality was

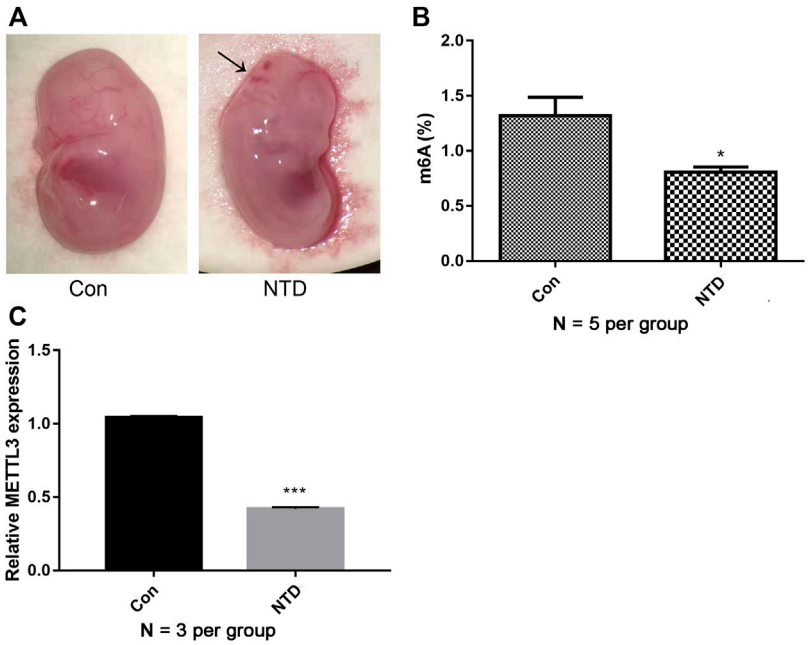


FIGURE 1
m^{6}A levels and *METTL3* expression were decreased in NTDs. **(A)** The dissecting microscope images for embryo in the control group and NTD groups on E16.5d. Arrows indicated the site of defects. **(B)** A quantitative m^{6}A analysis was conducted to explore m^{6}A enrichment in the brain tissues of NTD and control embryos (N = 5 per group). **(C)** qRT-PCR was carried out to investigate *METTL3* expression in the brain tissues of NTD and control embryos (N = 3 per group). Compared to control group, $*p < 0.05$ and $***p < 0.001$.

sufficiently high for subsequent analyses. After a differential expression analysis, 639 DElncRNAs (26 upregulated and 613 downregulated) and 1,132 DEGs (618 upregulated and 514 downregulated) were identified between the NTD and control groups (Figure 3A). As shown in a heatmap, the identified DElncRNAs (Figure 3B) and DEGs (Figure 3C) could distinguish NTD samples from control samples.

Identification of m^{6}A-related DElncRNAs and DEGs

Further integrative analyses of lncRNAs/mRNAs with differentially methylated m^{6}A sites and DElncRNAs/DEGs yielded 13 differentially m^{6}A-methylated DElncRNAs (corresponding to 20 transcripts) and 170 differentially m^{6}A-methylated DEGs (corresponding to 201 transcripts).

Analysis of lncRNA-mRNA co-expression and PPI networks

Based on PCC scores, 171 co-expression pairs were obtained, including nine differentially m^{6}A-methylated DElncRNAs and 58 differentially m^{6}A-methylated DEGs. Figure 4A shows the co-expression network. By analyzing the topological properties of nodes in the co-expression network, it was observed that lncRNAs, such as Mir100hg, A730017L22Rik, Gm10544, Gm5165, and Gm19265, had higher degrees than those of DEGs. In addition, a PPI network was established based on interactions between differentially m^{6}A-methylated DEGs and included 49 nodes and 131 interaction pairs (Figure 4B). All DEGs in the network were downregulated in NTD.

Functional enrichment analyses of nodes in the lncRNA-mRNA co-expression network

Functional enrichment analyses of the differentially m^{6}A-methylated DElncRNAs and DEGs were performed. In total, 1250 GO-BP terms and 23 KEGG pathways were significantly enriched for differentially m^{6}A-methylated DElncRNAs. For instance, A730017L22Rik, Gm19265, Gm29260, Gm5165, and Mir100 hg were significantly enriched in developmental cell growth (Figure 4C). Gm10545, Gm16096, Gm29260, and Mir100 hg were involved in mannose-type *O*-glycan biosynthesis. A730017L22Rik, Gm19265, and Gm5165 were involved in the Hippo signaling pathway (Figure 4D). Moreover, the differentially m^{6}A-methylated

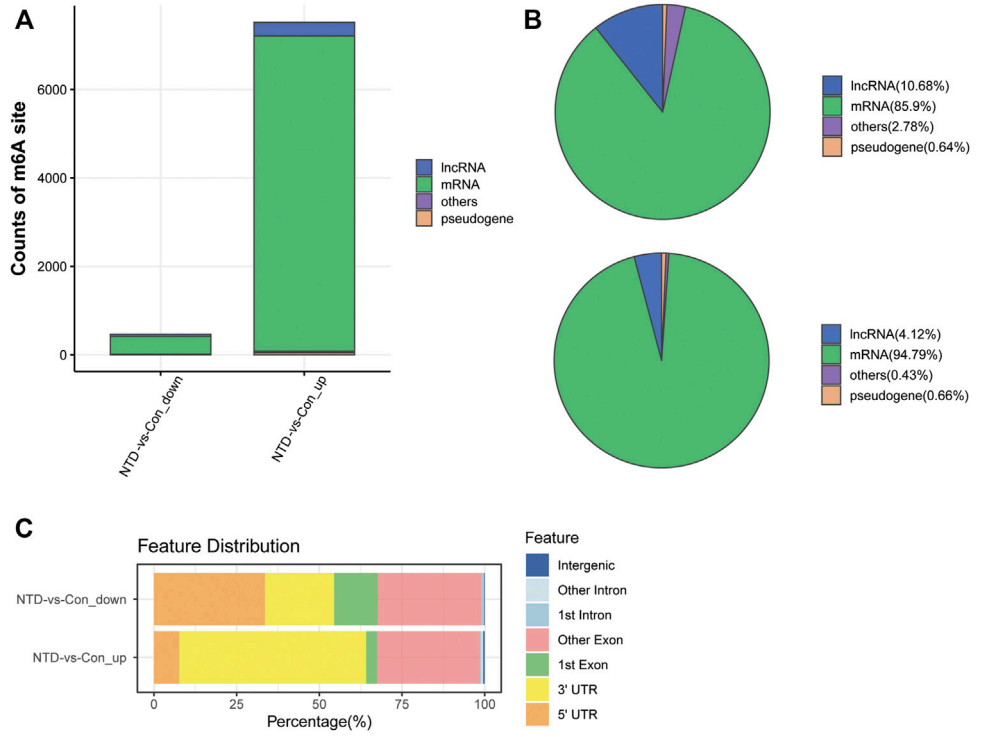


FIGURE 2
Analysis of differentially methylated m^{6}A sites based on MeRIP-seq data. **(A)** Counts of differentially methylated m^{6}A sites between the NTD and control groups. **(B)** RNA categories of hypomethylated m^{6}A sites. **(C)** RNA categories of hypermethylated m^{6}A sites. (D) Distribution of differentially methylated m^{6}A sites in functional regions.

DEGs were associated with 125 GO-BP terms, including neuron migration (Figure 4E), and one KEGG pathway, mannose-type *O*-glycan biosynthesis.

Validation by qRT-PCR

The expression levels of several DElncRNAs and DEGs were evaluated using qRT-PCR. The NTD group had dramatically lower expression levels of lncRNA, including Mir100hg, Gm19265, Gm10544, and Malat1, than those in the control group ($p < 0.05$, Figure 5A). Similarly, expression levels of key genes, such as *Zfp236*, *Erc2*, and *Hmg20a*, were lower in the NTD group than in the control group ($p < 0.05$, Figure 5B). These data were in line with the results of the bioinformatics analysis. There were no significant differences in Gm5165, A730017L22Rik, or Nudcd3 expression between the NTD and control groups.

Discussion

m^{6}A modifications contribute to the regulation of gene expression during embryonic development (Li C. et al., 2021). *METTL3* has been identified as a critical methyltransferase in m^{6}A modification and has functions in various biological processes (Liu S. et al., 2020). The *METTL3*-mediated m^{6}A modification is essential for mammalian embryonic development (Sui et al., 2020). A previous study has revealed that levels of m^{6}A modification and *METTL3* expression are lower in ethionine-induced NTD than in controls (Zhang L. et al., 2021). Consistent with these findings, we found that the retinoic acid-induced NTD mouse model had dramatically decreased overall m^{6}A levels and *METTL3* expression levels than those in control mice. These findings suggested that *METTL3*-mediated m^{6}A modifications may be involved in NTD development.

m^{6}A modifications can modulate the expression of RNAs, including lncRNAs (Meyer and Jaffrey, 2014). There is increasing evidence that m^{6}A-mediated epitranscriptomic changes can modulate lncRNAs in the developing cortex of mouse brains (Nie et al., 2021). We identified differentially m^{6}A-methylated DElncRNAs associated with NTDs *via* an integrative analysis of MeRIP-seq and RNA-seq data, including Mir100hg, Gm19265, Gm10544, and Malat1. The neurogenic lncRNA Mir100 hg is a miR-125b and let-7 precursor. Both are implicated in neural development

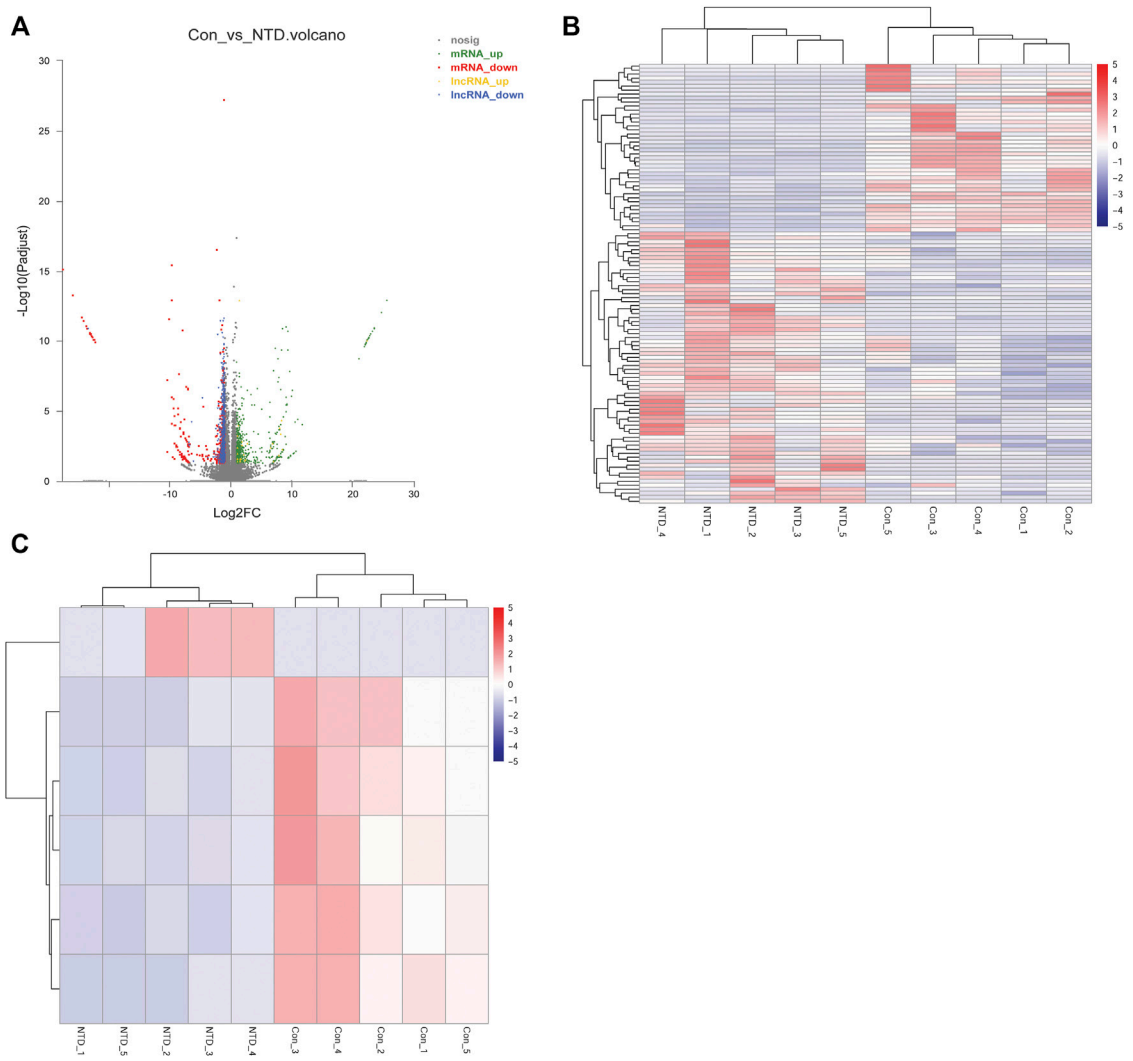


FIGURE 3
Analysis of DElncRNAs and DEGs between the NTD and control groups based on RNA-seq data. **(A)** Volcano plot of DElncRNAs and DEGs. Green, red, yellow, and blue nodes represent upregulated mRNAs, downregulated mRNAs, upregulated lncRNAs, and downregulated lncRNAs, respectively. **(B)** Heat map of DElncRNAs. **(C)** Heat map of DEGs.

(Rybak et al., 2008; Le et al., 2009). Malat1 is extremely abundant in brain tissues and is associated with neurological disorders, such as stroke, Alzheimer's disease, and retinal neurodegeneration (Zhang et al., 2017; Meng et al., 2019). In addition to lncRNAs, we identified m^{6}A-methylated DEGs, including *Zfp236*, *Erc2*, and *Hmg20a*. *Hmg20a* is mainly expressed in hypothalamic astrocytes and is crucial for neuronal integrity preservation (Lorenzo et al., 2021). *Erc2* participates in synaptic and neuronal functions (Lenihan et al., 2017). Nevertheless, the specific functions of *Gm19265*, *Gm10544*, and *Zfp236* in neuronal development and nervous system-related diseases have not been reported. Our results showed that these lncRNAs and genes are downregulated in NTD, implying that their dysregulation may be associated with NTD development. Further studies are needed to investigate the roles of these lncRNAs and genes in NTDs.

Notably, the differentially m^{6}A-methylated DElncRNAs and DEGs were significantly enriched in multiple GO terms and KEGG pathways. *O*-Mannose-linked glycans are highly enriched in the brain and are vital for nervous system functions (Morise et al., 2014). They are also involved in brain development and remyelination (Gao et al., 2019). Our analysis revealed that Gm10545, Gm16096, Gm29260, and Mir100 hg were involved in mannose-type *O*-glycan biosynthesis, providing a mechanism by which

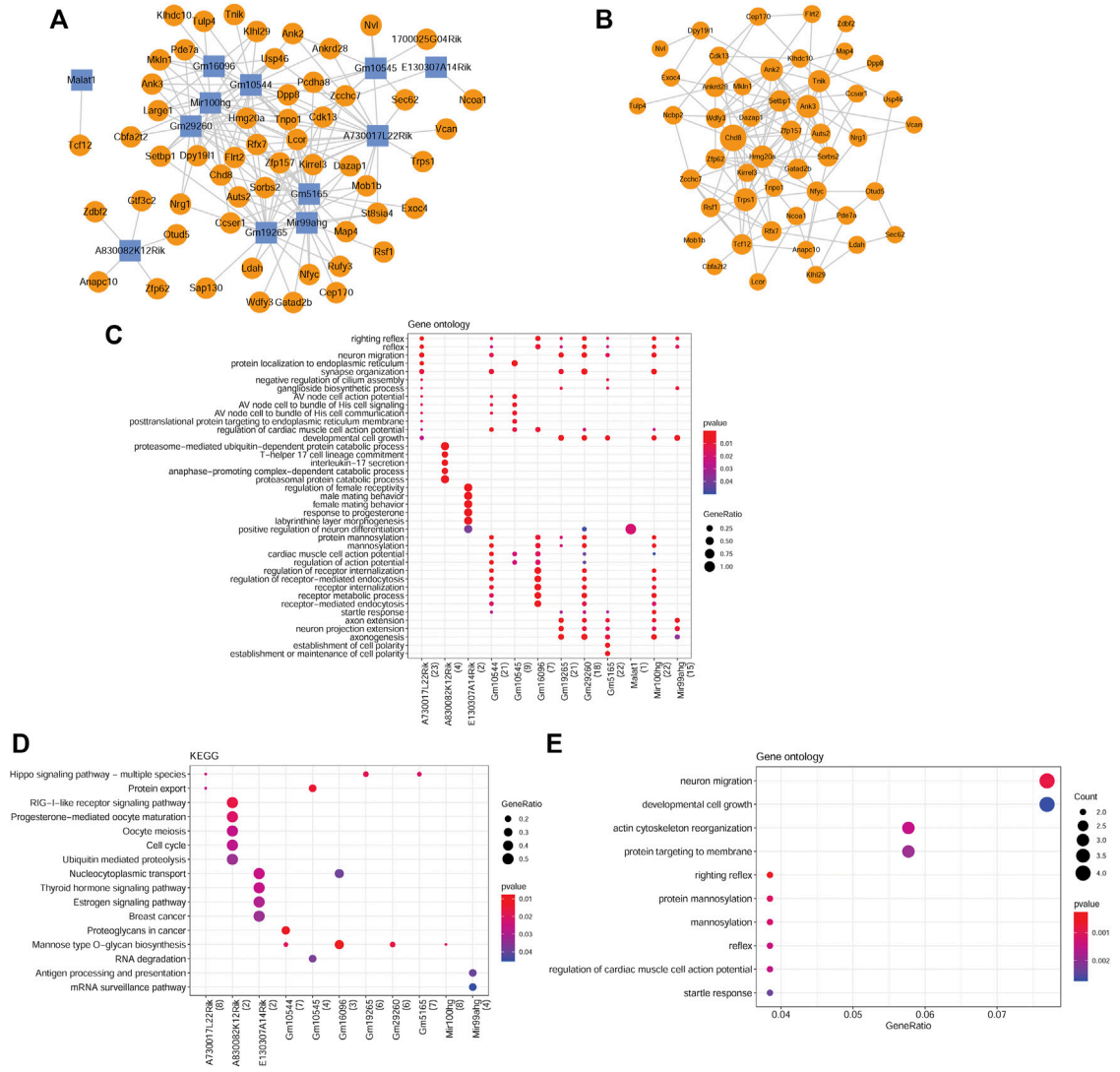


FIGURE 4
Analysis of co-expressed DElncRNAs and DEGs. **(A)** lncRNA-mRNA co-expression network. Yellow circles represent downregulated mRNAs. Blue squares represent downregulated lncRNAs. **(B)** PPI network constructed by co-expressed DEGs. **(C)** GO-BP terms enriched by co-expressed DElncRNAs. **(D)** KEGG pathways enriched by co-expressed DElncRNAs. **(E)** GO-BP terms enriched by co-expressed DEGs. The size of a node corresponds to the size of the degree.

m^{6}A-methylated DElncRNAs affected the development of NTDs. Moreover, A730017L22Rik, Gm19265, and Gm5165 were implicated in the Hippo signaling pathway, which has established roles in neuronal cell differentiation, neuroinflammation, and neuronal death (Cheng et al., 2020; Li X. et al., 2021). The Hippo/YAP signaling pathway has been implicated in the development of neurological diseases (Bao et al., 2017). Our findings revealed the potential role of the Hippo signaling pathway in NTD development. Moreover, the identified differentially m^{6}A-methylated DEGs were significantly enriched in GO-BP terms, including neuron migration, and one KEGG pathway, mannose-type *O*-glycan biosynthesis. Neuronal migration is a fundamental step in nervous system formation (Stockinger et al., 2011). These pathways and GO functions provide insight into the roles of key m^{6}A-methylated DElncRNAs and DEGs in the development of NTD.

In conclusion, our study is the first to identify differentially m^{6}A-methylated DElncRNAs and DEGs associated with NTDs. Our findings demonstrated that *METTL3*-mediated m^{6}A modifications may be involved in the development of NTD. In particular, decreased expression levels of lncRNAs, such as Mir100hg, Gm19265, Gm10544, and Malat1, as well as key genes, such as *Zfp236*, *Erc2*, and *Hmg20a*, may be crucial in

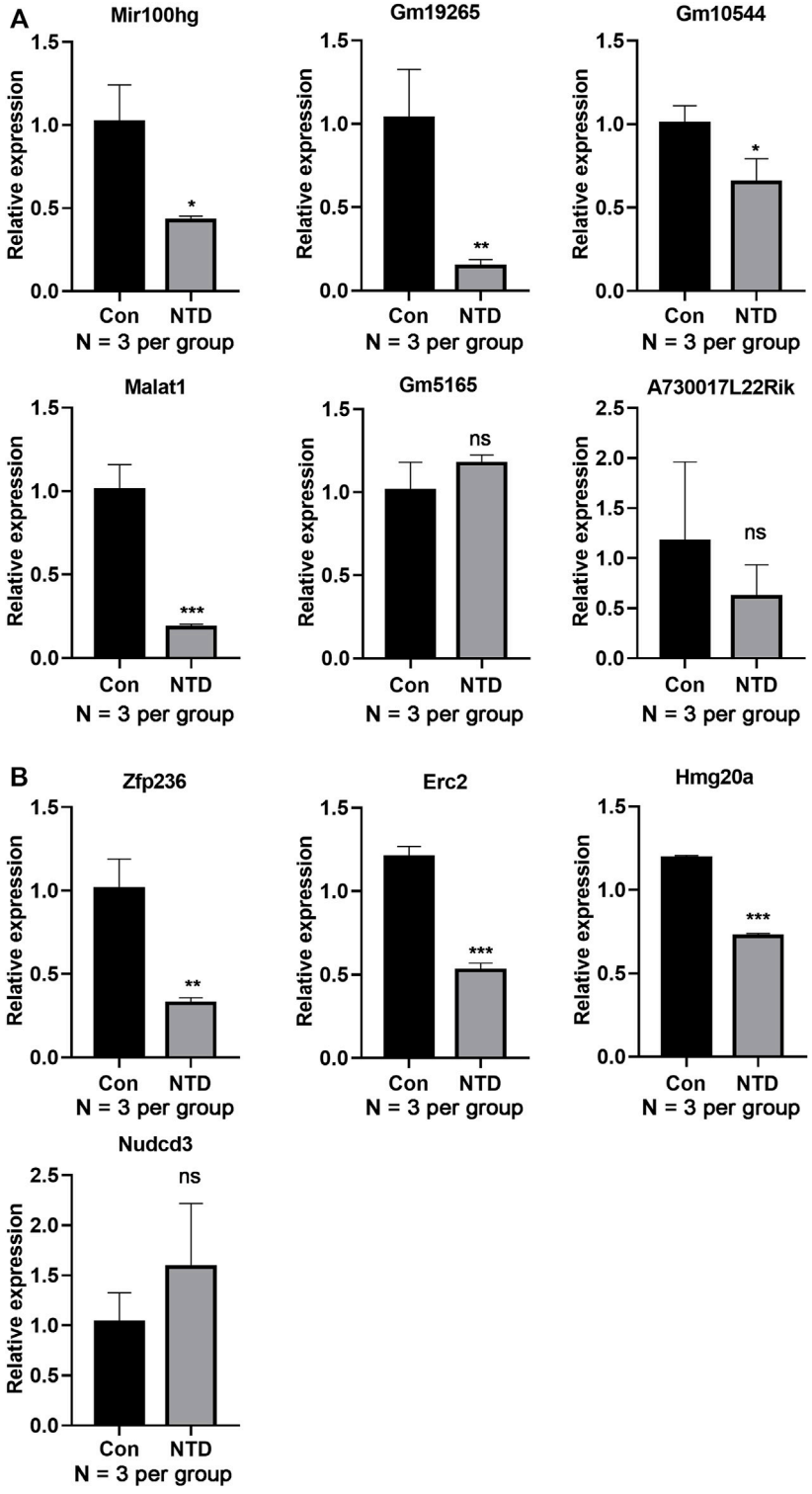


FIGURE 5
qRT-PCR assay verified the lncRNA and gene expression levels in the brain tissues of NTD and control embryos. **(A)** Expression levels of lncRNAs, such as Mir100hg, Gm19265, Gm10544, and Malat1 (N = 3 per group). **(B)** Expression levels of genes, such as *Zfp236*, *Erc2*, and *Hmg20a* (N = 3 per group). Compared to control group, $*p < 0.05$, $**p < 0.01$, and $***p < 0.001$.

NTD development. However, the sample size was small, and the expression levels of the identified m^{6}A-methylated lncRNAs and genes were not validated in neural tube cells isolated from fetal brains. Additional functional experiments are needed to reveal the regulatory mechanisms of key DElncRNAs and DEGs in NTDs.

Data availability statement

The raw data of MeRIP-seq and RNA-seq have been deposited in the NCBI Sequence Read Archive (SRA) database under accession number PRJNA879256, respectively.

Ethics statement

The animal study was reviewed and approved by The Animal Ethics Committee of the Third Affiliated Hospital of Guangzhou Medical University.

Author contributions

JY and MC designed the study. LZ and YL performed the experiment. JY and LZ analyzed the data. JY and MC wrote the manuscript. All authors read and approved the final article.

Funding

This study was supported by the Guangzhou Science and Technology Program (No. 202102010129) and the Hangzhou Science and Technology Program (No. 20211231Y121).

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/fgene.2022.974357/full#supplementary-material

References

Aliperti, V., Skonieczna, J., and Cerase, A. (2021). Long non-coding RNA (lncRNA) roles in cell biology, neurodevelopment and neurological disorders. *Noncoding. RNA* 7, 36. doi:10.3390/ncrna7020036

Avagliano, L., Massa, V., George, T. M., Qureshy, S., Bulfamante, G. P., and Finnell, R. H. (2019). Overview on neural tube defects: From development to physical characteristics. *Birth Defects Res.* 111, 1455–1467. doi:10.1002/bdr2.1380

Bao, X., He, Q., Wang, Y., Huang, Z., and Yuan, Z. (2017). The roles and mechanisms of the Hippo/YAP signaling pathway in the nervous system. *Yi chuan= Hered.* 39, 630–641. doi:10.16288/j.yczz.17-069

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinforma. Oxf. Engl.* 30, 2114–2120. doi:10.1093/bioinformatics/btu170

Cheng, J., Wang, S., Dong, Y., and Yuan, Z. (2020). The role and regulatory mechanism of Hippo signaling components in the neuronal system. *Front. Immunol.* 11, 281. doi:10.3389/fimmu.2020.00281

Cui, X., Meng, J., Zhang, S., Chen, Y., and Huang, Y. (2016). A novel algorithm for calling mRNA m6A peaks by modeling biological variances in MeRIP-seq data. *Bioinforma. Oxf. Engl.* 32, i378–i385. doi:10.1093/bioinformatics/btw281

Cui, X., Zhang, L., Meng, J., Rao, M. K., Chen, Y., and Huang, Y. (2018). MeTDiff: A novel differential RNA methylation analysis for MeRIP-seq data. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 15, 526–534. doi:10.1109/TCBB.2015.2403355

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2013). Star: Ultrafast universal RNA-seq aligner. *Bioinforma. Oxf. Engl.* 29, 15–21. doi:10.1093/bioinformatics/bts635

Finnell, R. H., Caiaffa, C. D., Kim, S. E., Lei, Y., Steele, J., Cao, X., et al. (2021). Gene environment interactions in the etiology of neural tube defects. *Front. Genet.* 12, 659612. doi:10.3389/fgene.2021.659612

Gao, T., Yan, J., Liu, C.-C., Palma, A. S., Guo, Z., Xiao, M., et al. (2019). Chemoenzymatic synthesis of O-mannose glycans containing sulfated or nonsulfated HNK-1 epitope. *J. Am. Chem. Soc.* 141, 19351–19359. doi:10.1021/jacs.9b08964

Gong, C., Fan, Y., and Liu, J. (2021). METTL14 mediated m6A modification to LncRNA ZFAS1/rab22a: A novel therapeutic target for atherosclerosis. *Int. J. Cardiol.* 328, 177. doi:10.1016/j.ijcard.2020.12.002

Gu, Y., Niu, S., Wang, Y., Duan, L., Pan, Y., Tong, Z., et al. (2021). DMDRMR-mediated regulation of m6A-modified CDK4 by m6A reader IGF2BP3 drives ccRCC progression. *Cancer Res.* 81, 923–934. doi:10.1158/0008-5472.CAN-20-1619

Huang, W., Huang, T., Liu, Y., Fu, J., Wei, X., Liu, D., et al. (2021). Nuclear factor I-C disrupts cellular homeostasis between autophagy and apoptosis via miR-200b-Ambra1 in neural tube defects. *Cell Death Dis.* 13, 17–04473. doi:10.1038/s41419-021-04473-2

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. doi:10.1038/ng.3192

Jiang, X., Liu, B., Nie, Z., Duan, L., Xiong, Q., Jin, Z., et al. (2021). The role of m6A modification in the biological functions and diseases. *Signal Transduct. Target. Ther.* 6, 74. doi:10.1038/s41392-020-00450-x

Kakebeen, A. D., and Niswander, L. (2021). Micronutrient imbalance and common phenotypes in neural tube defects. *Genesis* 59, e23455. doi:10.1002/dvg.23455

Kanehisa, M., and Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Kim, D., Langmead, B., and Salzberg, S. L. (2015). Hisat: A fast spliced aligner with low memory requirements. *Nat. Methods* 12, 357–360. doi:10.1038/nmeth.3317

Le, M. T., Xie, H., Zhou, B., Chia, P. H., Rizk, P., Um, M., et al. (2009). MicroRNA-125b promotes neuronal differentiation in human cells by repressing multiple targets. *Mol. Cell. Biol.* 29, 5290–5305. doi:10.1128/MCB.01694-08

Lenihan, J. A., Saha, O., Heimer-McGinn, V., Cryan, J. F., Feng, G., and Young, P. W. (2017). Decreased anxiety-related behaviour but apparently unperturbed NUMB function in ligand of NUMB protein-X (LNX) 1/2 double knockout mice. *Mol. Neurobiol.* 54, 8090–8109. doi:10.1007/s12035-016-0261-0

Li, B., and Dewey, C. N. (2011). Rsem: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinforma.* 12, 323. doi:10.1186/1471-2105-12-323

Li, C., Jiang, Z., Hao, J., Liu, D., Hu, H., Gao, Y., et al. (2021a). Role of N6-methyladenosine modification in mammalian embryonic development. *Genet. Mol. Biol.* 44, e20200253–e20204685. doi:10.1590/1678-4685-GMB-2020-0253

Li, G., Ma, L., He, S., Luo, R., Wang, B., Zhang, W., et al. (2022). WTAP-mediated m(6)A modification of lncRNA NORAD promotes intervertebral disc degeneration. *Nat. Commun.* 13, 1469–28990. doi:10.1038/s41467-022-28990-6

Li, X., Li, K., Chen, Y., and Fang, F. (2021b). The role of Hippo signaling pathway in the development of the nervous system. *Dev. Neurosci.* 43, 263–270. doi:10.1159/000515633

Liu, P., Zhang, B., Chen, Z., He, Y., Du, Y., Liu, Y., et al. (2020a). m(6)A-induced lncRNA MALAT1 aggravates renal fibrogenesis in obstructive nephropathy through the miR-145/FAK pathway. *Aging* 12, 5280–5299. doi:10.18632/aging.102950

Liu, S., Zhuo, L., Wang, J., Zhang, Q., Li, Q., Li, G., et al. (2020b). METTL3 plays multiple functions in biological processes. *Am. J. Cancer Res.* 10, 1631–1646.

Lorenzo, P. I., Martin Vazquez, E., López-Noriega, L., Fuente-Martín, E., Mellado-Gil, J. M., Franco, J. M., et al. (2021). The metabesity factor HMG20A potentiates astrocyte survival and reactive astrogliosis preserving neuronal integrity. *Theranostics* 11, 6983–7004. doi:10.7150/thno.57237

Meng, C., Yang, X., Liu, Y., Zhou, Y., Rui, J., Li, S., et al. (2019). Decreased expression of lncRNA Malat1 in rat spinal cord contributes to neuropathic pain by increasing neuron excitability after brachial plexus avulsion. *J. Pain Res.* 12, 1297–1310. doi:10.2147/JPR.S195117

Meyer, K. D., and Jaffrey, S. R. (2014). The dynamic epitranscriptome: N6-methyladenosine and gene expression control. *Nat. Rev. Mol. Cell Biol.* 15, 313–326. doi:10.1038/nrm3785

Morise, J., Kizuka, Y., Yabuno, K., Tonoyama, Y., Hashii, N., Kawasaki, N., et al. (2014). Structural and biochemical characterization of O-mannose-linked human natural killer-1 glycan expressed on phosphacan in developing mouse brains. *Glycobiology* 24, 314–324. doi:10.1093/glycob/cwt116

Nie, Y., Tian, G. G., Zhang, L., Lee, T., Zhang, Z., Li, J., et al. (2021). Identifying cortical specific long noncoding RNAs modified by m(6)A RNA methylation in mouse brains. *Epigenetics* 16, 1260–1276. doi:10.1080/15592294.2020.1861170

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E. E., Nitsch, R., et al. (2008). A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat. Cell Biol.* 10, 987–993. doi:10.1038/ncb1759

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi:10.1101/gr.1239303

Stockinger, P., Maître, J.-L., and Heisenberg, C.-P. (2011). Defective neuroepithelial cell cohesion affects tangential branchiomotor neuron migration in the zebrafish neural tube. *Development* 138, 4673–4683. doi:10.1242/dev.071233

Sui, X., Hu, Y., Ren, C., Cao, Q., Zhou, S., Cao, Y., et al. (2020). METTL3-mediated m(6)A is required for murine oocyte maturation and maternal-to-zygotic transition. *Cell Cycle* 19, 391–404. doi:10.1080/15384101.2019.1711324

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi:10.1093/nar/gku1003

Tang, Y., Chen, K., Song, B., Ma, J., Wu, X., Xu, Q., et al. (2021). m6A-Atlas: a comprehensive knowledgebase for unraveling the N6-methyladenosine (m6A) epitranscriptome. *Nucleic Acids Res.* 49, D134–d143. doi:10.1093/nar/gkaa692

Tang, Y., Li, M., Wang, J., Pan, Y., and Wu, F. X. (2015). CytoNCA: A cytoscape plugin for centrality analysis and evaluation of protein interaction networks. *Biosystems.* 127, 67–72. doi:10.1016/j.biosystems.2014.11.005

The Gene Ontology Consortium (2019). The gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* 47, D330–D338. doi:10.1093/nar/gky1055

Xie, S. J., Tao, S., Diao, L. T., Li, P. L., Chen, W. C., Zhou, Z. G., et al. (2021). Characterization of long non-coding RNAs modified by m(6)A RNA methylation in skeletal myogenesis. *Front. Cell Dev. Biol.* 9, 762669. doi:10.3389/fcell.2021.762669

Yadav, U., Kumar, P., and Rai, V. (2021). Maternal biomarkers for early prediction of the neural tube defects pregnancies. *Birth Defects Res.* 113, 589–600. doi:10.1002/bdr2.1842

Yang, X., Wu, W., Pan, Y., Zhou, Q., Xu, J., and Han, S. (2020). Immune-related genes in tumor-specific CD4(+) and CD8(+) T cells in colon cancer. *BMC Cancer* 20, 585–07075. doi:10.1186/s12885-020-07075-x

Yen, Y. P., and Chen, J. A. (2021). The m(6)A epitranscriptome on neural development and degeneration. *J. Biomed. Sci.* 28, 40–00734. doi:10.1186/s12929-021-00734-6

Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics a J. Integr. Biol.* 16, 284–287. doi:10.1089/omi.2011.0118

Yu, G., Wang, L. G., and He, Q. Y. (2015). ChIPseeker: An R/bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinforma. Oxf. Engl.* 31, 2382–2383. doi:10.1093/bioinformatics/btv145

Yu, J., Mu, J., Guo, Q., Yang, L., Zhang, J., Liu, Z., et al. (2017). Transcriptomic profile analysis of mouse neural tube development by RNA-Seq. *IUBMB life* 69, 706–719. doi:10.1002/iub.1653

Zhang, L., Cao, R., Li, D., Sun, Y., Zhang, J., Wang, X., et al. (2021a). Ethionine-mediated reduction of S-adenosylmethionine is responsible for the neural tube defects in the developing mouse embryo-mediated m6A modification and is involved in neural tube defects via modulating Wnt/β-catenin signaling pathway. *Epigenetics Chromatin* 14, 52–00426. doi:10.1186/s13072-021-00426-3

Zhang, W., Qian, Y., and Jia, G. (2021b). The detection and functions of RNA modification m(6)A based on m(6)A writers and erasers. *J. Biol. Chem.* 297, 100973. doi:10.1016/j.jbc.2021.100973

Zhang, X., Hamblin, M. H., and Yin, K. J. (2017). The long noncoding RNA Malat1: Its physiological and pathophysiological functions. *RNA Biol.* 14, 1705–1714. doi:10.1080/15476286.2017.1358347

Zhao, J., Zhao, Y., Ma, X., Feng, H., and Jia, L. (2022). Outstanding prognostic value of novel ferroptosis-related genes in chemoresistance osteosarcoma patients. *Sci. Rep.* 12, 5029–09080. doi:10.1038/s41598-022-09080-5

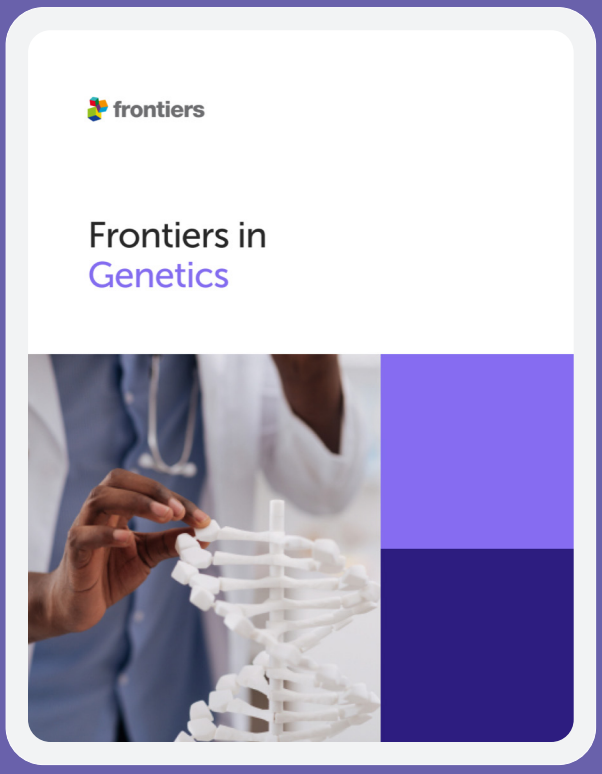
Frontiers in Genetics

Highlights genetic and genomic inquiry relating to all domains of life

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact