GENETIC AND ENVIRONMENTAL FACTORS IN THE OCCURRENCE OF NEONATAL DISORDERS

EDITED BY: Wenhao Zhou, Shoo Kim Lee, Fan Xia, Maria Elisabetta Baldassarre and Mingbang Wang

PUBLISHED IN: Frontiers in Genetics and Frontiers in Pediatrics

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88976-414-3
DOI 10.3389/978-2-88976-414-3

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

GENETIC AND ENVIRONMENTAL FACTORS IN THE OCCURRENCE OF NEONATAL DISORDERS

Topic Editors:
Wenhao Zhou, Fudan University, China
Shoo Kim Lee, University of Toronto, Canada
Fan Xia, Baylor College of Medicine, United States
Maria Elisabetta Baldassarre, University of Bari Aldo Moro, Italy
Mingbang Wang, South China Hospital of Shenzhen University, China

Citation: Zhou, W., Lee, S. K., Xia, F., Baldassarre, M. E., Wang, M., eds. (2022). Genetic and Environmental Factors in the Occurrence of Neonatal Disorders. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-414-3

Table of Contents

frontiers in Genetics

ORIGINAL RESEARCH
published: 02 July 2021
doi: 10.3389/fgene.2021.689071

Bronchopulmonary Dysplasia Predicted by Developing a Machine Learning Model of Genetic and Clinical Information

Dan Dai[1†], Huiyao Chen[2†], Xinran Dong[2], Jinglong Chen[1], Mei Mei[1], Yulan Lu[2], Lin Yang[2,3], Bingbing Wu[2], Yun Cao[4], Jin Wang[4], Wenhao Zhou[2,3,4] and Liling Qian[1,3*]*

[1] Division of Pulmonary Medicine, Children's Hospital of Fudan University, Shanghai, China, [2] Molecular Medical Center, Children's Hospital of Fudan University, Shanghai, China, [3] Shanghai Key Laboratory of Birth Defects, Shanghai, China, [4] Department of Neonatology, Children's Hospital of Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Fan Jin,
Zhejiang University, China

Reviewed by:
Jinchen Li,
Central South University, China
Chiara Di Resta,
Vita-Salute San Raffaele University, Italy

***Correspondence:**
Wenhao Zhou
zhouwenhao@fudan.edu.cn
Liling Qian
llqian@126.com

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

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: 31 March 2021
Accepted: 31 May 2021
Published: 02 July 2021

Citation:
Dai D, Chen H, Dong X, Chen J, Mei M, Lu Y, Yang L, Wu B, Cao Y, Wang J, Zhou W and Qian L (2021) Bronchopulmonary Dysplasia Predicted by Developing a Machine Learning Model of Genetic and Clinical Information. Front. Genet. 12:689071. doi: 10.3389/fgene.2021.689071

Background: An early and accurate evaluation of the risk of bronchopulmonary dysplasia (BPD) in premature infants is pivotal in implementing preventive strategies. The risk prediction models nowadays for BPD risk that included only clinical factors but without genetic factors are either too complex without practicability or provide poor-to-moderate discrimination. We aim to identify the role of genetic factors in BPD risk prediction early and accurately.

Methods: Exome sequencing was performed in a cohort of 245 premature infants (gestational age <32 weeks), with 131 BPD infants and 114 infants without BPD as controls. A gene burden test was performed to find risk genes with loss-of-function mutations or missense mutations over-represented in BPD and severe BPD (sBPD) patients, with risk gene sets (RGS) defined as BPD–RGS and sBPD–RGS, respectively. We then developed two predictive models for the risk of BPD and sBPD by integrating patient clinical and genetic features. The performance of the models was evaluated using the area under the receiver operating characteristic curve (AUROC).

Results: Thirty and 21 genes were included in BPD–RGS and sBPD–RGS, respectively. The predictive model for BPD, which combined the BPD–RGS and basic clinical risk factors, showed better discrimination than the model that was only based on basic clinical features (AUROC, 0.915 *vs*. AUROC, 0.814, $P = 0.013$, respectively) in the independent testing dataset. The same was observed in the predictive model for sBPD (AUROC, 0.907 *vs*. AUROC, 0.826; $P = 0.016$).

Conclusion: This study suggests that genetic information contributes to susceptibility to BPD. The predictive model in this study, which combined BPD–RGS with basic clinical risk factors, can thus accurately stratify BPD risk in premature infants.

Keywords: bronchopulmonary dysplasia, machine learning, prediction model, exome sequencing, premature infants

INTRODUCTION

Bronchopulmonary dysplasia (BPD)—a disorder arising from genetic and environmental risk factors—is one of the most serious complications in premature infants and is responsible for large economic and healthcare burdens. The incidence of BPD in preterm infants [those born with birth weight (BW) between 501 and 1,500 g] varied from 4 to 58.3% in 2003 according to the Vermont Oxford Network data (Payne et al., 2006) and has continued to rise especially among extremely premature infants (Stoll et al., 2010, 2015) due to the improvement in the survival rates of premature infants that benefited from the developments in perinatal medicine and neonatal intensive care. Multiple clinical factors have been implicated in BPD risk, including intrauterine growth restriction (Gortner et al., 1999), low gestational age (GA), low BW, male sex (Korhonen et al., 1999), neonatal respiratory distress syndrome, invasive mechanical ventilation (IMV) (Zhang et al., 2016), sepsis, asphyxia, and—inconsistently—chorioamnionitis (Hartling et al., 2012), race or ethnicity (Rojas et al., 1995; Vesoulis et al., 2020), and mode of delivery (Wang et al., 2014; Chen et al., 2019).

In addition to clinical factors, a considerable amount of research has revealed a genetic basis for BPD and a modulation in the susceptibility of BPD development associated with environmental factors. Genome-wide association studies have identified several single-nucleotide polymorphisms (SNPs) such as rs1245560 in the *SPOCK2* gene of Caucasian and African populations that influence BPD risk (Hadchouel et al., 2011) and other SNPs in SP-A1 (Weber et al., 2000) and SP-B genes (Rova et al., 2004). Numerous rare variants in risk gene sets (RGS) of BPD (BPD–RGS) that are located in pathways related to lung development are also closely associated with the onset and progression of BPD in numerous candidate gene or pathway studies (Carrera et al., 2015; Li et al., 2015; Hadchouel et al., 2020). However, these findings on the genetic etiology of BPD are inconsistent, both reflecting the etiologic molecular heterogeneity of the disease and also showing it to be a complex disease with significant clinical heterogeneity. Despite this, a role for the genetic susceptibility of BPD has been quantified through statistical approaches in two important twin studies (Bhandari, 2006; Lavoie et al., 2008).

Many investigators have carried out research studies in an attempt to predict the development of BPD using clinical factors and respiratory parameters in the neonatal intensive care unit (NICU) of various centers. However, most clinical prediction models are poor to moderate predictors of BPD, e.g., the area under the receiver operating characteristic curve (AUROC) ranged from 0.50 to 0.76 for BPD in an external validation study of a systematic review (Onland et al., 2013). Additionally, the ratio of tidal expiratory flow at 50% of expired volume to peak tidal expiratory flow (as one of the mechanical ventilation parameters) gave an AUROC for the development of moderate/severe BPD (sBPD) of 0.774 (Bentsen et al., 2018); mechanical ventilation at 1 week provided an AUROC for the development of BPD and sBPD of 0.77 and 0.83 (Hunt et al., 2018), respectively; and a multifactorial model that included BW, GA, sex, presence of a hemodynamically significant patent ductus arteriosus (as diagnosed by an echocardiogram), respiratory distress syndrome, hypotension within the first 72 h of life, and intraventricular hemorrhage (IVH) delivered an AUROC for the development of BPD of 0.930—which showed a noticeably improved discriminatory performance, but without favorable clinical maneuverability (Gursoy et al., 2015).

The addition of genetic factors to BPD risk prediction models might improve the ability to accurately predict which infants will develop this significant and serious complication. We therefore developed a risk model for the prediction of BPD and tested the hypothesis that a combined clinical and genetic model that incorporated BPD–RGS would be superior to a clinical-only model.

Abbreviations: ASD, atrial septal defects; AUROC, area under the curve; AUROC, area under the receiver operating characteristic curve; BW, birth weight; BPD, bronchopulmonary dysplasia; CES, clinical exome sequencing; NSV, variant not in unaffected siblings; GA, gestational age; IMV, invasive mechanical ventilation; IUGR, intrauterine growth restriction; IVH, intraventricular hemorrhage; LOF, loss-of-function variants; MIS, missense variants; NICU, neonatal intensive care unit; NRDS, neonatal respiratory distress syndrome; RGS, risk gene sets; sBPD, severe bronchopulmonary dysplasia; SNPs, single-nucleotide polymorphisms; VSD, ventricular septal defects; WES, whole-exome sequencing.

MATERIALS AND METHODS

Design, Setting, and Participants

We conducted a case–control analysis based on a prospective preterm cohort and consecutively recruited 245 infants from the Children's Hospital of Fudan University from January 2017 to May 2019 using the following inclusion criteria: premature infants (defined as born with a GA of less than 32 weeks) who required supplemental oxygen on the first day of postnatal life and who were admitted to the NICU. The exclusion criteria were as follows: (1) infants with significant diseases (e.g., major congenital malformations, clinical syndromes, chromosomal abnormalities, systemic infections and shock, or other definite diseases beyond respiratory), (2) refusal for an infant to participate in the study or withdrawal of an infant from intensive care before tracheal extubation was attempted, (3) infants who did not undergo exome sequencing, and (4) infants who died within 7 days after birth.

Clinical Diagnosis of BPD and Follow-up Collection

Briefly, BPD was diagnosed with respect to a requirement of supplemental oxygen for at least 28 days, with subsequent severity assessment at 36 weeks of postmenstrual age (Jobe and Bancalari, 2001). Perinatal and postnatal information during inpatient hospitalizations and follow-up data were collected until discharge or death, and the timepoints for the collection of clinical characteristics and the corresponding evaluation of outcomes are shown in **Figure 1B**. Data were doubly entered by two

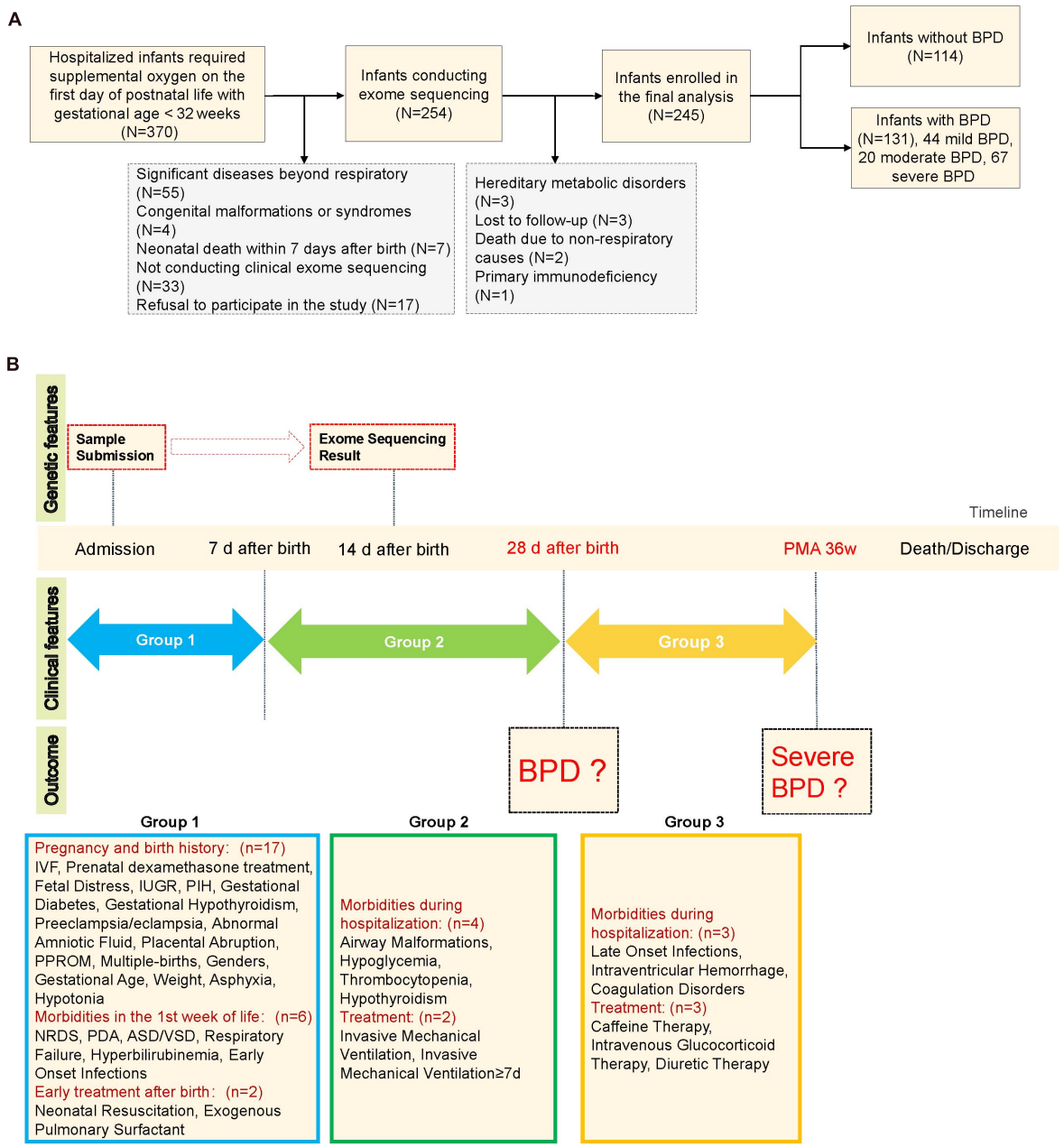


FIGURE 1 | Schematic diagram of the study design. **(A)** Study flow chart. **(B)** Schematic of clinical characteristic collection at each timepoint. Asphyxia was defined as APGAR score of less than seven at 1 min or/and APGAR score of less than seven at 5 min. IVF, *in vitro* fertilization; IUGR, intrauterine growth restriction; PIH, pregnancy-induced hypertension; PPROM, preterm premature rupture of membranes; NRDS, neonatal respiratory distress syndrome; ASD/VSD, atrial septal defect/ventricular septal defect. Patent ductus arteriosus was defined by clinical signs supported by echocardiographic confirmation. Airway malformations: bronchomalacia, tracheomalacia, laryngomalacia, or subglottic stenosis. Early-onset infections: prenatal infection within 72 h of delivery or onset of neonatal pneumonia/sepsis within 7 days of birth. Late-onset infections: infections after 7 days of birth. PMA 36w: postmenstrual age 36 weeks.

clinical physicians, and the inconsistent data were re-evaluated to reach a consensus.

Process of Exome Sequencing

Exome sequencing was performed within 7 days after birth for all 245 infants included in the study, and the sequencing results were obtained within 2 weeks after birth. We received a total of 245 infant samples that came from 206 families—including 169 singleton infants, 35 twins, and two triplets. The samples underwent clinical exome sequencing (CES; n = 234) and whole-exome sequencing (WES, n = 11). In brief, DNA was extracted from peripheral blood specimens using the QIAamp DNA Mini Kit (Qiagen, Pennsylvania, United States) according to the manufacturer's instructions. We subjected the samples

to the Agilent ClearSeq Inherited Disease Kit (for CES) or Agilent Sureselect All Exons Human V5 Kit (for WES) and ran them on the Illumina HiSeq X10 machine, with 150-bp pair-end sequencing. Our detailed sequencing strategy and variant filtration processes were described in detail in our previously published article (Dong et al., 2020). The quality statistics for the 245 samples can be found in **Supplementary Materials**.

Variant Not in Unaffected Siblings Analysis

There were 24 infant samples from 11 families (nine twins and two triplets) that each had BPD infants and infants without BPD. For each comparison (BPD *vs*. control or sBPD *vs*. others), we defined a variant in an infant sample as NSV (variant not in unaffected siblings) if it met both of the following criteria: (1) the sample was involved in the case group of the corresponding comparison and (2) the variant did not exist in its control-sibling samples (who were in the control group of the corresponding comparison). Then, for each of the comparisons, we scored a NSVn value for each of the genes as the total number of variants matching the above-mentioned criteria.

Gene Burden Test

A total of 245 infant samples were used for the gene burden test. For each comparison (BPD *vs*. control or sBPD *vs*. others), we used Fisher's exact test to compare the number of loss-of-function (LOF) variants (either a frameshift, stop–gain, or canonical splice donor/acceptor variant) and missense variants (MIS) of each gene from cases to those from control subjects. In addition, we also considered the pathogenicity of variants and only selected the variants that were predicted as potentially deleterious variants or deleterious variants by any of the three variant effect prediction tools [SIFT (19561590), PolyPhen-2 (20354512), and MutationTaster (24681721)] for Fisher's exact test. A threshold of $P < 0.05$ was considered to be statistically significant, and all tests were one-sided (so as to only find genes that are over-represented in the case groups). Thus, each gene produced two separate *P*-values for LOF and MIS.

Gene Scoring System and Risk Gene Set

For each comparison, we developed a gene scoring system to measure the gene's contribution by combining the results from the gene burden test and NSV analysis. The score for each gene was defined as follows:

$$\text{Score} = \text{NSVn} + 2 \times \left(-\log_{10}\left(P_{\text{LOF}}\right)\right) + \left(-\log_{10}\left(P_{\text{MIS}}\right)\right).$$

We then defined RGS as the genes with a score greater than two. For two comparisons, the gene set was marked as BPD–RGS (BPD *vs*. control) and sBPD–RGS (sBPD *vs*. others).

The genetic predictors were treated as follows: (1) we extracted genes from the corresponding RGS (BPD–RGS or sBPD–RGS), (2) samples with LOF variants from RGS were scored as 1, and the others were scored as 0 (the feature was recorded as RGS_LOF), and (3) samples with MIS variants from RGS were scored as 1, and the others were scored as 0 (this feature was recorded as RGS_MIS).

Clinical Predictors

As the clinical diagnoses of BPD and sBPD occur at different time periods with different sets of clinical features available (**Figure 1B**), we used a total of 31 clinical features (pregnancy and birth history, 17 features; morbidities in the first week of life, six features; early treatment after birth, two features; morbidities during hospitalization, four features; and treatment, two features) for BPD prediction and 37 features for sBPD prediction (morbidities during hospitalization, three features; and treatment, three features). The clinical features were selected based on previous reports or our own experience as likely risk factors for BPD. Missing data were manually imputed by clinical experts. Additionally, we defined the set of basic clinical prediction features with three clinical characteristics—GA, BW, and IMV—as the three most significant predictors for BPD as reported in published studies (Jobe and Bancalari, 2001; Onland et al., 2013).

Sample Size

According to the previously reported clinical prediction models of BPD, the AUROC is about 0.8 (Onland et al., 2013; Bentsen et al., 2018; Hunt et al., 2018). We hypothesized that adding the genetic information into the prediction model of BPD can improve the AUROC value by 0.1. A required sample size consisting of 103 BPD cases and 103 control cases for a power of 80%, with an alpha value of 0.05, was calculated with PASS, version 11.0. Considering potential loss by insufficient data, we increased the sample size by 15%. We had two tasks: the first one was to predict BPD (cases = 131, controls = 114) and the second was to predict the group of sBPD (cases = 67, controls = 178). We collected 245 samples in this study, which we split into a training dataset ($N = 172$) and testing dataset ($N = 73$) randomly.

Statistical Analyses

Distributions of continuous variables were evaluated using the Kruskal–Wallis H, and Shapiro–Wilk's test and Fisher's exact probability test were used to compare the differences in categorical variables. Data are expressed as medians and interquartile ranges (Q1–Q3) for non-parametric distributions, and categorical data are expressed as numbers and percentages. We clustered clinical characteristics by the Ward agglomeration method, and distance was calculated using the Canberra method. The number of clusters was specified as three. A multivariable logistic regression analysis was used to assess independent associations between clinical characteristics and BPD or sBPD, adjusting for GA and BW in 245 infants. The odds ratios (ORs) and 95% confidence intervals were estimated.

We constructed three predictive models for both BPD and sBPD by using different sets of predictors as follows: (1) the set of three basic clinical features (GA, BW, and history of IMV), (2) three basic clinical features with two genetic features (RGS_LOF and RGS_MIS), and (3) all clinical features selected by LASSO (R package: glmnet, 15 features for BPD prediction and 19 for sBPD

prediction, **Supplementary Figure 1**) with two genetic features (RGS_LOF and RGS_MIS).

For each predictive model, we applied LASSO (R package: glmnet) to generate the predictive model on the training dataset. A 10-fold cross-validation was performed to select the optimal lambda (penalty for the number of characteristics), which determined the performance of the lasso model (number of features included in the model and predictive deviations). We then evaluated the predictive performance for the three lasso models in independent testing datasets. The area under the receiver operating characteristic curve (AUROC) was used for model evaluation, and the statistical differences between models were tested using Delong's test. We performed all analyses using R software (version 3.6.0, [1]).

Ethics and Informed Consent

This study was approved by the Ethics Committee of the Children's Hospital of Fudan University (no. 2016-97), and informed consent for DNA analysis of peripheral blood cells was obtained from the study participants in accordance with the time of collection.

[1]http://cran.r-project.org

RESULTS

Patient Cohort of Preterm Infants

A consecutively enrolled total of 370 premature infants with a GA of <32 weeks were admitted to our NICU during the study period. One hundred sixteen infants were ruled out by the exclusion criteria, and nine (four for clear etiology, three lost to follow-up, and two died of non-respiratory causes) were excluded from further analysis; ultimately, 245 infants were included in the final analysis (**Figure 1A**). There were 131 (53.5%) infants diagnosed with BPD, including 67 (51.1%) with sBPD (including nine infants who died due to respiratory dysfunction before postmenstrual age 36 weeks), 20 (15.3%) with moderate BPD, and 44 (33.6%) with mild BPD. The infants with BPD exhibited lower GA (28.1 *vs*. 30 weeks, $P < 0.001$) and BW (1,090 *vs*. 1,332.5 g, $P < 0.001$) compared to infants with no BPD. Furthermore, the BPD infants manifested a higher rate of IMV (77.9 *vs*. 35.1%, $P < 0.001$), IMV ≥7 days (55 *vs*. 6.1%, $P < 0.001$), early-onset infections (<7 days) (87 *vs*. 70.2%, $P = 0.002$), late-onset infections (≥7 days) (45.8 *vs*. 27.2%, $P = 0.004$), grades III and IV IVH (42 *vs*. 17.5%, $P < 0.001$), and death (14.5 *vs*. 0, $P < 0.001$, **Table 1**). We noted that nine of the 19 infants who died before the evaluation of BPD severity had been assigned

TABLE 1 | Clinical characteristics of 245 premature infants.

Characteristic	Total infants (*n* = 245)	BPD (*n* = 131)	No BPD (*n* = 114)	*P*-value	Mild BPD (*n* = 44)	Moderate BPD (*n* = 20)	Severe BPD (*n* = 67)	*P*-value
Gestational age [week, median (IQR)]	29.1 (27.9–30.6)	28.1 (27.1–29.8)	30 (28.6–31.1)	< 0.001	29 (27.8–29.7)	28.8 (27.5–30.6)	27.6 (26.6–29.4)	< 0.001
Birth weight [g, median (IQR)]	1,200 (1,020–1,410)	1,090 (945–1,260)	1,332.5 (1,131.3–1,570)	< 0.001	1,160 (1,072.5–1,280)	1,110 (1,017.5–1,271.3)	1,035 (915–1,200)	< 0.001
Sex (male), no. (%)	140 (57.1%)	80 (61.1%)	60 (52.6%)	0.229	26 (59.1%)	13 (65%)	41 (61.2%)	0.579
Multiple births, no. (%)	117 (47.8%)	72 (55%)	45 (39.5%)	0.022	27 (61.4%)	13 (65%)	32 (47.8%)	0.032
Neonatal respiratory distress syndrome, no. (%)	194 (79.2%)	113 (86.3%)	81 (71.1%)	0.006	40 (90.9%)	18 (90%)	55 (82.1%)	0.019
Respiratory failure, no. (%)	229 (93.5%)	131 (100%)	98 (86%)	< 0.001	44 (100%)	20 (100%)	67 (100%)	< 0.001
Atrial septal defect or ventricular septal defect, no. (%)	63 (25.7%)	41 (31.3%)	22 (19.3%)	0.046	6 (13.6%)	9 (45%)	26 (38.8%)	0.001
Patent ductus arteriosus, no. (%)	205 (83.7%)	118 (90.1%)	87 (76.3%)	0.006	40 (90.9%)	19 (95%)	59 (88.1%)	0.029
Thrombocytopenia, no. (%)	27 (11%)	23 (17.6%)	4 (3.5%)	0.001	2 (4.5%)	4 (20%)	17 (25.4%)	< 0.001
Coagulation disorders, no. (%)	73 (29.8%)	52 (39.7%)	21 (18.4%)	< 0.001	10 (22.7%)	9 (45%)	33 (49.3%)	< 0.001
Intraventricular hemorrhage (grades III and IV), no. (%)	75 (30.6%)	55 (42%)	20 (17.5%)	< 0.001	13 (29.5%)	10 (50%)	32 (47.8%)	< 0.001
Early-onset infections (<7 days), no. (%)	194 (79.2%)	114 (87%)	80 (70.2%)	0.002	35 (79.5%)	16 (80%)	63 (94%)	0.002
Late-onset infections (≥7 days), no. (%)	91 (37.1%)	60 (45.8%)	31 (27.2%)	0.004	14 (31.8%)	13 (65%)	33 (49.3%)	0.001
Airway malformations, no. (%)	11 (4.5%)	10 (7.6%)	1 (0.9%)	0.025	1 (2.3%)	1 (5%)	8 (11.9%)	0.005
Invasive mechanical ventilation (IMV), no. (%)	142 (58%)	102 (77.9%)	40 (35.1%)	< 0.001	24 (54.5%)	17 (85%)	61 (91%)	< 0.001
IMV ≥7 days, no. (%)	79 (32.2%)	72 (55%)	7 (6.1%)	< 0.001	10 (22.7%)	11 (55%)	51 (76.1%)	< 0.001
Exogenous pulmonary surfactant, no. (%)	170 (69.4%)	102 (77.9%)	68 (59.6%)	0.003	32 (72.7%)	15 (75%)	55 (82.1%)	0.013
Intravenous glucocorticoid therapy, no. (%)	26 (10.6%)	25 (19.1%)	1 (0.9%)	< 0.001	2 (4.5%)	1 (5%)	22 (32.8%)	< 0.001
Caffeine therapy, no. (%)	197 (80.4%)	116 (88.5%)	81 (71.1%)	0.001	40 (90.9%)	19 (95%)	57 (85.1%)	0.005
Death, no. (%)	19 (7.8%)	19 (14.5%)	0	< 0.001	0	0	19 (28.4%)	< 0.001

All summary data are medians (25–75% percentile) or counts (%).
wk, week; gr, gram.

to the sBPD group. Other clinical characteristics are depicted in **Supplementary Table 2**.

Clinical Risk Factors for BPD and Severity of BPD in Preterm Infants

Twenty-one clinical characteristics that significantly differed between BPD and controls or between any two BPD levels ($p < 0.05$) were selected. We grouped these clinical features into three clusters according to their co-existence with sBPD and mBPD (mild and moderate BPD), as shown in **Figure 2A**. Generally, Cluster1 features were related to the general situation at birth and early complications within 7 days after birth. Congenital airway anomaly and congenital heart disease were principally shown in Cluster2 features. Cluster3 features were primarily involved in late complications beyond 7 days after birth and included a condition of IMV. We then validated the independent contribution to the clinical features between case and control groups in three comparisons using multivariable logistic regression and adjusting for GA and BW. The OR for each feature in each of the comparisons is shown in **Figure 2B**. IMV ≥7 days and airway malformations were both the top two clinical factors for BPD (OR = 14.209, 95% CI, 6.102–38.112 and OR = 10.485, 95% CI, 1.739–202.955, respectively) and sBPD development (OR = 11.686, 95% CI, 5.724–24.867 and OR = 10.954, 95% CI, 2.604–58.863, respectively). Furthermore, the need for intravenous glucocorticoid therapy was higher in infants with sBPD (OR = 21.308, 95% CI, 6.792–84.288) relative to the other groups. Atrial septal defects and/or ventricular septal defects and coagulation disorders were risk factors

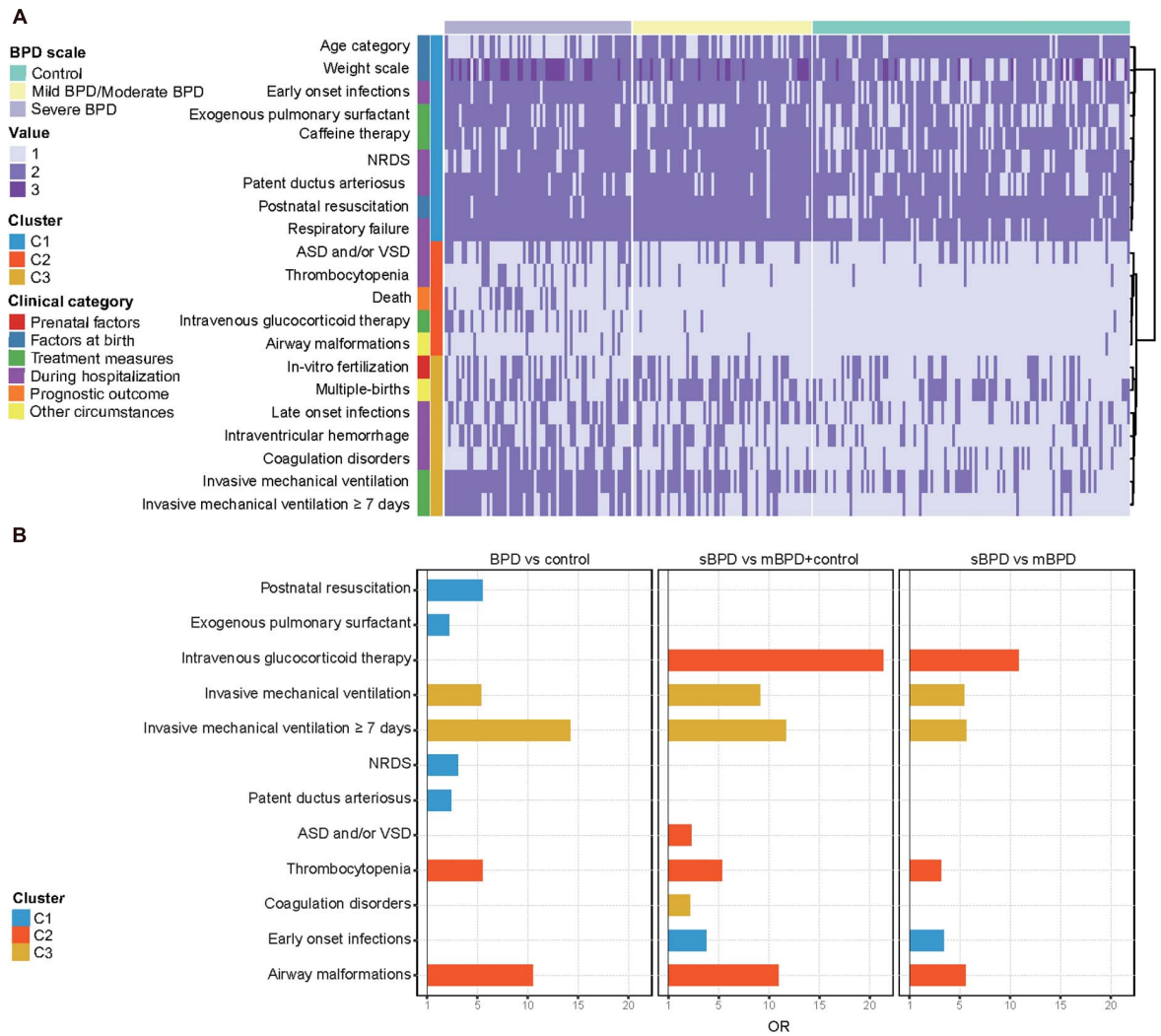


FIGURE 2 | Distribution of infant clinical characteristics. **(A)** Cluster analysis of clinical characteristics of all infants. **(B)** Significant clinical characteristics in three comparisons. Twenty-one clinical characteristics that significantly differed between bronchopulmonary dysplasia (BPD) and controls or between any two BPD levels ($p < 0.05$) were selected. Weight scale including low BW (1,500–2,499 g), very low BW (1,000–1,499 g), and extremely low BW (<1,000 g) were represented by values 1–3, respectively. Age category including extremely preterm (<28 weeks) and very preterm (28 to 32 weeks) was represented by values 1 and 2, respectively. ASD/VSD, atrial septal defect/ventricular septal defects; NRDS, neonatal respiratory distress syndrome; sBPD, severe BPD; mBPD, mild and moderate BPD.

for sBPD compared to other groups, but not in sBPD *vs*. mBPD comparison.

Risk Genes for BPD and Severe BPD

A total of 30 BPD–RGS and 21 sBPD–RGS (with a score of significant genes >2) were identified based on the gene burden test and NSV analysis in the comparison of BPD *vs*. control and sBPD *vs*. others, respectively (**Table 2** and **Supplementary Table 2**). Most of these genes (such as *OBSL1*, *GNAS*, *TCIRG1*, and *C5*) are involved in susceptibility to infection and inflammatory response, cellular and immune regulation, cellular biologic function, and metabolic biologic processes closely related to early development and organogenesis, which are important in the occurrence of biologic dysfunction in BPD development (**Supplementary Figure 2**). In addition, there were 16 (53.3%) and eight (38.1%) genes that also appeared in the set of significant genes, respectively, when using potentially deleterious variants or deleterious variants for the burden test (**Supplementary Table 3**). The RGS obtained in the study showed a very little overlap with other reported studies and was similar to that for studies that supported a large genetic heterogeneity (**Supplementary Figure 3**).

Machine Learning Model Generation and Testing for the Prediction of BPD or Severe BPD

Herein we integrated clinical and genetic features to predict the risk of BPD and sBPD by using a machine learning model. Different sets of clinical features were used for each task (**Figure 1B**) in combination with two genetic predictors (BPD–RGS or sBPD–RGS for LOF and MIS), and ROC analysis was used to evaluate the performance of the predictive models. We obtained an excellent predictive result using basic clinical characteristics combined with BPD–RGS (AUROC of 0.915, 0.843–0.987) compared with only using basic clinical characteristics (AUROC of 0.814, 0.718–0.911) (P = 0.013), and the results were similar when using all clinical characteristics combined with the BPD–RGS variant burden (AUROC, 0.953; 0.911–0.996) (P = 0.183; **Figure 3A**). For the sBPD prediction model, the model results revealed a higher AUROC value (0.907; 0.830–0.984) using basic clinical characteristics combined with the BPD–RGS variant burden compared with using basic clinical characteristics (0.826, 0.712–0.939) (P = 0.016; **Figure 3B**). Surprisingly, when we built prediction models using the associated genes found by potentially deleterious variants or deleterious variants, we did not find a significant improvement in the accuracy of the two models (model BPD: AUROC, 0.872 *vs*. AUROC, 0.814, P = 0.125; model sBPD: AUROC, 0.881 *vs*. AUROC, 0.827, P = 0.32) (**Supplementary Figure 4**). Additionally, we found that genetic factors also contributed to the prediction model with respect to deaths in infants with BPD (AUROC, 0.891 *vs*. AUROC, 0.859, P = 0.258) (**Supplementary Figure 5**); however, we posit that its predictive power was not robust due to the limited number of samples from deceased infants, and more evidence in a larger cohort is therefore needed to support this finding.

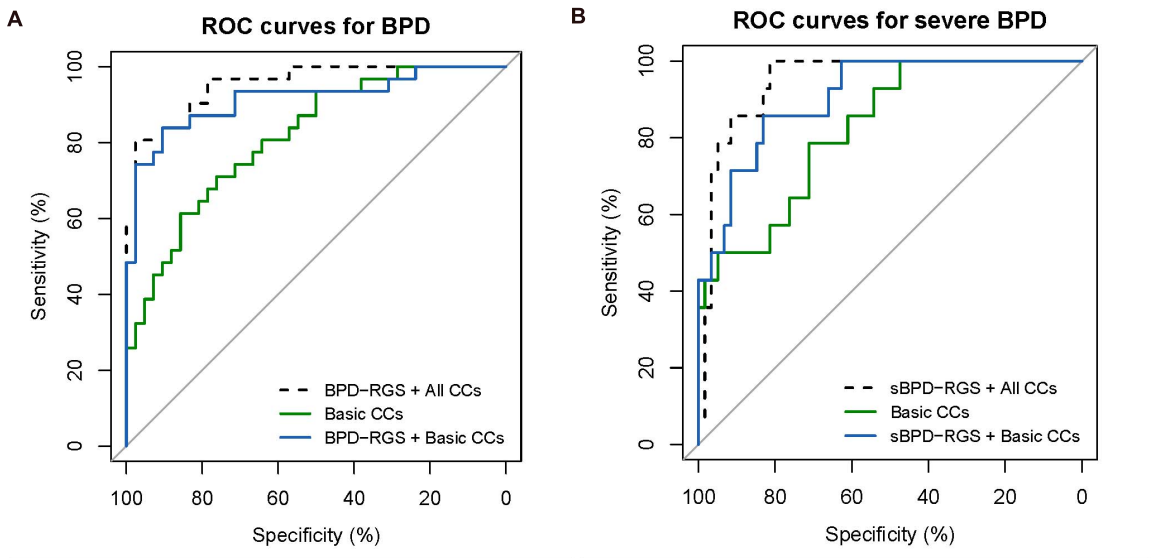
DISCUSSION

Bronchopulmonary dysplasia remains the most common long-term respiratory morbidity affecting prematurely born infants and has been a significant burden on families and public healthcare resources. Its multi-factorial pathogenesis arises from

TABLE 2 | Genes with a significant burden for LOF/MIS variants.

Comparison	Gene	Samples	Case (*n*) LOF	Case (*n*) MIS	Control (*n*) LOF	Control (*n*) MIS	Control (*n*)	Case (*n*)	LOF *P*-value	MIS *P*-value	NSV (*n*)	Score
BPD *vs*. control	OBSL1	31	11	18	0	13	114	131	< 0.001	0.362	1	7.600
BPD *vs*. control	NTRK1	34	1	15	1	2	114	131	0.785	0.002	2	4.884
BPD *vs*. control	CHRNA4	10	0	10	1	0	114	131	1.000	0.002	1	3.791
BPD *vs*. control	PDE11A	24	7	14	2	3	114	131	0.124	0.011	0	3.766
BPD *vs*. control	FRG1	24	9	10	2	3	114	131	0.049	0.070	0	3.764
BPD *vs*. control	SPTAN1	26	0	19	0	4	114	131	1.000	0.002	1	3.613
BPD *vs*. control	DCC	17	1	12	0	2	114	131	0.535	0.011	1	3.507
BPD *vs*. control	BDP1	14	3	10	0	2	114	131	0.151	0.030	0	3.159
BPD *vs*. control	C5	10	1	7	1	0	114	131	0.785	0.012	1	3.147
sBPD *vs*. mBPD/control	ACADSB	10	5	1	3	6	178	67	0.037	0.897	2	4.907
sBPD *vs*. mBPD/control	TCIRG1	15	3	5	0	13	178	67	0.020	0.578	1	4.646
sBPD *vs*. mBPD/control	OBSL1	31	7	10	4	21	178	67	0.011	0.323	0	4.408
sBPD *vs*. mBPD/control	FGFR3	17	0	12	0	9	178	67	1.000	0.003	1	3.594
sBPD *vs*. mBPD/control	BDP1	14	3	2	0	10	178	67	0.020	0.887	0	3.459
sBPD *vs*. mBPD/control	RBBP8	9	0	7	0	3	178	67	1.000	0.005	1	3.290
sBPD *vs*. mBPD/control	SPG7	20	0	8	1	5	178	67	1.000	0.008	1	3.077
sBPD *vs*. mBPD/control	GNAS	17	0	11	0	6	178	67	1.000	< 0.001	0	3.028

LOF/MIS, loss-of-function variants and missense variants; sBPD, severe bronchopulmonary dysplasia; mBPD, mild and moderate bronchopulmonary dysplasia; NSV, variant not in unaffected siblings.



	AUC	95% CI	*P*	*P*
BPD-RGS + All CCs	0.953	0.911–0.996		Ref
Basic CCs	0.814	0.718–0.911	Ref	
BPD-RGS + Basic CCs	0.915	0.843–0.987	0.013	0.183

	AUC	95% CI	*P*	*P*
sBPD-RGS + All CCs	0.948	0.898–0.998		Ref
Basic CCs	0.826	0.712–0.939	Ref	
sBPD-RGS + Basic CCs	0.907	0.830–0.984	0.016	0.161

FIGURE 3 | Receiver operating characteristic (ROC) analyses of predictive models for infants with bronchopulmonary dysplasia (BPD) or severe BPD (sBPD). The comparisons of predictive models for BPD and sBPD. *P*-values show the areas under the ROC curves (AUROCs) between the different models. Clinical characteristics include the variables in a separate lasso model (**Supplementary Figure 1**). **(A)** ROC analyses of the prediction of BPD by the combination of BPD–risk gene set (RGS) and all clinical characteristic model, the basic clinical characteristic model, and the combined BPD–RGS and basic clinical characteristics model. **(B)** ROC analyses of the prediction of severe BPD by the combined sBPDâĂŞRGS and all clinical characteristic model, the basic clinical characteristic model, and the combined sBPD–RGS and basic clinical characteristics model. CCs, clinical characteristics.

a complex interaction between genetic and environmental influences, and a plethora of studies are devoted to the early prediction of the development of BPD as well as the severity of BPD and poor prognosis. Clinical characteristics such as GA, BW, race and ethnicity, gender, APGAR score, and amount of oxygen administered/positive inspiratory pressure/mean airway pressure are the common predictors in a variety of predictive models described in previous studies. Despite these efforts, no reliable and reproducible risk stratification model has been found that would allow an early diagnosis and ease of application within NICUs. A single predictive factor cannot accurately predict the BPD outcome because BPD is a multifactorial disease, and therefore current multivariate modeling demonstrates that the greater the number of clinical predictor variables, the better the modeling results tend to be (Laughon et al., 2011; Gursoy et al., 2015). However, this also increases complexity and impairs practicality in clinical practice. In our cohort of premature infants, we found that low BW, low GA, IMV, IMV ≥7 days, IVH, and early- and/or late-onset infections were risk factors for BPD—and this was also demonstrated in numerous other studies (Lapcharoensap et al., 2015). In addition, we uncovered a higher proportion of intravenous glucocorticoid therapy use in infants with sBPD, which may be explained by the severe fundamental lung conditions observed in infants who later developed sBPD.

That there is a significant contribution of genetic factors to the development of BPD is well established, and considerable research effort has been directed toward elucidating the genetic etiologies and pathogenic genetic mechanisms involved. Although these studies yielded inconsistent and sometimes controversial results in different cohorts due to the vast heterogeneity of infants with BPD, they still provided very valuable knowledge and increased our understanding of the nature of BPD development at the DNA level. To overcome this dilemma, several studies have already focused on the genetic basis for BPD in infants with extreme phenotypes of early respiratory morbidity and infants who exhibited a good response to certain therapies (Hamvas et al., 2018; Torgerson et al., 2018). Advancements in the care of patients with BPD over the next decades are dependent upon improved understanding and the use of disease phenotyping in infants with BPD to enable better risk stratification and targeted therapeutic interventions. Thus, to better predict BPD development and the severity of BPD—and to identify those infants at greatest risk for poor outcomes—there is a need to identify perinatal risk factors on the basis of genetic predisposition.

Unfortunately, there are a few studies that focus on genetic aspects combined with clinical risk factors in the prediction of BPD development. In the present study, we developed a scoring system to identify the gene burden in infants with BPD and obtained BPD–RGS (with scores >2) with most of the genes tightly correlated with drug metabolism, immune response, and organ development as described in the functional enrichment analysis—and thus playing essential roles in the

proper development and function of the postnatal lung. We further constructed predictive models of BPD development using BPD–RGS combined with clinical factors; intriguingly, the use of gene burdens resulted in a significant increase of 10% in the AUROC of the model using only basic clinical characteristics (GA, BW, and IMV) for BPD and 8% for sBPD. These results suggested that we can classify the probable status of the disease in admitted patients using a small amount of clinical information and the patients' BPD–RGS sequencing information and that this can then be further validated in a larger independent cohort. However, the findings of significant improvement in the accuracy of the two predictive models mentioned above are no longer present when prediction models were built using the associated genes found by potentially deleterious or deleterious variants, suggesting that some variants of undetermined significance may also have certain effects on protein function especially considering the overall consequences of multiple variants.

Our findings would allow an early and accurate identification of infants who could potentially benefit from focused therapy and would take a significant step forward in the comprehensive and personalized care of individuals with BPD. However, despite our encouraging results, we still need to accrue additional genetic study cohorts or enlarge our present cohort to develop a pathway scoring system that allows the progression of BPD–RGS into the pathway level, overcoming and stratifying the tremendous genetic heterogeneity that is present.

Some limitations to this study should also be acknowledged. A single-center design leads to missing data and unavoidable biases in identifying and recruiting participants. Further validation beyond this initial exploratory cohort is warranted. In order to demonstrate the robustness of the results, we intend to further verify this predictive model in another external cohort (currently recruiting). Despite these limitations, this study was the first one designed to combine clinical factors and genetic variations so as to predict BPD occurrence and severity. Thus, our model might assist clinicians in the earlier diagnosis of BPD, guide clinical therapy and prognosis, allow for appropriate decision-making, and optimize the use of hospital resources.

CONCLUSION

We first determined the genetic contribution in predictive models of BPD development and showed that three basic clinical characteristics combined with BPD–RGS achieved a high prediction accuracy of models that predicted BPD development and its severity.

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 the Children's Hospital of Fudan University (approval number: 2016-97). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Informed consent was obtained from the parents of patients who were recruited in the present study.

AUTHOR CONTRIBUTIONS

LQ and WZ contributed to research conception and design and take responsibility for the accuracy and integrity of the data analysis. YC, JW, and MM contributed to the patient enrollment, sample collection, and follow-up. BW, YL, LY, and XD contributed to the genetic testing and bioinformatics analysis and drafted the report. DD, HC, and JC contributed to the data acquisition, analysis, and interpretation. DD and HC contributed to the drafting of the manuscript. LQ contributed to the critical revision of the manuscript. All authors approved the final manuscript and had full access to all data in the study.

FUNDING

This work was supported by a grant from the National Natural Science Foundation of China (no. 8167060762).

ACKNOWLEDGMENTS

We thank all the patients and their families for their participation in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.689071/full#supplementary-material

REFERENCES

Bentsen, M. H., Markestad, T., and Halvorsen, T. (2018). Ventilator flow data predict bronchopulmonary dysplasia in extremely premature neonates. *ERJ Open Res.* 4, 00099–2017. doi: 10.1183/23120541.00099-2017

Bhandari, V. (2006). Familial and genetic susceptibility to major neonatal morbidities in preterm twins. *Pediatrics* 117, 1901–1906. doi: 10.1542/peds.2005-1414

Carrera, P., Di Resta, C., Volonteri, C., Castiglioni, E., Bonfiglio, S., Lazarevic, D., et al. (2015). BPD and genetics study group. exome sequencing and pathway analysis for identification of genetic variability relevant for bronchopulmonary dysplasia (BPD) in preterm newborns: a pilot study. *Clin. Chim. Acta* 451(Pt A), 39–45. doi: 10.1016/j.cca.2015.01.001

Chen, X., Lin, B., Xiong, X., Sun, P., Kong, Y., and Yang, C. (2019). The Utility of Comprehensive metabolic panel tests for the prediction of bronchopulmonary

dysplasia in extremely premature infants. *Dis. Markers* 2019: 5681954.

Dong, X., Liu, B., Yang, L., Wang, H., Wu, B., Liu, R., et al. (2020). Clinical exome sequencing as the first-tier test for diagnosing developmental disorders covering both CNV and SNV: a Chinese cohort. *J. Med. Genet.* 57, 558–566. doi: 10.1136/jmedgenet-2019-106377

Gortner, L., Wauer, R. R., Stock, G. J., Reiter, H. L., Reiss, I., Jorch, G., et al. (1999). Neonatal outcome in small for gestational age infants: do they really better? *J. Perinat. Med.* 27, 484–489.

Gursoy, T., Hayran, M., Derin, H., and Ovali, F. (2015). A clinical scoring system to predict the development of bronchopulmonary dysplasia. *Am. J. Perinatol.* 32, 659–666. doi: 10.1055/s-0034-1393935

Hadchouel, A., Decobert, F., Besmond, C., and Delacourt, C. (2020). Exome sequencing of extreme phenotypes in bronchopulmonary dysplasia. *Eur. J. Pediatr.* 179, 579–586. doi: 10.1007/s00431-019-03535-0

Hadchouel, A., Durrmeyer, X., Bouzigon, E., Incitti, R., Huusko, J., Jarreau, P. H., et al. (2011). Identification of SPOCK2 as a susceptibility gene for bronchopulmonary dysplasia. *Am. J. Respir. Crit. Care Med.* 184, 1164–1170.

Hamvas, A., Feng, R., Bi, Y., Wang, F., Bhattacharya, S., Mereness, J., et al. (2018). Exome sequencing identifies gene variants and networks associated with extreme respiratory outcomes following preterm birth. *BMC Genet.* 19:94.

Hartling, L., Liang, Y., and Lacaze-Masmonteil, T. (2012). Chorioamnionitis as a risk factor for bronchopulmonary dysplasia: a systematic review and meta-analysis. *Arch. Dis. Child. Fetal. Neonatal. Ed.* 97, F8–F17.

Hunt, K. A., Dassios, T., Ali, K., and Greenough, A. (2018). Prediction of bronchopulmonary dysplasia development. *Arch. Dis. Child Fetal. Neonatal. Ed.* 103, F598–F599.

Jobe, A. H., and Bancalari, E. (2001). Bronchopulmonary dysplasia. *Am. J. Respir. Crit. Care Med.* 163, 1723–1729.

Korhonen, P., Tammela, O., Koivisto, A. M., Laippala, P., and Ikonen, S. (1999). Frequency and risk factors in bronchopulmonary dysplasia in a cohort of very low birth weight infants. *Early Hum. Dev.* 54, 245–258. doi: 10.1016/s0378-3782(98)00101-7

Lapcharoensap, W., Gage, S. C., Kan, P., Profit, J., Shaw, G. M., Gould, J. B., et al. (2015). Hospital variation and risk factors for bronchopulmonary dysplasia in a population-based cohort. *JAMA Pediatr.* 169:e143676. doi: 10.1001/jamapediatrics.2014.3676

Laughon, M. M., Langer, J. C., Bose, C. L., Smith, P. B., Ambalavanan, N., Kennedy, K. A., et al. (2011). Prediction of bronchopulmonary dysplasia by postnatal age in extremely premature infants. *Am. J. Respir. Crit. Care Med.* 183, 1715–1722.

Lavoie, P. M., Pham, C., and Jang, K. L. (2008). Heritability of bronchopulmonary dysplasia, defined according to the consensus statement of the national institutes of health. *Pediatrics* 122, 479–485. doi: 10.1542/peds.2007-2313

Li, J., Yu, K. H., Oehlert, J., Jeliffe-Pawlowski, L. L., Gould, J. B., Stevenson, D. K., et al. (2015). Exome sequencing of neonatal blood spots and the identification of genes implicated in bronchopulmonary dysplasia. *Am. J. Respir. Crit. Care Med.* 192, 589–596. doi: 10.1164/rccm.201501-0168oc

Onland, W., Debray, T. P., Laughon, M. M., Miedema, M., Cools, F., Askie, L. M., et al. (2013). Clinical prediction models for bronchopulmonary dysplasia: a systematic review and external validation study. *BMC Pediatr.* 13:207.

Payne, N. R., LaCorte, M., Karna, P., Chen, S., Finkelstein, M., Goldsmith, J. P., et al. (2006). Reduction of bronchopulmonary dysplasia after participation in the breathsavers group of the vermont oxford network neonatal intensive care quality improvement collaborative. *Pediatrics* 118(Suppl. 2), S73–S77.

Rojas, M. A., Gonzalez, A., Bancalari, E., Claure, N., Poole, C., and Silva-Neto, G. (1995). Changing trends in the epidemiology and pathogenesis of neonatal chronic lung disease. *J. Pediatr.* 126, 605–610. doi: 10.1016/s0022-3476(95)70362-4

Rova, M., Haataja, R., Marttila, R., Ollikainen, V., Tammela, O., and Hallman, M. (2004). Data mining and multiparameter analysis of lung surfactant protein genes in bronchopulmonary dysplasia. *Hum. Mol. Genet.* 13, 1095–1104. doi: 10.1093/hmg/ddh132

Stoll, B. J., Hansen, N. I., Bell, E. F., Shankaran, S., Laptook, A. R., Walsh, M. C., et al. (2010). Neonatal outcomes of extremely preterm infants from the NICHD neonatal research network. *Pediatrics* 126, 443–456.

Stoll, B. J., Hansen, N. I., Bell, E. F., Walsh, M. C., Carlo, W. A., Shankaran, S., et al. (2015). Eunice trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA* 314, 1039–1051.

Torgerson, D. G., Ballard, P. L., Keller, R. L., Oh, S. S., Huntsman, S., Hu, D., et al. (2018). Ancestry and genetic associations with bronchopulmonary dysplasia in preterm infants. *Am. J. Physiol. Lung Cell Mol. Physiol.* 315, L858–L869.

Vesoulis, Z. A., McPherson, C. C., and Whitehead, H. V. (2020). Racial disparities in calculated risk for bronchopulmonary dysplasia: a dataset. *Data Brief.* 11:105674. doi: 10.1016/j.dib.2020.105674

Wang, K., Huang, X., Lu, H., and Zhang, Z. (2014). A comparison of KL-6 and Clara cell protein as markers for predicting bronchopulmonary dysplasia in preterm infants. *Dis. Markers* 2014, 736536.

Weber, B., Borkhardt, A., Stoll-Becker, S., Reiss, I., and Gortner, L. (2000). Polymorphisms of surfactant protein A genes and the risk of bronchopulmonary dysplasia in preterm infants. *Turk J. Pediatr.* 42, 181–185.

Zhang, H., Zhang, J., and Zhao, S. (2016). Airway damage of prematurity: the impact of prolonged intubation, ventilation, and chronic lung disease. *Semin. Fetal. Neonatal. Med.* 21, 246–253. doi: 10.1016/j.siny.2016.04.001

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 in Genetics

ORIGINAL RESEARCH
published: 04 August 2021
doi: 10.3389/fgene.2021.677748

High-Frequency Exon Deletion of DNA Cross-Link Repair 1C Accounting for Severe Combined Immunodeficiency May Be Missed by Whole-Exome Sequencing

Feifan Xiao[1], Yulan Lu[1], Bingbing Wu[1], Bo Liu[1], Gang Li[1], Ping Zhang[1], Qinhua Zhou[2], Jinqiao Sun[2], Huijun Wang[1] and Wenhao Zhou[1,3]**

[1]Center for Molecular Medicine, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [2]Department of Immunology, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China, [3]Key Laboratory of Neonatal Diseases, Ministry of Health, Department of Neonates, Children's Hospital of Fudan University, National Children's Medical Center, Shanghai, China

Next-generation sequencing (NGS) has been used to detect severe combined immunodeficiency (SCID) in patients, and some patients with DNA cross-link repair 1C (*DCLRE1C*) variants have been identified. Moreover, some compound variants, such as copy number variants (CNV) and single nucleotide variants (SNV), have been reported. The purpose of this study was to expand the genetic data related to patients with SCID carrying the compound *DCLRE1C* variant. Whole-exome sequencing (WES) was performed for genetic analysis, and variants were verified by performing Sanger sequencing or quantitative PCR. Moreover, we searched PubMed and summarized the data of the reported variants. Four SCID patients with *DCLRE1C* variants were identified in this study. WES revealed a homozygous deletion in the *DCLRE1C* gene from exons 1–5 in patient 1, exons 1–3 deletion and a novel rare variant (c.92T>C, p.L31P) in patient 2, exons 1–3 deletion and a novel rare variant (c.328C>G, p.L110V) in patient 3, and exons 1–4 deletion and a novel frameshift variant (c.449dup, p.His151Alafs*20) in patient 4. Based on literature review, exons 1–3 was recognized as a hotspot region for deletion variation. Moreover, we found that compound variations (CNV + SNV) accounted for approximately 7% variations in all variants. When patients are screened for T-cell receptor excision circles (TRECs), NGS can be used to expand genetic testing. Deletion of the *DCLRE1C* gene should not be ignored when a variant has been found in patients with SCID.

Keywords: severe combined immunodeficiency, *DCLRE1C* gene, copy number variation, single nucleotide variation, sequencing

OPEN ACCESS

Edited by:
Desheng Liang,
Central South University, China

Reviewed by:
María Eugenia De La Morena-Barrio,
University of Murcia, Spain
Lifeng Tian,
University of Pennsylvania,
United States

****Correspondence:***
Huijun Wang
huijunwang@fudan.edu.cn
Wenhao Zhou
zhouwenhao@fudan.edu.cn

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *08 March 2021*
Accepted: *28 June 2021*
Published: *04 August 2021*

Citation:
Xiao F, Lu Y, Wu B, Liu B, Li G, Zhang P, Zhou Q, Sun J, Wang H and Zhou W (2021) High-Frequency Exon Deletion of DNA Cross-Link Repair 1C Accounting for Severe Combined Immunodeficiency May Be Missed by Whole-Exome Sequencing. Front. Genet. 12:677748. doi: 10.3389/fgene.2021.677748

INTRODUCTION

Severe combined immunodeficiency (SCID), one of the most severe forms of primary immunodeficiency diseases, is characterized by a deficiency of T-cell, B-cell, and sometimes NK-cell responses to infections. The reported incidence of SCID ranges from one per 50,000 to one per 1,00,000 live births (Lipstein et al., 2010). The affected patients fail to clear infections and usually die early in life, even with treatment. Studies have shown that genetic abnormalities are associated with the development of SCID. To date, more than 40 genes have been reported to be associated with SCID.

The DNA cross-link repair 1C (*DCLRE1C*) gene, also known as *ARTEMIS*, is located on chromosome 10p13. It encodes the nuclease ARTEMIS, a protein with 5'–3' exonuclease activity for single-stranded DNA. ARTEMIS plays an important role in V(D)J recombination that occurs during B- and T-cell development. Variations in the *DCLRE1C* gene is associated with autosomal recessive SCID by affecting the V(D)J recombination. In 2001, Moshous et al. (2001) reported 13 patients with SCID who carried *DCLRE1C* gene point variants/exon deletions. Since then, an increasing number of patients have been reported, some of which were found during newborn screening (Rechavi et al., 2017). Notably, some compound variations [copy number variation (CNV) + single nucleotide variation (SNV)] in the *DCLRE1C* were identified by Sanger sequencing or PCR (Moshous et al., 2001, 2003). However, these traditional methods (Sanger or PCR) can only sequence short DNA fragments. The development of next-generation sequencing (NGS) has accelerated genomics studies. NGS can simultaneously sequence more than 100 genes and has a shorter turnaround time. To date, several studies (Luk et al., 2017; Sundin et al., 2018; Strand et al., 2020) have reported the use of NGS for the detection of SCID patients, and some patients with *DCLRE1C* variants were identified. However, the detection of CNV by NGS remains a challenge because of the issues intrinsic to the technology (such as short read lengths, etc.; Moreno-Cabrera et al., 2020). Thus, some deletions in *DCLRE1C* may be missed by NGS.

Here, we present four cases of SCID caused by *DCLRE1C* variants. All patients were diagnosed with SCID by a physician (two of them were identified during newborn screening). The genetic information of SCID patients carrying the *DCLRE1C* variants was collected from reported studies. The purpose of this study was to expand knowledge related to the genetic information of SCID patients carrying the *DCLRE1C* variant.

MATERIALS AND METHODS

Patients

Four patients were included in the present study. Informed consent was obtained from the parents or guardians of the patients. The study was approved by the Ethics Committee of the Children's Hospital of Fudan University.

Whole-Exome Sequencing

At least 2 ml peripheral blood was obtained from the patients and their parents into tubes with EDTA. Genomic DNA was extracted using a TIANGEN DNA Blood Mini Kit. Subsequently, the DNA fragments were enriched using the Agilent SureSelect XT Human All Exon V5 kit. Sequencing was performed on an Illumina HiSeq X10 platform. The average on-target sequencing depth was 120X. Burrows-Wheeler Aligner was used to align the sequencing reads to the reference genome hg19 (Li and Durbin, 2009). The variants were called based on the genome analysis toolkit Best Practices, and a variant call format file was generated. The detailed methods for the data analysis are present in our published study (Yang et al., 2019). For CNV analysis, pipeline for clinical NGS-involved CNV detection (PICNIC) was used to detect CNV from whole-exome sequencing (WES) data. The PICNIC can filter out high-frequency gene deletions/duplications (which occurred in >10% of the internal samples). In addition, we used Database of Genomic Variants, the Database of Chromosomal Imbalance and Phenotype in Humans using Ensemble Resources, and our internal database (Wang et al., 2014) for region-level annotation. More information about CNV detection can be seen in our previous study (Dong et al., 2020).

Quantitative PCR to Detect DNA

Deletions in the *DCLRE1C* were confirmed using quantitative PCR (qPCR). The primers were designed using Primer Premier 5.0. The details on the primers are shown in **Supplementary Table S1**. The qPCR mixture (10 μl) contained 5.0 μl SYBR Green mix (2X), 0.2 μl ROX Reference Dye II (50X), 0.5 μmol/L of each primer for the target region and for lactate dehydrogenase A as the reference gene, 20 μg DNA, and ddH_2O. The following conditions were used for qPCR: an initial 45-s 95°C period followed by 40 cycles of amplification (5 s at 95°C, 30 s at 60°C, and 30 s at 72°C) using the real-time PCR system (Applied Biosystems StepOnePlus). Relative quantification was performed using the $2^{-\Delta\Delta Ct}$ method. The copy numbers that were less than 0.8, between 0.8 and 1.4, and more than 1.4 mean deletion, normal, and duplication, respectively.

Sanger Sequencing

Sanger sequencing was performed to validate the identified variants. Sequencing was conducted using an automated sequencer (3500XL Genetic Analyzer, Applied Biosystems, United States). The primers were designed using Primer Premier 5.0. The Mutation Surveyor (SoftGenetics®, State College, PA, United States) was used for sequence analysis.

RESULTS

Clinical Characteristics

All the clinical features of our patients are listed in **Table 1**.

Patient 1 (P1) had severe pneumonia at 3 months of age. Her sputum test revealed an infection with *Pneumocystis carinii* and *Candida albicans*. Flow cytometry revealed a decrease in

TABLE 1 | Clinical characteristics of the included four patients.

	Patient 1	Patient 2	Patient 3	Patient 4
Clinical characteristic				
Age	3 months	2 months	3 months	2 months
Gender	Female	Male	Female	Female
Birth weight	NA	3,000 g	2,750 g	3,350 g
Symptoms	Severe pneumonia	Fever, infection	Acute upper respiratory infection, septicemia	Fever, cough
Clinical diagnosis	SCID	SCID	SCID	SCID
Clinical immuno-function test				
$CD3^+$% (64–73%)	Decreased	52.06%	54.46%	39.24%
$CD4^+$% (29–36%)	NA	25.46%	41.14%	36.16%
$CD8^+$% (24–34%)	NA	26.63%	7.97%	2.97%
$CD19^+$% (14–21%)	Decreased	0.19%	17.95%	0
$CD16^+CD56$% (11–23%)	NA	46.71%	22.59%	56.95%
IgG (3.7–8.3 g/L)	NA	6.1	2.2	7.6
IgA (0.14–0.5 g/L)	NA	2.66	0.08	0.04
IgM (0.33–1.25 g/L)	NA	0.16	0.74	0.04
TREC	Abnormal	NA	NA	Abnormal
Genotype features				
Region of deletion	Exons 1–5	Exons 1–3	Exons 1–3	Exons 1–4
Zygote of deletion	Homozygous	Heterozygous	Heterozygous	Heterozygous
Mutations	/	Exon1: c.92T>C (p.L31P)	Exon5: c.328C>G (p.L110V)	Exon6: c.449_450insT
Zygote of mutation	/	Heterozygous	Heterozygous	Heterozygous
Clinical outcome				
Prognosis	Died	Refused HSCT	Refused HSCT	HSCT
Last follow-up	Died at 4 months old	10 months and 14 days	1 year and 7 months	1 year and 3 months

NA means not available.

the expression of CD3 and CD19. She died at 4 months of age after unclear treatment at a local hospital. Blood samples were sent to the laboratory for genetic testing.

Patient 2 (P2) developed eczema all over the body without any known cause at 2 months of age. At 4 months of age, the patient was diagnosed with pneumonia due to cough and a recurrent fever. Subsequently, he was diagnosed with *Acinetobacter baumannii* infection and had a recurrent fever, which was identified at a local hospital. He was admitted to our hospital for further treatment at 6 months and 15 days of age. However, the family refused hemopoietic stem cell transplantation at 10 months and 14 days of age.

Patient 3 (P3) presented with fever and vomiting at 4 months of age. She was diagnosed with acute upper respiratory infection, septicemia, and low white blood cell count at a local hospital. She was treated with antibiotics, but the patient frequently had a fever. She visited our hospital for further treatment at 9 months and 3 days of age. However, her parents refused hematopoietic stem cell transplantation, and there was no further follow-up after 1 year and 10 months of age.

Patient 4 (P4) started coughing without a known reason at 2 months of age. Her parents used Chinese herbs, but her symptoms did not improve. At 4 months of age, she had a fever, which increased up to 39.3°C. In addition, an approximately 4-cm mass was found in the left axilla. She was diagnosed with immunodeficiency and tuberculous infection. Her genetic test at the local hospital showed no variation in *RAG1* and *RAG2*. The patient was subsequently referred to our hospital for further treatment at the age of 6 months. She underwent hematopoietic stem cell transplantation at the age of 7 months. Follow-up at 15 months showed normal growth and development.

Genetic Testing Results

Whole-exome sequencing was performed on peripheral blood sample obtained from the patients and their biological parents. In this study, both the CNV and SNV results were generated. Genetic testing revealed a homozygous deletion of the *DCLRE1C* gene from exons 1–5 in P1 (**Figure 1A**). P2 had exons 1–3 deletion and a mosaic novel rare variant (c.92T>C, p.L31P; **Figure 1B**); the short tandem repeat test showed that maternal cells accounted for 31.36% in patient 2. P3 had a compound heterozygous exons 1–3 deletion and a novel rare variant (c.328C>G, p.L110V; **Figure 1C**). P4 had a heterozygous exons 1–4 deletion and a novel frameshift variant (c.449dup, p.His151Alafs*20; **Figure 1D**). All variants and deletions of each patient were confirmed by Sanger sequencing and qPCR, respectively (**Figure 1**). **Supplementary Table S2** shows the prediction of pathogenicity for novel variants.

Genetic Features in Published Studies

Combining Human Gene Mutation Database (HGMD) and PubMed, a total of 26 publications that reported SCID related to *DCLRE1C* until March 2021 were included for analysis. The genetic characteristics of patients from published studies (Moshous et al., 2001, 2003; Li et al., 2002; Kobayashi et al., 2003; Noordzij et al., 2003; Musio et al., 2005; Evans et al., 2006; Darroudi et al., 2007; van der Burg et al., 2007; Lagresle-Peyrou et al., 2008; Alsmadi et al., 2009; Pannicke et al., 2010; Woodbine et al., 2010; Ijspeert et al., 2011; Lee et al., 2011, 2013; Tomashov-Matar et al., 2012; Bajin et al., 2013; Halbrich et al., 2013; Schuetz et al., 2014; Felgentreff et al., 2015; Lobachevsky et al., 2015; Volk et al., 2015; Al-Mousa et al., 2016; Dvorak et al., 2017; Luk et al., 2017; Rechavi et al., 2017; Stray-Pedersen et al., 2017; Tahiat et al., 2017; Al-Herz et al., 2018; Sundin et al., 2018, 2019; Krantz et al., 2019; Wu et al., 2019; Dasouki et al., 2020; Fayez et al., 2020; Firtina et al., 2020; Kalina et al., 2020; Simon et al., 2020; Strand et al., 2020; Vignesh et al., 2020) are listed in **Supplementary Table S3**. In total, 87 variants were recorded in HGMD as SCID. A combined literature review was performed on studies published on PubMed, and 153 patients were reported with the *DCLRE1C* variants. A total of 116 patients had detailed

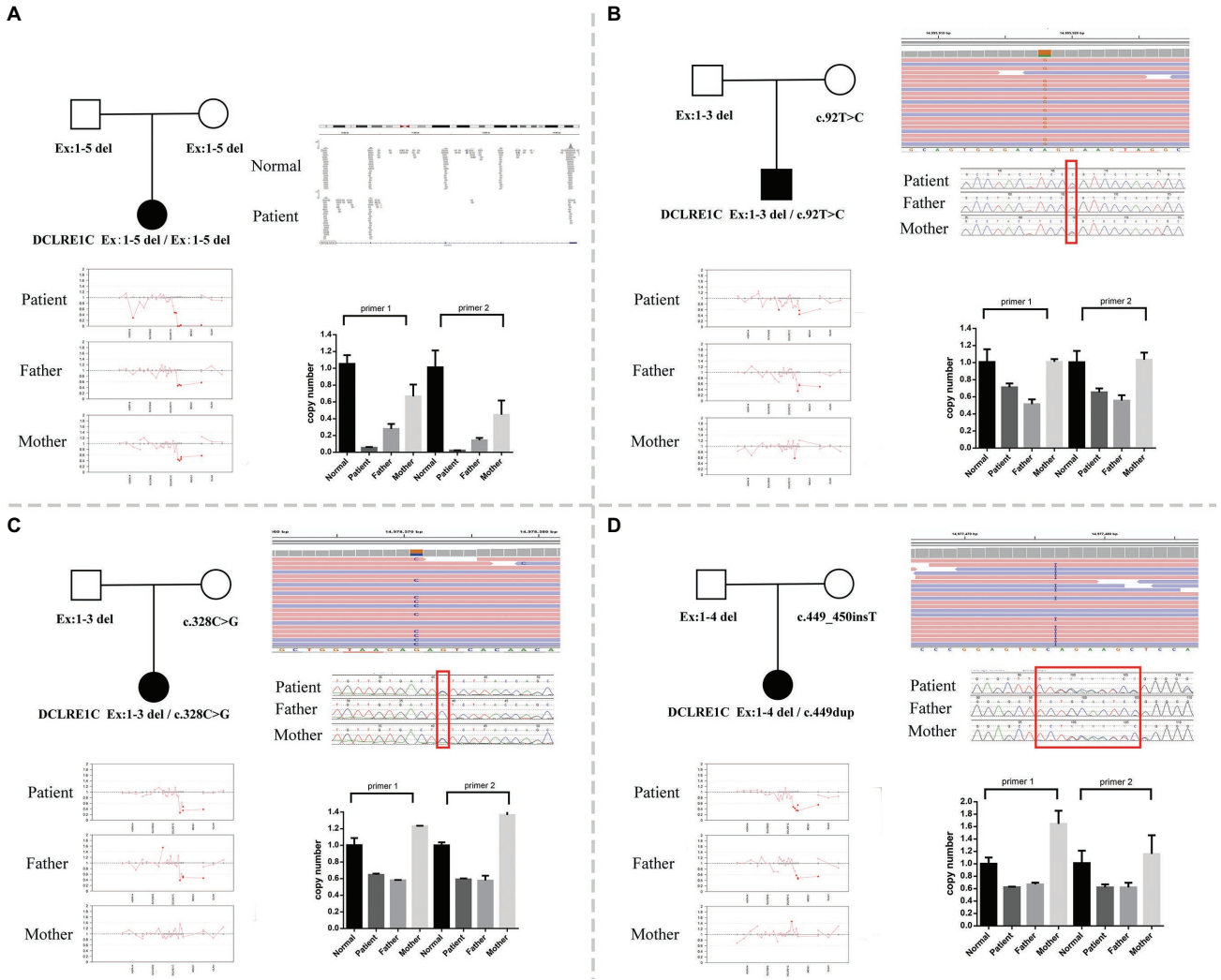


FIGURE 1 | (A) A family pedigree of patient 1. Whole-exome sequencing (WES) revealed a homozygous deletion of DNA cross-link repair 1C (*DCLRE1C*) gene from exons 1–5 in this patient, and quantitative PCR (qPCR) verified it. **(B)** A family pedigree of patient 2. WES revealed exons 1–3 deletion and a novel rare variant (c.92T>C, p.L31P). The qPCR verified exons 1–3 deletion, and Sanger sequencing confirmed the variant (c.92T>C, p.L31P). **(C)** A family pedigree of patient 3. WES revealed heterozygous exons 1–3 deletion and a novel rare variant (c.328C>G, p.L110V). The qPCR verified exons 1–3 deletion, and Sanger sequencing confirmed the variant (c.328C>G, p.L110V). **(D)** A family pedigree of patient 4. WES revealed heterozygous exons 1–4 deletion and a novel frameshift variant (c.449_450insT). The qPCR verified exons 1–4 deletion, and Sanger sequencing confirmed the variant (c.449dup, p.His151Alafs*20).

zygosity information. Of them, 14 patients were compound heterozygous, 10 patients were heterozygous, and 92 patients were homozygous. In total, 37 studies indicated the test methods. Of these studies, 18 studies used PCR, 15 studies (including eight WES) used NGS, two studies used Sanger sequencing, one study used a genome-wide scan, and one study used *multiplex ligation-dependent probe amplification*. We observed that the exon deletions were approximately one-third (55/153) of the total variations in SCID patients carrying *DCLRE1C* variants. Among these deletions, more than 50% (33/55) were exons 1–3 deletion; exons 1–3 was recognized as a hotspot region for deletion variation. Moreover, compound variations (CNV + SNV) accounted for approximately 7% (11/153) of variations in all variants.

DISCUSSION

In this study, four SCID patients with deletions in exons 1–5 were identified. P1 had homozygous exon deletions, while P2, P3, and P4 carried compound heterozygous variants (CNV + SNV). In all the patients, these deletions were inherited from the father, while the point variants were inherited from the mother. We propose the following as a possible mechanism for the observations made in this study. P1 had homozygous exons 1–5 deletion, which is caused by homologous recombination and results in SCID. P2 and P3 carried compound heterozygous variants with exons 1–3 deletion and a missense variant. Exons 1–3 deletion can account for the absence of ARTEMIS protein expression due to homology and should be considered a null

allele (Pannicke et al., 2010). These two novel missense variants (p.L31P and p.L110V) are located on exons 1 and 5, respectively. Exons 1 and 2 are important for ARTEMIS endonucleolytic features (Pannicke et al., 2004). Exons 5 and 6 are necessary for maintaining the ARTEMIS structure (Pannicke et al., 2004). Since exons 1–3 deletion is a null allele, p.L31P probably influences the active site of ARTEMIS, and p.L110V likely affects the protein structure, thereby leading to SCID. P4 had a heterozygous exons 1–4 deletion along with a frameshift variant (c.449dup) in exons 6. Felgentreff et al. (2015) verified that a frameshift variant can disrupt domains and almost completely abrogate the ARTEMIS function. We speculated that this frameshift variant probably affects the balance of the ARTEMIS structure and results in SCID.

Our literature review analyzed 26 publications and summarized that exons 1–3 is a hotspot region for deletion. The major mechanism for this hotspot deletion is homologous recombination (Meek et al., 2004). Moreover, Pannicke et al. (2010) found that there was a series of short interspersed nuclear element, long interspersed nuclear element, long terminal repeat, simple repeat, and low complexity repeat sequences in this region that can enhance the recombination events. Although the above-mentioned mechanisms have been identified, more evidence from functional studies are needed.

In addition, our literature review found that compound variations (CNV + SNV) accounted for approximately 7% of the total variants, and deletion in *DCLRE1C* gene was mainly tested by PCR. Currently, NGS is a more powerful tool for obtaining genetic test results than PCR. However, small CNV calling remains challenging due to variable coverage and data analysis pipelines (Teo et al., 2012). WES analysis pipelines still leave a gap of small deletion calling between SNVs/small insertions/deletions and CNVs larger than 1 Mb. In recent years, the development of long-read sequencing, directly sequencing single molecules of DNA in real time, can help for the detection of complex chromosomal rearrangements including deletions, inversions, insertions, and duplications (McGinty et al., 2017). However, long-read sequencing has several limitations, such as high sequencing error rate, high systematic error, and higher cost than NGS (Xiao and Zhou, 2020).

Traditional exome sequencing mainly concentrates on the detection of SNVs. The small deletion may be missed by these methods. Subsequently, several algorithms, including CANOES (Backenroth et al., 2014), XHMM (Fromer et al., 2012), and Co NIFER (Krumm et al., 2012), have been developed for CNV analysis, and recent studies (Yang et al., 2013; Retterer et al., 2015; Ellingford et al., 2018) showed the good performance of CNV analysis through WES. In this study, we used PICNIC (Qin et al., 2018), which demonstrated 100% specificity and sensitivity to pathogenic/likely pathogenic CNVs, for the detection of CNVs. Totally, four cases were diagnosed with SCID, with three compounds heterozygous and one homozygous deletion of the *DCLRE1C* gene, which was detected by WES. Without PICNIC, these cases may have been left undiagnosed if the deletion calling was missed. In addition, qPCR can be further used for deletion validation due to the limitations of NGS. Notably, healthy individuals can also harbor variations in the *DCLRE1C* gene. Thus, healthcare professionals should consider the possibility of deletion when a variant is found in the *DCLRE1C* gene.

Currently, newborns are being screened for SCID in some countries by assessing T-cell receptor excision circles (TREC). Krantz et al. (2019) reported two familial patients diagnosed with SCID who carried the same compound *DCLRE1C* deletions. However, the outcomes for these patients were different: the one who underwent TREC screening and early genetic diagnosis survived with SCID prior to the onset of infections. Although TREC is useful for SCID screening, SCID diagnosis should be confirmed by genetic testing. Although only 5% of SCID patients carry a variant of *DCLRE1C*, these patients (87.5%, 7/8) had 10-year survival rate higher than that of *RAG1/2* SCID patients (64.4%, 9/13) after hematopoietic stem cell transplantation (Hamid et al., 2018). Therefore, the early diagnosis of *DCLRE1C* SCID may provide patients with effective treatment and a better chance of survival.

In conclusion, we reported four SCID cases with five allelic deletions in exons 1–5. When patients are screened for the presence of TREC to diagnose SCID, NGS can be used for further genetic testing. Gross deletions in the *DCLRE1C* should not be ignored when a variant has been identified.

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 authors.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

HW, FX, and WZ conceptualized the study and analyzed the data. YL, BW, BL, GL, PZ, QZ, and JS co-conceptualized the study and interpreted the data. All authors contributed to the drafting of the article and revised it critically for important intellectual content. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.677748/full#supplementary-material

REFERENCES

Al-Herz, W., Massaad, M. J., Chou, J., Notarangelo, L. D., and Geha, R. S. (2018). DNA recombination defects in Kuwait: clinical, immunologic and genetic profile. *Clin. Immunol.* 187, 68–75. doi: 10.1016/j.clim.2017.10.006

Al-Mousa, H., Abouelhoda, M., Monies, D. M., Al-Tassan, N., Al-Ghonaium, A., Al-Saud, B., et al. (2016). Unbiased targeted next-generation sequencing molecular approach for primary immunodeficiency diseases. *J. Allergy Clin. Immunol.* 137, 1780–1787. doi: 10.1016/j.jaci.2015.12.1310

Alsmadi, O., Al-Ghonaium, A., Al-Muhsen, S., Arnaout, R., Al-Dhekri, H., Al-Saud, B., et al. (2009). Molecular analysis of $T^-B^-NK^+$ severe combined immunodeficiency and Omenn syndrome cases in Saudi Arabia. *BMC Med. Genet.* 10:116. doi: 10.1186/1471-2350-10-116

Backenroth, D., Homsy, J., Murillo, L. R., Glessner, J., Lin, E., Brueckner, M., et al. (2014). CANOES: detecting rare copy number variants from whole exome sequencing data. *Nucleic Acids Res.* 42:e97. doi: 10.1093/nar/gku345

Bajin, İ. Y., Ayvaz, D., Ünal, S., Özgür, T. T., Çetin, M., Gümrük, F., et al. (2013). Atypical combined immunodeficiency due to Artemis defect: a case presenting as hyperimmunoglobulin M syndrome and with LGLL. *Mol. Immunol.* 56, 354–357. doi: 10.1016/j.molimm.2013.05.004

Darroudi, F., Wiegant, W., Meijers, M., Friedl, A. A., van der Burg, M., Fomina, J., et al. (2007). Role of Artemis in DSB repair and guarding chromosomal stability following exposure to ionizing radiation at different stages of cell cycle. *Mutat. Res.* 615, 111–124. doi: 10.1016/j.mrfmmm.2006.11.029

Dasouki, M., Jabr, A., AlDakheel, G., Elbadaoui, F., Alazami, A. M., Al-Saud, B., et al. (2020). TREC and KREC profiling as a representative of thymus and bone marrow output in patients with various inborn errors of immunity. *Clin. Exp. Immunol.* 202, 60–71. doi: 10.1111/cei.13484

Dong, X., Liu, B., Yang, L., Wang, H., Wu, B., Liu, R., et al. (2020). Clinical exome sequencing as the first-tier test for diagnosing developmental disorders covering both CNV and SNV: a Chinese cohort. *J. Med. Genet.* 57, 558–566. doi: 10.1136/jmedgenet-2019-106377

Dvorak, C. C., Patel, K., Puck, J. M., Wahlstrom, J., Dorsey, M. J., Adams, R., et al. (2017). Unconditioned unrelated donor bone marrow transplantation for IL7Rα- and Artemis-deficient SCID. *Bone Marrow Transplant.* 52, 1036–1038. doi: 10.1038/bmt.2017.74

Ellingford, J. M., Horn, B., Campbell, C., Arno, G., Barton, S., Tate, C., et al. (2018). Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *J. Med. Genet.* 55, 114–121. doi: 10.1136/jmedgenet-2017-104791

Evans, P. M., Woodbine, L., Riballo, E., Gennery, A. R., Hubank, M., and Jeggo, P. A. (2006). Radiation-induced delayed cell death in a hypomorphic Artemis cell line. *Hum. Mol. Genet.* 15, 1303–1311. doi: 10.1093/hmg/ddl050

Fayez, E. A., Qazvini, F. F., Mahmoudi, S. M., Khoei, S., Vesaltalab, M., and Teimourian, S. (2020). Diagnosis of radiosensitive severe combined immunodeficiency disease (RS-SCID) by comet assay, management of bone marrow transplantation. *Immunobiology* 225:151961. doi: 10.1016/j.imbio.2020.151961

Felgentreff, K., Lee, Y. N., Frugoni, F., Du, L., van der Burg, M., Giliani, S., et al. (2015). Functional analysis of naturally occurring DCLRE1C mutations and correlation with the clinical phenotype of ARTEMIS deficiency. *J. Allergy Clin. Immunol.* 136, 140.e7–150.e7. doi: 10.1016/j.jaci.2015.03.005

Firtina, S., Ng, Y. Y., Ng, O. H., Kiykim, A., Aydiner, E., Nepesov, S., et al. (2020). Mutational landscape of severe combined immunodeficiency patients from Turkey. *Int. J. Immunogenet.* 47, 529–538. doi: 10.1111/iji.12496

Fromer, M., Moran, J. L., Chambert, K., Banks, E., Bergen, S. E., Ruderfer, D. M., et al. (2012). Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. *Am. J. Hum. Genet.* 91, 597–607. doi: 10.1016/j.ajhg.2012.08.005

Halbrich, M., Ben-Shoshan, M., and McCusker, C. (2013). Autoimmune hemolytic anemia in a teenager: a wolf in sheep's clothing. *Eur. J. Haematol.* 91, 262–264. doi: 10.1111/ejh.12148

Hamid, I. J. A., Slatter, M. A., McKendrick, F., Pearce, M. S., and Gennery, A. R. (2018). Long-term health outcome and quality of life post-HSCT for IL7Ralpha-, Artemis-, RAG1- and RAG2-deficient severe combined immunodeficiency: a single center report. *J. Clin. Immunol.* 38, 727–732. doi: 10.1007/s10875-018-0540-9

Ijspeert, H., Lankester, A. C., van den Berg, J. M., Wiegant, W., van Zelm, M. C., Weemaes, C. M., et al. (2011). Artemis splice defects cause atypical SCID and can be restored in vitro by an antisense oligonucleotide. *Genes Immun.* 12, 434–444. doi: 10.1038/gene.2011.16

Kalina, T., Bakardjieva, M., Blom, M., Perez-Andres, M., Barendregt, B., Kanderová, V., et al. (2020). Euroflow standardized approach to diagnostic immunopheneotyping of severe PID in newborns and young children. *Front. Immunol.* 11:371. doi: 10.3389/fimmu.2020.00371

Kobayashi, N., Agematsu, K., Sugita, K., Sako, M., Nonoyama, S., Yachie, A., et al. (2003). Novel Artemis gene mutations of radiosensitive severe combined immunodeficiency in Japanese families. *Hum. Genet.* 112, 348–352. doi: 10.1007/s00439-002-0897-x

Krantz, M. S., Stone, C. A., Connelly, J. A., Norton, A. E., and Khan, Y. W. (2019). The effect of delayed and early diagnosis in siblings, and importance of newborn screening for SCID. *Ann. Allergy Asthma Immunol.* 122, 211–213. doi: 10.1016/j.anai.2018.11.002

Krumm, N., Sudmant, P. H., Ko, A., O'Roak, B. J., Malig, M., Coe, B. P., et al. (2012). Copy number variation detection and genotyping from exome sequence data. *Genome Res.* 22, 1525–1532. doi: 10.1101/gr.138115.112

Lagresle-Peyrou, C., Benjelloun, F., Hue, C., Andre-Schmutz, I., Bonhomme, D., Forveille, M., et al. (2008). Restoration of human B-cell differentiation into NOD-SCID mice engrafted with gene-corrected $CD34^+$ cells isolated from Artemis or RAG1-deficient patients. *Mol. Ther.* 16, 396–403. doi: 10.1038/sj.mt.6300353

Lee, P. P., Chan, K. W., Chen, T. X., Jiang, L. P., Wang, X. C., Zeng, H. S., et al. (2011). Molecular diagnosis of severe combined immunodeficiency--identification of IL2RG, JAK3, IL7R, DCLRE1C, RAG1, and RAG2 mutations in a cohort of Chinese and southeast Asian children. *J. Clin. Immunol.* 31, 281–296. doi: 10.1007/s10875-010-9489-z

Lee, P. P., Woodbine, L., Gilmour, K. C., Bibi, S., Cale, C. M., Amrolia, P. J., et al. (2013). The many faces of Artemis-deficient combined immunodeficiency—two patients with DCLRE1C mutations and a systematic literature review of genotype-phenotype correlation. *Clin. Immunol.* 149, 464–474. doi: 10.1016/j.clim.2013.08.006

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, L., Moshous, D., Zhou, Y., Wang, J., Xie, G., Salido, E., et al. (2002). A founder mutation in Artemis, an SNM1-like protein, causes SCID in Athabascan-speaking native Americans. *J. Immunol.* 168, 6323–6329. doi: 10.4049/jimmunol.168.12.6323

Lipstein, E. A., Vorono, S., Browning, M. F., Green, N. S., Kemper, A. R., Knapp, A. A., et al. (2010). Systematic evidence review of newborn screening and treatment of severe combined immunodeficiency. *Pediatrics* 125, e1226–e1235. doi: 10.1542/peds.2009-1567

Lobachevsky, P., Woodbine, L., Hsiao, K. C., Choo, S., Fraser, C., Gray, P., et al. (2015). Evaluation of severe combined immunodeficiency and combined immunodeficiency pediatric patients on the basis of cellular radiosensitivity. *J. Mol. Diagn.* 17, 560–575. doi: 10.1016/j.jmoldx.2015.05.004

Luk, A. D. W., Lee, P. P., Mao, H., Chan, K. W., Chen, X. Y., Chen, T. X., et al. (2017). Family history of early infant death correlates with earlier age at diagnosis but not shorter time to diagnosis for severe combined immunodeficiency. *Front. Immunol.* 8:808. doi: 10.3389/fimmu.2017.00808

McGinty, R. J., Rubinstein, R. G., Neil, A. J., Dominska, M., Kiktev, D., Petes, T. D., et al. (2017). Nanopore sequencing of complex genomic rearrangements in yeast reveals mechanisms of repeat-mediated double-strand break repair. *Genome Res.* 27, 2072–2082. doi: 10.1101/gr.228148.117

Meek, K., Gupta, S., Ramsden, D. A., and Lees-Miller, S. P. (2004). The DNA-dependent protein kinase: the director at the end. *Immunol. Rev.* 200, 132–141. doi: 10.1111/j.0105-2896.2004.00162.x

Moreno-Cabrera, J. M., Del Valle, J., Castellanos, E., Feliubadaló, L., Pineda, M., Brunet, J., et al. (2020). Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. *Eur. J. Hum. Genet.* 28, 1645–1655. doi: 10.1038/s41431-020-0675-z

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., et al. (2001). Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 105, 177–186. doi: 10.1016/S0092-8674(01)00309-9

Moshous, D., Pannetier, C., de Chasseval Rd, R., le Deist Fl, F., Cavazzana-Calvo, M., Romana, S., et al. (2003). Partial T and B lymphocyte immunodeficiency

and predisposition to lymphoma in patients with hypomorphic mutations in Artemis. *J. Clin. Invest.* 111, 381–387. doi: 10.1172/JCI16774

Musio, A., Marrella, V., Sobacchi, C., Rucci, F., Fariselli, L., Giliani, S., et al. (2005). Damaging-agent sensitivity of Artemis-deficient cell lines. *Eur. J. Immunol.* 35, 1250–1256. doi: 10.1002/eji.200425555

Noordzij, J. G., Verkaik, N. S., van der Burg, M., van Veelen, L. R., de Bruin-Versteeg, S., Wiegant, W., et al. (2003). Radiosensitive SCID patients with Artemis gene mutations show a complete B-cell differentiation arrest at the pre-B-cell receptor checkpoint in bone marrow. *Blood* 101, 1446–1452. doi: 10.1182/blood-2002-01-0187

Pannicke, U., Hönig, M., Schulze, I., Rohr, J., Heinz, G. A., Braun, S., et al. (2010). The most frequent DCLRE1C (ARTEMIS) mutations are based on homologous recombination events. *Hum. Mutat.* 31, 197–207. doi: 10.1002/humu.21168

Pannicke, U., Ma, Y., Hopfner, K. P., Niewolik, D., Lieber, M. R., and Schwarz, K. (2004). Functional and biochemical dissection of the structure-specific nuclease ARTEMIS. *EMBO J.* 23, 1987–1997. doi: 10.1038/sj.emboj.7600206

Qin, Q., Liu, B., Yang, L., Wu, B., Wang, H., Dong, X., et al. (2018). Application of copy number variation screening analysis process based on high-throughput sequencing technology. *Chin. J. Evid. Based Pediatr.* 13, 275–279. doi: 10.3969/j.issn.1673-5501.2018.04.007

Rechavi, E., Lev, A., Simon, A. J., Stauber, T., Daas, S., Saraf-Levy, T., et al. (2017). First year of Israeli newborn screening for severe combined immunodeficiency-clinical achievements and insights. *Front. Immunol.* 8:1448. doi: 10.3389/fimmu.2017.01448

Retterer, K., Scuffins, J., Schmidt, D., Lewis, R., Pineda-Alvarez, D., Stafford, A., et al. (2015). Assessing copy number from exome sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. *Genet. Med.* 17, 623–629. doi: 10.1038/gim.2014.160

Schuetz, C., Neven, B., Dvorak, C. C., Leroy, S., Ege, M. J., Pannicke, U., et al. (2014). SCID patients with Artemis vs. RAG deficiencies following HCT: increased risk of late toxicity in Artemis-deficient SCID. *Blood* 123, 281–289. doi: 10.1182/blood-2013-01-476432

Simon, A. J., Golan, A. C., Lev, A., Stauber, T., Barel, O., Somekh, I., et al. (2020). Whole exome sequencing (WES) approach for diagnosing primary immunodeficiencies (PIDs) in a highly consanguineous community. *Clin. Immunol.* 214:108376. doi: 10.1016/j.clim.2020.108376

Strand, J., Gul, K. A., Erichsen, H. C., Lundman, E., Berge, M. C., Trømborg, A. K., et al. (2020). Second-tier next generation sequencing integrated in nationwide newborn screening provides rapid molecular diagnostics of severe combined immunodeficiency. *Front. Immunol.* 11:1417. doi: 10.3389/fimmu.2020.01417

Stray-Pedersen, A., Sorte, H. S., Samarakoon, P., Gambin, T., Chinn, I. K., Akdemir, Z. H. C., et al. (2017). Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. *J. Allergy Clin. Immunol.* 139, 232–245. doi: 10.1016/j.jaci.2016.05.042

Sundin, M., Marits, P., Ramme, K., Kolios, A. G. A., and Nilsson, J. (2019). Severe combined immunodeficiency (SCID) presenting in childhood, with agammaglobulinemia, associated with novel compound heterozygous mutations in DCLRE1C. *Clin. Immunol.* 200, 16–18. doi: 10.1016/j.clim.2018.12.019

Sundin, M., Uhlin, M., Gaballa, A., Ramme, K., Kolios, A. G., Marits, P., et al. (2018). Late presenting atypical severe combined immunodeficiency (SCID) associated with a novel missense mutation in DCLRE1C. *Pediatr. Allergy Immunol.* 29, 108–111. doi: 10.1111/pai.12812

Tahiat, A., Badran, Y. R., Chou, J., Cangemi, B., Lefranc, G., Labgaa, Z. M., et al. (2017). Epidermodysplasia verruciformis as a manifestation of ARTEMIS deficiency in a young adult. *J. Allergy Clin. Immunol.* 139, 372.e4–375.e4. doi: 10.1016/j.jaci.2016.07.024

Teo, S. M., Pawitan, Y., Ku, C. S., Chia, K. S., and Salim, A. (2012). Statistical challenges associated with detecting copy number variations with next-generation sequencing. *Bioinformatics* 28, 2711–2718. doi: 10.1093/bioinformatics/bts535

Tomashov-Matar, R., Biran, G., Lagovsky, I., Kotler, N., Stein, A., Fisch, B., et al. (2012). Severe combined immunodeficiency (SCID): from the detection of a new mutation to preimplantation genetic diagnosis. *J. Assist. Reprod. Genet.* 29, 687–692. doi: 10.1007/s10815-012-9765-3

van der Burg, M., Verkaik, N. S., den Dekker, A. T., Barendregt, B. H., Pico-Knijnenburg, I., Tezcan, I., et al. (2007). Defective Artemis nuclease is characterized by coding joints with microhomology in long palindromic-nucleotide stretches. *Eur. J. Immunol.* 37, 3522–3528. doi: 10.1002/eji.200737624

Vignesh, P., Rawat, A., Kumrah, R., Singh, A., Gummadi, A., Sharma, M., et al. (2020). Clinical, immunological, and molecular features of severe combined immune deficiency: a multi-institutional experience from India. *Front. Immunol.* 11:619146. doi: 10.3389/fimmu.2020.619146

Volk, T., Pannicke, U., Reisli, I., Bulashevska, A., Ritter, J., Björkman, A., et al. (2015). DCLRE1C (ARTEMIS) mutations causing phenotypes ranging from atypical severe combined immunodeficiency to mere antibody deficiency. *Hum. Mol. Genet.* 24, 7361–7372. doi: 10.1093/hmg/ddv437

Wang, H., Bi, W., Wu, B., Zhou, W., and Cheung, S. (2014). Application and explanation of chromosome microarray (CMA) analysis in the diagnosis of clinical genetic disease. *Chin. J. Evid. Based Pediatr.* 9, 227–235. doi: 10.3969/j.issn.1673-5501.2014.03.013

Woodbine, L., Grigoriadou, S., Goodarzi, A. A., Riballo, E., Tape, C., Oliver, A. W., et al. (2010). An Artemis polymorphic variant reduces Artemis activity and confers cellular radiosensitivity. *DNA Repair* 9, 1003–1010. doi: 10.1016/j.dnarep.2010.07.001

Wu, Z., Subramanian, N., Jacobsen, E. M., Sampaio, K. L., van der Merwe, J., Hönig, M., et al. (2019). NK cells from RAG- or DCLRE1C-deficient patients inhibit HCMV. *Microorganisms* 7:546. doi: 10.3390/microorganisms7110546

Xiao, T., and Zhou, W. (2020). The third generation sequencing: the advanced approach to genetic diseases. *Transl. Pediatr.* 9, 163–173. doi: 10.21037/tp.2020.03.06

Yang, L., Kong, Y., Dong, X., Hu, L., Lin, Y., Chen, X., et al. (2019). Clinical and genetic spectrum of a large cohort of children with epilepsy in China. *Genet. Med.* 21, 564–571. doi: 10.1038/s41436-018-0091-8

Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A., et al. (2013). Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N. Engl. J. Med.* 369, 1502–1511. doi: 10.1056/NEJMoa1306555

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 Pediatrics

ORIGINAL RESEARCH
published: 04 October 2021
doi: 10.3389/fped.2021.682160

Single Nucleotide Polymorphisms Interactions of the Surfactant Protein Genes Associated With Respiratory Distress Syndrome Susceptibility in Preterm Infants

Shaili Amatya[1], *Meixia Ye*[2], *Lili Yang*[3], *Chintan K. Gandhi*[1], *Rongling Wu*[4], *Beth Nagourney*[5] *and Joanna Floros*[1,6]*

[1] Department of Pediatrics, Center for Host Defense, Inflammation, and Lung Disease (CHILD) Research, Pennsylvania State University College of Medicine, Hershey, PA, United States, [2] Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [3] School of First Clinical Medicine, Nanjing University of Chinese Medicine, Nanjing, China, [4] Public Health Science, Pennsylvania State University College of Medicine, Hershey, PA, United States, [5] Albert Einstein College of Medicine, New York, NY, United States, [6] Obstetrics and Gynecology, Pennsylvania State University College of Medicine, Hershey, PA, United States

OPEN ACCESS

Edited by:
Maria Elisabetta Baldassarre, University of Bari Aldo Moro, Italy

Reviewed by:
Uday Kishore, Brunel University London, United Kingdom
Ga Won Jeon, Inje University Busan Paik Hospital, South Korea
Anthony George Tsolaki, Brunel University London, United Kingdom

***Correspondence:**
Joanna Floros
jfloros@pennstatehealth.psu.edu

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Pediatrics

Received: *17 March 2021*
Accepted: *06 September 2021*
Published: *04 October 2021*

Citation:
Amatya S, Ye M, Yang L, Gandhi CK, Wu R, Nagourney B and Floros J (2021) Single Nucleotide Polymorphisms Interactions of the Surfactant Protein Genes Associated With Respiratory Distress Syndrome Susceptibility in Preterm Infants. Front. Pediatr. 9:682160. doi: 10.3389/fped.2021.682160

Background: Neonatal respiratory distress syndrome (RDS), due to surfactant deficiency in preterm infants, is the most common cause of respiratory morbidity. The surfactant proteins (*SFTP*) genetic variants have been well-studied in association with RDS; however, the impact of SNP-SNP (single nucleotide polymorphism) interactions on RDS has not been addressed. Therefore, this study utilizes a newer statistical model to determine the association of *SFTP* single SNP model and SNP-SNP interactions in a two and a three SNP interaction model with RDS susceptibility.

Methods: This study used available genotype and clinical data in the Floros biobank at Penn State University. The patients consisted of 848 preterm infants, born <36 weeks of gestation, with 477 infants with RDS and 458 infants without RDS. Seventeen well-studied *SFTPA1*, *SFTPA2*, *SFTPB*, *SFTPC*, and *SFTPD* SNPs were investigated. Wang's statistical model was employed to test and identify significant associations in a case-control study.

Results: Only the rs17886395 (C allele) of the *SFTPA2* was associated with protection for RDS in a single-SNP model (Odd's Ratio 0.16, 95% CI 0.06–0.43, adjusted $p = 0.03$). The highest number of interactions ($n = 27$) in the three SNP interactions were among *SFTPA1* and *SFTPA2*. The three SNP models showed intergenic and intragenic interactions among all *SFTP* SNPs except *SFTPC*.

Conclusion: The single SNP model and SNP interactions using the two and three SNP interactions models identified *SFTP*-SNP associations with RDS. However, the large number of significant associations containing *SFTPA1* and/or *SFTPA2* SNPs point to the importance of *SFTPA1* and *SFTPA2* in RDS susceptibility.

Keywords: epistasis, neonatal, genetic variants, pulmonary, allele

INTRODUCTION

Neonatal respiratory distress syndrome (RDS) is the most common cause of respiratory failure in premature infants due to surfactant deficiency (1). However, the infant mortality rate due to RDS was 11.4 per 100,000 live births and accounted for 2% of all infant deaths in 2017 in the United States (2) despite the judicious use of postnatal surfactant along with antenatal steroids (3).

Major risk factors, such as prematurity and low birth weight (BW) along with sex and race (4–7) have been implicated in RDS. Genetic factors have also been associated with RDS by various twins' studies (8, 9). Thus, the susceptibility to RDS is considered multifactorial and/or polygenic (10), with ample evidence in the literature that gene–host-environment interactions may play a large role in the morbidity and mortality associated with this syndrome. The understanding of gene interactions in RDS may help identify novel therapeutic targets for susceptible infants.

Furthermore, it has been noted that infants dying with RDS have low levels of surfactant proteins (SP) (11, 12). SP-A and SP-D are hydrophilic proteins and play an important role in innate immunity and the regulation of inflammatory processes and host defense (13–17). SP-B and SP-C are hydrophobic proteins that enhance the adsorption and spreading of surfactant phospholipid (18). In addition, SP-B is essential for lung function by reducing surface tension and preventing alveolar collapse (19–21). SP B and SP-C are present in the exogenous surfactant used to treat RDS. However, SP-A and SP-D (SP-D co-isolates with the surfactant complex) are not included in the formulation, even though a major complication in prematurely born infants with RDS is infection. In addition to its host defense function, SP-A, along with SP-B, is important for the formation of tubular myelin (an extracellular surfactant structure) (22–24). Moreover, SP-A is involved in surfactant-related functions (17, 25) and lung airway function (26).

Multiple genetic variants and single nucleotide polymorphisms (SNP) of the surfactant protein gene (*SFTP)* have been shown to associate with RDS (10, 27–40). Human SP-A, consisting of SP-A1 and SP-A2 proteins, is encoded by two functional genes *SFTPA1* and *SFTPA2*, respectively (41). The *SFTPA1* and *SFTPA2* genes share a high degree of sequence similarity but differ at various splice variants at the 5′ untranslated region (UTR) and exhibit sequence variability within coding and non-coding regions (17). Prior studies have also found intragenic and intergenic haplotypes between *SFTPA1* and/or *SFTPA2* (42) and *SFTPB* and/or *SFTPD* haplotypes associated with risk or protective effect in RDS (43).

However, the impact of SNP-SNP interactions on RDS susceptibility has not been addressed before. The synergistic (epistatic) interactions among genetic variants of the surfactant proteins may alter disease susceptibility (44, 45), but this was not possible to study earlier due to the limitation of statistical approaches at the time. However, current more advanced statistical models may help identify the intricate epistatic interaction among multiple gene variants that play a significant role in multifactorial and complex diseases, such as RDS. Such analysis is likely to be beneficial to understand the impact of genetics on complex diseases, especially as we move toward personalized medicine.

In the present study, we studied intergenic and intragenic SNP-SNP interactions of the *SFTP* genes. We hypothesized that epistatic interactions among *SFTP* gene variants are associated with RDS susceptibility in preterm infants.

Abbreviations: SNP, Single nucleotide polymorphism; RDS, respiratory distress syndrome; *SFTP*, surfactant protein gene; BW, birth weight; SP, surfactant protein.

MATERIALS AND METHODS

Study Samples

This study used available genotype data and clinical information in the Floros biobank at Penn State University, College of Medicine. These were collected and processed under an approved protocol by the institutional review board from the human subject protection office of the Pennsylvania State University (PSU) College of Medicine as well as the institutional review board of the respective centers where samples were collected in other Institutions other than PSU, as described previously (12, 29, 31, 32, 46, 47). The clinical and demographic data of the study samples are given in **Table 1**. The patients consisted of 848 preterm infants born <36 weeks of gestation, stratified by RDS, where 458 infants were diagnosed with RDS, and 477 infants did not develop RDS. RDS was diagnosed by clinical features of respiratory distress such as retractions, grunting, and flaring after birth. Chronic lung disease was diagnosed as needing supplemental oxygen at 28 days of life or 36 weeks postmenstrual age (50). Chorioamnionitis was diagnosed by clinical features such as maternal fever. The use of antenatal steroids was variable with betamethasone or dexamethasone.

A total of 17 SNPs of the SP genes *SFTPA1, SFTPA2, SFTPB, SFTPC*, and *SFTPD* were studied. These included five SNPs from *SFTPA1*: rs1059047, rs1136450, rs1136451, rs1059057, and rs4253527; four SNPs from *SFTPA2*: rs1059046, rs17886395, rs1965707, and 1965708; four SNPs from *SFTPB*: rs1130866, rs7316, rs2077079, and rs3024798; two SNPs from *SFTPC*: rs4715 and rs1124; and two SNPs from *SFTPD*: rs721917 and rs2243639. Polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) was used to analyze the *SFTP* gene polymorphisms as described (49, 51, 52).

Statistical Analysis

Wang et al. (53) developed a general multi-locus model for analyzing genetic associations in a case-control study. This model has three characteristics. First, it integrates classic quantitative genetic principles into a categorical data analysis framework, allowing epistatic interactions to be interpreted on a solid genetic basis. Second, this model can not only detect the genetic effects of single SNPs and pairwise genetic interactions, but also characterize high-order genetic interactions. That is, the model dissects genotypic differences into additive (a) and dominant (d) genetic effects at individual SNPs: additive × additive (aa), additive × dominant (ad), dominant × additive (da), and dominant × dominant (dd) epistatic effects at a pair of SNPs, and additive × additive × additive (aaa), additive × additive

× dominant (aad), additive × dominant × additive (ada), additive × dominant × dominant (add), dominant × additive × additive (daa), dominant × additive × dominant (dad), dominant × dominant × additive (dda), dominant × dominant × dominant (ddd) epistatic effects at a triad of SNPs. Mounting evidence shows that high-order interactions play an important role in mediating complex traits and complex human diseases (54). Third, while the precise detection of a pairwise genetic interaction requires a huge number of samples, such as 5,000 (55), which may be hardly met in general studies, Wang et al.'s model is less sample size-reliant by coalescing case and control samples into a 2 × 2 contingency table for the detection of epistasis at any order. The statistical properties of Wang et al.'s model have been extensively studied through computer simulation, with results, presented in the original article, demonstrating its usefulness and robustness in a small-sample case-control study. Also, a detailed computational procedure of this model was given in the original article, allowing the readers to understand and repeat the model.

For each type of data analysis, case-control genotype observations were sorted into a 2 × 2 contingency table to test each of the genetic effects described above. For example, consider a SNP with three genotypes AA, Aa, and aa. To estimate its dominant effect, the effect size was compared to that of the heterozygote Aa against the average size of each of the two homozygotes AA and aa in cases and controls, respectively. Based on the resulting 2 × 2 contingency table, the logistic regression model was implemented to estimate the dominant effect of this SNP, and the effects were adjusted for age and sex. The odds ratio (OR) was estimated to assess the magnitude of the dominant/additive effect.

To estimate the additive effect, the size was compared as below,

Odds of genotype for cases = number of cases with AA/number of cases with aa

Odds of genotype for controls = number of controls with AA/number of controls with aa

$$\text{OR} = \text{odds for cases/odds for controls} = \frac{(\text{number of cases with AA} \times \text{number of controls with aa})}{(\text{number of controls with AA} \times \text{number of cases with aa})}$$

For example-

OR = 1: Genotype difference is not associated with the disease;

OR > 1.0: Genotype AA is "more risky" (i.e., associated with higher risk for the disease than genotype aa)

OR < 1.0: Genotype aa is "more risky" for the disease than genotype AA

A similar procedure was applied to analyze all other genetic effects.

The significance of each effect was adjusted for multiple comparisons using the false discovery rate (FDR) controlled at 1%. Wang et al.'s simulation data indicate that a 100 × 100 sample size combination in an epistatic case-control model has a power of > 0.80 to detect significant associations in a 2 × 2 contingency table analysis (53). Thus, our current sample size provides adequate power to detect all the significant epistatic interactions.

RESULTS

Clinical Characteristics of Infants With and Without RDS

Table 1 shows the demographic and clinical characteristics of infants with and without RDS. There were 458 infants without RDS and 477 infants who developed RDS. Infants with RDS were younger as assessed by gestational age at birth (30 vs. 33 weeks) and had lower birth weight (1,474 ± 606 gram vs. 1,818 ± 515 gram) compared to infants without RDS. Infants with RDS were predominantly male (58 vs. 48%, *p*-value 0.02). The two risk factors for RDS (gestational age and sex) were corrected in the analysis. Gestational age and birth weight are co-linear variables, and only one (gestational age) was chosen to be corrected in the analysis. As expected, infants who developed RDS had increased use of surfactant and a higher incidence of chronic lung disease than infants who did not have RDS. These outcomes are related to RDS rather than predictors (surfactant use and chronic lung disease); therefore, we did not correct them in the SNP-SNP interaction model. The use of antenatal steroids was significantly different between the two groups. However, ∼40% of the antenatal steroid data were missing and may have caused bias in estimating this parameter.

Association of *SFTP* SNP-SNP Interaction With RDS

Description

The associations of single SNP and intergenic/intragenic two and three SNP interactions with RDS are shown in **Tables 2**–**4**, respectively. The tables show the specific SNPs of the *SFTP* genes and their effect, either additive (a) or dominant (d). The additive effect of the SNP indicates that one of the homozygous alleles (one or two copies) is associated with the disease compared to the other homozygous allele. The dominant effect of the SNP indicates that the heterozygous genotype is associated with the disease compared to the mean of either homozygous genotype. The numbers 1, 2, or 3 are for SNP1, SNP2, or SNP3, respectively. For example, (a) a1d2 (**Table 3**) interaction means that the presence of any minor allele genotype of SNP1 and the heterozygous genotype of SNP2 is significant. (b) d1d2d3 (**Table 4**) interaction indicates that the combination of the heterozygous genotype at the first, second, and third SNP is associated with the disease.

Association of Single *SFTP* SNPs With RDS

Out of the 17 SNPs of the five *SFTP* genes, only the rs17886395 of the *SFTPA2* was associated by itself with RDS (**Table 2**). This SNP exhibited an additive effect on RDS susceptibility (OR 0.16, 95% CI 0.06–0.43, adjusted $p = 0.03$). This particular SNP is also noted to interact with other SNPs in the two and three SNP interactions models, as shown in **Tables 3**, **4**. No other *SFTP* SNP by itself was associated with RDS at the adjusted value $p < 0.01$.

TABLE 1 | Clinical Characteristics of the cohort with and without RDS.

Variables	No RDS ($n = 458$)	RDS ($n = 477$)	*P*-value
Gestational age (weeks): median (IQR)	33 (31, 35)	30 (26, 34)	<0.001*
Sex: *n* (%)			
Female	236 (51)	198 (41)	0.02*
Male	220 (48)	277 (58)	
Race: *n* (%)			
Non-Hispanic white	328 (71)	343 (72)	
Non-Hispanic black	64 (14)	82 (17)	0.09
Hispanic	20 (4)	25 (5)	
Asian-pacific islander	23 (5)	13 (2)	
Other/mixed parents	22 (4)	13 (2)	
Infant birth weight (g) ± SD	1,818 ± 515	1,474 ± 606	<0.001*
Preterm labor: *n* (%)			
Absent	64 (14)	74 (15)	0.36
Present	203 (44)	196 (41)	
Maternal diabetes mellitus: *n* (%)			
No	419 (92)	412 (94)	0.27
Yes	33 (7)	21 (5)	
Chorioamnionitis: *n* (%)			
No	161 (35)	204 (43)	0.26
Yes	35 (8)	33 (7)	
Antenatal steroid: *n* (%)			
No	1 (0.6%)	16 (3%)	0.0003*
Yes	280 (61%)	273 (57%)	
Surfactant use: *n* (%)			
No	448 (97)	167 (35)	<0.001*
Yes	8 (2)	305 (64)	
Chronic lung disease: *n* (%)			
No	297 (65)	238 (50)	<0.001*
Yes	16 (4)	92 (20)	

The infants with RDS had younger gestational age at birth, lower birth weight, predominantly male, and had increased use of surfactant and higher incidence of chronic lung disease. The two groups (RDS, no RDS) did not differ in race, incidence of preterm labor, maternal diabetes mellitus, chorioamnionitis***.*

***Chronic lung disease included infants treated with oxygen at 28 days of life or at 36 weeks postmenstrual age (48).*

****Chorioamnionitis is diagnosed based on clinical features such as maternal fever (49).*

TABLE 2 | Single SNP associated with RDS.

Gene	SNP	Effect	Odd ratio	95% CI	*P*-value	*P*-value Adjusted*
SFTPA2	rs17886395	Additive	0.16	0.06–0.43	0.0006	0.03

**P-value is adjusted for gestational age, sex, as well as for multiple comparisons by FDR, $P < 0.05$.*

Association of Intragenic SNP-SNP Interactions With RDS in Two- and Three-SNP Interaction Model

Two SNP Model Intragenic Interactions

Among the two SNP interactions, the only intragenic interaction included *SFTPA1* SNPs; rs1136450 and rs4253527 (**Table 3**), and this combination exhibited two effects, where the d1d2 interaction was associated with increased risk for RDS (OR 1.77, 96% CI 1.42–2.19, adjusted $P = 0.0001$), and the d1a2 was associated with protection for RDS (OR 0.54, 95% CI 0.41–0.72, adjusted $P = 0.004$) (**Figure 1**).

Three SNP Model Intragenic Interactions

There were five intragenic interactions associated with RDS. Three interactions were among SNPs of the *SFTPA1* and two involved the *SFTPA2* and *SFTPB* genes. The *SFTPA2* SNPs: rs1059046, rs1965707, and rs1965708 exhibited an effect, d1d2d3, that was protective for RDS (OR = 0.55, 95% CI 0.46–0.55, adjusted $p < 0.01$). The *SFTPA1* gene variants: rs1059047 (SNP1), rs1136451 (SNP2), rs1059057 (SNP3) in a three-SNP interaction (d1a2d3) increased the risk for RDS (OR 4.09, 95% CI 2.39–7.00, adjusted $p = 0.0012$) (**Table 3**). The other intragenic interaction, d1d2d3, was found among *SFTPB* SNPs: rs2077079

TABLE 3 | The two SNP interactions associated with RDS susceptibility.

Gene1	SNP 1	Gene2	SNP2	Effect	Odds ratio	95% CI	*P*-value	*P*-value adjusted
SFTPA2	rs17886395	*SFTPD*	rs721917	d1d2	0.56	0.45–0.69	9.33E-08	9.77E-05
		SFTPA1	rs4253527	d1d2	1.69	1.32–2.07	8.88E-06	0.003097
**SFTPA1*	rs1136450	*SFTPA1*	rs4253527	d1d2	1.77	1.42–2.19	3.08E-07	0.000161
				d1a2	0.54	0.41–0.72	2.91E-05	0.004226
SFTPA2	rs1965708	*SFTPA1*	rs1059047	d2	0.43	0.29–0.62	1.61E-05	0.004507
				d1d2	1.69	1.32–2.17	2.85E-05	0.004507
SFTPB	rs2077079	*SFTPC*	rs4715	a1d2	0.19	0.09–0.38	3.04E-05	0.004507
SFTPB	rs3024798			a1d2	5.7	2.56–12.65	3.44E-05	0.004507

Interaction effect: a- additive, d-dominant, ad-additive × dominant, dd-dominant × dominant between the two SNPs. The intragenic interaction is marked with an asterisk (*).
The interactions associated with risk are highlighted in yellow.
Numbers 1 and 2 in the effect column represent SNP1 and SNP2, respectively.
The a1d2 stands for additive effect for SNP1 and dominant effect for SNP2.
The d1d2 stands for dominant effect for SNP1 and dominant effect for SNP2.
P-value is adjusted for gestational age, sex, and corrected for multiple comparisons by FDR, P-value adjusted <0.01.

(SNP1), rs3024798 (SNP2), and rs7316 (SNP3), as d1d2d3, and this was protective for RDS (OR = 0.63, 95% CI 0.52–0.76, adjusted P 0.001).

Association of Intergenic Interactions Among the Surfactant Protein Genes SNPs With RDS in a Two- and Three-SNP Model

Two SNP Model Intergenic Interactions

The two SNP interactions are shown in **Table 3**. The combination of *SFTPA2* rs17886395 (SNP1) with (i) *SFTPA1* rs4253527 (SNP2) as d1d2, increased risk of RDS (OR 1.69, 95% CI 1.32–2.17, adjusted $p = 0.004$), and (ii) *SFTPA1* rs1059047 (SNP2) as d2 without any epistatic effect from SNP1 was protective (OR 0.43, 95% CI 0.29–0.62, adjusted $p = 0.004$). The *SFTPA2* SNP rs17886395 interaction with the *SFTPD* SNP rs721917 was protective when both had a dominant effect (OR 0.56, 95% CI 0.45–0.69 adjusted $p < 0.01$). Intergenic SNP-SNP interactions were also noted between each of the two of the *SFTPB* SNPs (rs2077079 or rs3024798) and one *SFTPC* SNP rs4715 associated with protection or risk against RDS, as shown in **Table 3**.

Three SNP Model Intergenic Interactions

Table 4 shows the intergenic three SNP interactions of the *SFTP* genes associated with RDS. There were a total of 28 intergenic interactions. There were four *SFTPA2* SNPs studied. Among them, the rs17886395 SNP, found to have an additive effect and be protective for RDS by itself in the single SNP model, was present in 7 out of the 28 intergenic interactions and in 5 out of the 7 interactions were noted to be protective.

The five *SFTPA1* gene SNPs exhibited mainly a dominant effect. The rs1136450 was involved in the highest number of interactions (10 intergenic interactions), and the other *SFTPA1* SNPs had fewer than 5 interactions showing either protective or risk effect. An example of a three intergenic SNP interaction is shown diagrammatically in **Figure 2**. This figure depicts an interaction among three SNPs of *SFTPA1* and *SFTPA2*. In this intergenic interaction, the additive effect of SNP1, rs17886395, G variant that codes for alanine interacts with SNP2 (rs1059047) and SNP3 (rs1059047) of *SFTPA1* in a dominant effect. This interaction, based on odd's ratios, is associated with increased disease susceptibility. It has the highest odd's ratio (OR 4.76, 95% CI 2.67–8.47) compared to the odd's ratios of the other three SNP interactions.

The *SFTPB* SNPs (rs7316, rs1130866, rs2077079) were involved in 5 intergenic interactions, and the *SFTPD* SNPs (rs721917, rs2243639) were involved in a total of 6 intergenic interactions, and they were mainly in a dominant effect.

Hydrophobic vs. Hydrophilic Surfactant Protein Gene SNP Interactions

Figure 3 shows that the SNPs of the hydrophobic *SFTPB* and *SFTPC* interacted with each other in the two-SNP model, and the SNPs of the hydrophilic *SFTPA1, SFTPA2,* and *SFTPD* SNPs also interacted with each other. There was no interaction between any of the hydrophobic and the hydrophilic SPs SNPs. The three-SNP model (**Figure 4**) depicted an intricate network of interactions among all the *SFTP* genes, except for *SFTPC*. A total of 28 three SNP interactions were identified. The *SFTPA1* and *SFTPA2* have the maximum number of interactions and, along with *SFTPD,* interacted with *SFTPB*. All three SNP interactions, except for one intragenic interaction of *SFTPB* (rs2077079-SNP1, rs3024798-SNP2, rs7316-SNP3 as d1d2d3), involved either *SFTPA1* or *SFTPA2*. This highlights the impact and importance of *SFTPA1* and *SFTPA2* in RDS.

DISCUSSION

Although *SFTP* variants have been implicated in RDS (10, 27, 39), the statistical method used at the time had a limited ability to detect complex epistatic interactions among multiple SNPs. However, a more recent methodology by Wang et al. (53) enables investigation of complex SNP-SNP interactions by employing SNP interaction models. As one of very few statistical models that can analyze high-order interactions,

TABLE 4 | Three SNP-SNP-SNP interactions of surfactant protein genes associated with RDS.

Gene1	SNP1	Gene2	SNP2	Gene3	SNP3	Effect	Odd's ratio	95% CI	*P*-value adjusted
**SFTPA2*	rs1059046	*SFTPA2*	rs1965707	*SFTPA2*	rs1965708	d1d2d3	0.55	0.46–0.65	7.74E-08
SFTPA2	rs1965707	*SFTPA2*	rs1965708	*SFTPA1*	rs1136450	d1d2d3	0.55	0.46–0.65	1.30E-07
						d1d3	1.92	1.47–2.51	0.001018
SFTPA2	rs1059046	*SFTPA2*	rs17886395	*SFTPA1*	rs1059047	d1d2d3	0.57	0.47–0.69	3.54E-05
						d1d2d3	0.59	0.49–0.72	0.000159
SFTPA2	rs17886395	*SFTPA2*	rs1965707	*SFTPA1*	rs1136451	d1d2d3	1.57	1.3–1.89	0.001033
SFTPA2	rs1059046	*SFTPA1*	rs1136451	*SFTPA1*	rs1059057	d1d2d3	0.54	0.44–0.65	8.12E-07
SFTPA2	rs17886395	*SFTPA1*	rs1059047	*SFTPA1*	rs1059057	a1d2d3	4.76	2.67–8.47	0.001024
SFTPA2	rs17886395	*SFTPA1*	rs1136450	*SFTPA1*	rs1059057	d1d2d3	0.57	0.47–0.69	0.000401
SFTPA2	rs17886395	*SFTPA1*	rs1059047	*SFTPA1*	rs1136450	d1d2d3	0.53	0.44–0.65	8.12E-07
SFTPA2	rs1059046	*SFTPA1*	rs1136450	*SFTPA1*	rs4253527	d1d2d3	1.53	1.28–1.81	0.001018
SFTPA2	rs17886395	*SFTPA1*	rs1059047	*SFTPA1*	rs1136451	d1d2d3	0.62	0.51–0.75	0.001235
**SFTPA1*	rs1059047	*SFTPA1*	rs1136450	*SFTPA1*	rs1136451	d1d2d3	0.53	0.43–0.64	2.82E-07
**SFTPA1*	rs1136450	*SFTPA1*	rs1136451	*SFTPA1*	rs1059057	d1d2d3	0.57	0.47–0.69	1.77E-05
**SFTPA1*	rs1059047	*SFTPA1*	rs1136451	*SFTPA1*	rs1059057	d1a2d3	4.09	2.39–7.00	0.0012
SFTPA2	rs1059046	*SFTPD*	rs721917	*SFTPB*	rs7316	d1d2	0.53	0.41–0.67	0.000197
						d1d2a3	0.51	0.40–0.64	6.71E-05
SFTPA1	rs1136450	*SFTPA1*	rs4253527	*SFTPB*	rs7316	d1d2	2.01	1.56–2.60	6.71E-05
						d1d2a3	1.96	1.52–2.52	0.000196
SFTPA2	rs1965708	*SFTPD*	rs721917	*SFTPB*	rs1130866	d2d3	0.52	0.40–0.67	0.000362
SFTPA2	rs1059046	*SFTPA1*	rs4253527	*SFTPD*	rs721917	d1d3	0.51	0.40–0.64	3.54E-05
						d1a2d3	0.49	0.39–0.62	1.40E-05
SFTPA2	rs1059046	*SFTPA1*	rs1136450	*SFTPD*	rs721917	d1d2d3	0.53	0.45–0.63	1.12E-08
SFTPA2	rs17886395	*SFTPA1*	rs1136451	*SFTPD*	rs721917	d1d2d3	0.61	0.49–0.73	0.000467
SFTPA2	rs1965708	*SFTPA1*	rs1136450	*SFTPD*	rs2243639	d1d2d3	1.62	1.34–1.95	0.000273
SFTPA2	rs1965708	*SFTPA1*	rs1059057	*SFTPB*	rs2077079	d1d2d3	1.64	1.33–2.01	0.001275
SFTPA2	rs1059046	*SFTPB*	rs2077079	*SFTPB*	rs1130866	d1d2d3	0.67	0.57–0.79	0.001295
**SFTPB*	rs2077079	*SFTPB*	rs3024798	*SFTPB*	rs7316	d1d2d3	0.63	0.52–0.76	0.001029

Interaction effect: a- additive, d-dominant, for example, dda-dominant × dominant × additive among the three SNPs. The intragenic interactions are marked with asterisks (*). The interactions associated with risk are highlighted in yellow.

Numbers 1, 2 and 3 in the effect column represent SNP1, SNP2, and SNP3, respectively.

The d1 stands for dominant effect for SNP1, d2 stands for dominant effect for SNP2, and a3 stands for additive effect for SNP3.

P-value adjusted is for gestational age, sex, and corrected for multiple comparisons by FDR, $P < 0.01$.

Wang et al.'s (53) model has been used in a variety of case-control studies, showing its elegance and robustness. For example, using this model, interactions among *SFTP* SNPs were detected to impact cystic fibrosis (57), pediatric acute respiratory failure (58), and hypersensitivity pneumonitis (59). In this study, we used 17 well-studied SNPs of *SFTPA1, SFTPA2, SFTPB, SFTPC*, and *SFTPD* to investigate the impact of individual SNPs and SNP interactions among SNPs in the same gene or between two or among three different genes. This approach revealed that (a) the highest number of the two and three SNP interactions were among SNPs of the *SFTPA1* and *SFTPA2,* and these were associated with risk or protection for RDS. (b) Only the rs17886395 (C allele) of the *SFTPA2* was protective for RDS in a single-SNP model. (c) In the two SNP models, there was no interaction between the hydrophilic SFTPA1, SFTPA2, *SFTPD* SNPs, and the hydrophobic *SFTPB* or *SFTPC* SNPs. (d) the three SNP models showed intricate intergenic and intragenic interactions among SNPs of the *SFTPA1, SFTPA2, SFTPB*, and *SFTPD;* however, *SFTPC* did not interact with any of the other *SFTP* SNPs. Thus, in the present study we show not only association of a single SNP but also of two and three SNP interactions to associate with RDS susceptibility.

Association of an *SFTPA2* SNP With RDS in a Single-SNP Model

Using the stringent criteria of FDR correction with 1% ($p < 0.01$), none of the single *SFTP* SNPs was associated with RDS. When the FDR correction was set at 5% ($p < 0.05$), the rs17886395 G allele of the *SFTPA2* gene exhibited an additive effect and increased risk for neonatal RDS compared to the C allele. The $1A^3$ haplotype that includes the G allele increased the risk of TB in Mexicans (60). However, the C allele of the same SNP, found to be protective of RDS (present study), has also been protective against infection, such as RSV in Finnish infants (61). In contrast, in an Ethiopian study group, the C allele

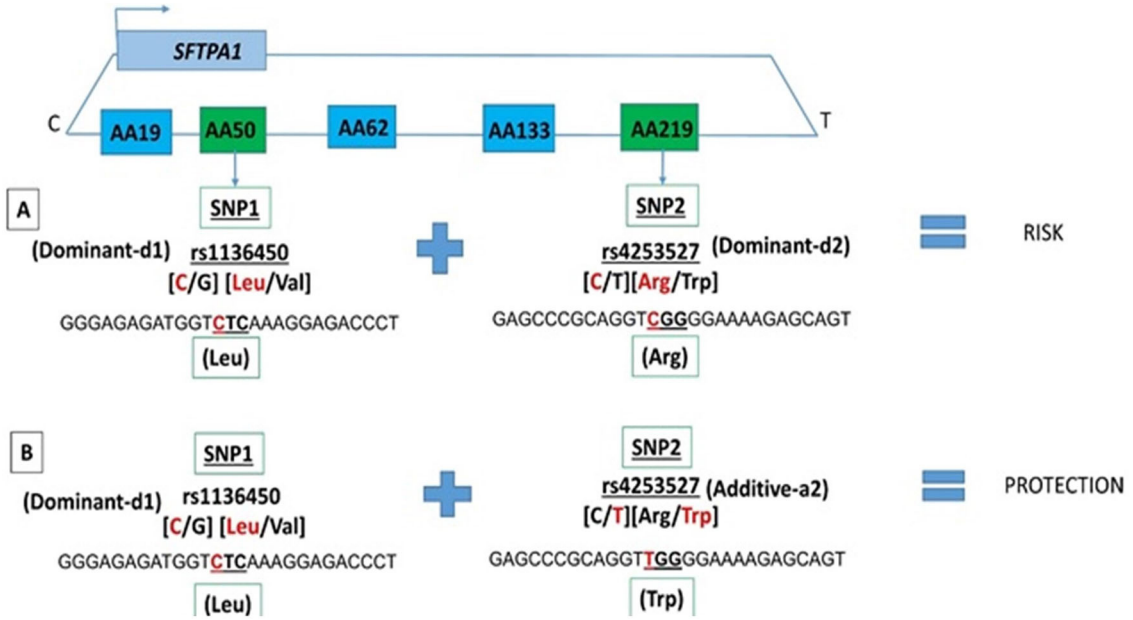


FIGURE 1 | *SFTPA1* intragenic two SNP interactions and RDS susceptibility. It shows the schematic presentation of *SFTPA1* on the top and the arrow depicts the transcriptional orientation. The relative location of SNPs are shown from centromere (C) to telomere (T) and each box represents the amino acid number that includes the particular SNP. For example, AA50 denotes the rs1136450 SNP and AA219 denotes the rs4253527 SNP. The amino acid numbering is based on the precursor molecule and thus includes the signal peptide (56). In this two SNP intragenic interaction, underneath the green boxes are the SNP ID and the actual SNPs involved. **(A)** The dominant effect, dl of SNP1, is highlighted by red, as noted for the rs1134650 C variant that codes for leucine. This interacts with the dominant effect of SNP2, rs4253527 C variant that codes for arginine (in red) and increases risk of RDS. **(B)** The dominant effect, dl of SNP1, is highlighted by red and this SNP interacts with the additive effect of SNP2 rs4253527 T variant that codes for tryptophan and this interaction is protective for RDS.

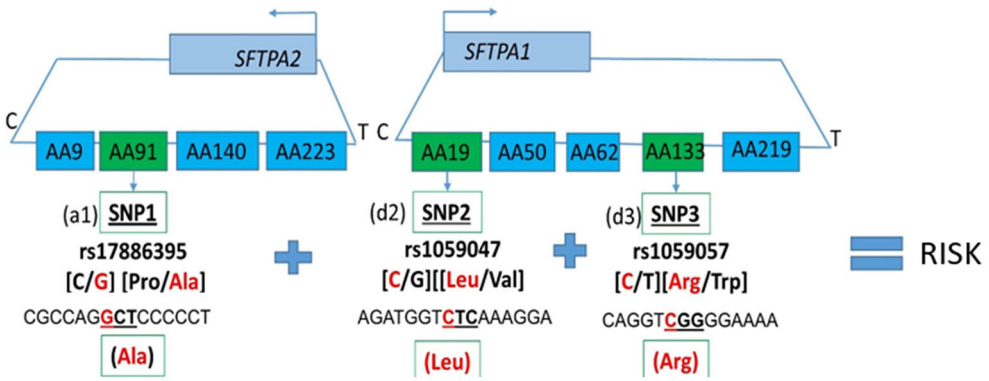


FIGURE 2 | Intergenic three SNP interaction and RDS susceptibility. It shows the schematic presentation of *SFTPA2* and *SFTPA1* on the top and the arrows depict the opposite transcriptional orientation. The relative location of SNPs is shown from centromere (C) to telomere (T) and each box represents the amino acid number that includes the particular SNP. For example, AA91 denotes the rs17886395 SNP, AA19 denotes the rs1059047 SNP, and AA133 denotes the rs1059057 SNP. In this three SNP intergenic interaction, underneath the green boxes are the SNP ID and the SNPs involved. The additive effect of SNP1, rs17886395 G variant that codes for alanine (highlighted in red) interacts with SNP2 and SNP3 of *SFTPA1* in a dominant effect and increases risk of RDS.

was associated with increased risk of TB (62), and this allele as part of 6A/1A genotype was associated with risk in community-acquired pneumonia in a Spanish study group (63). Several haplotypes of *SFTPA1* and *SFTPA2* have been well-characterized (39, 64) and the most common haplotype, $6A^{2}/1A^{0}$, has been associated with low SP-A protein expression in a study of patients with sudden infant death syndrome (65). It is of interest that the C allele of the rs17886395 SNP in pediatric diseases (i.e., RDS, RSV) is associated with protection, but in diseases likely to occur in adults (i.e., TB, community-acquired pneumonia) is associated with risk. Whether disease susceptibility by the C allele of the rs17886395 SNP is influenced by the lung environment in an age-dependent manner remains to be determined. The association of this particular SNP (rs17886395)

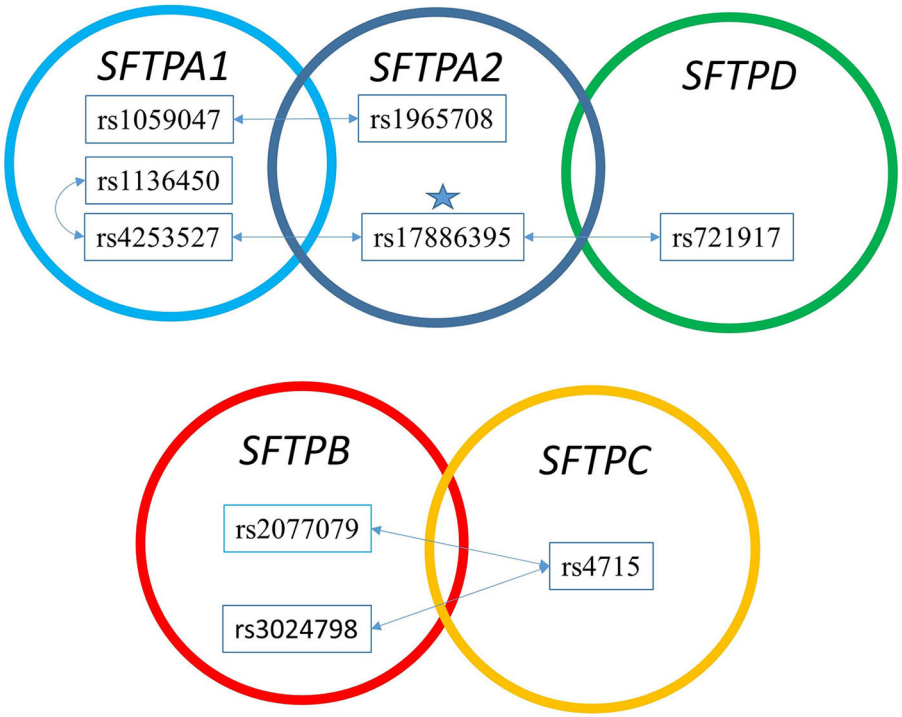


FIGURE 3 | The two SNP interaction in RDS susceptibility. Associations between RDS and the two SNP-SNP interactions are shown. The star marks the *SFTPA2* SNP shown to associate with RDS by itself. **(A)** depicts the two SNP-SNP intergenic and intragenic interactions of the hydrophilic SP genes associated with RDS. **(B)** depicts the two SNP-SNP intergenic and intragenic interactions of the hydrophobic SP genes associated with RDS.

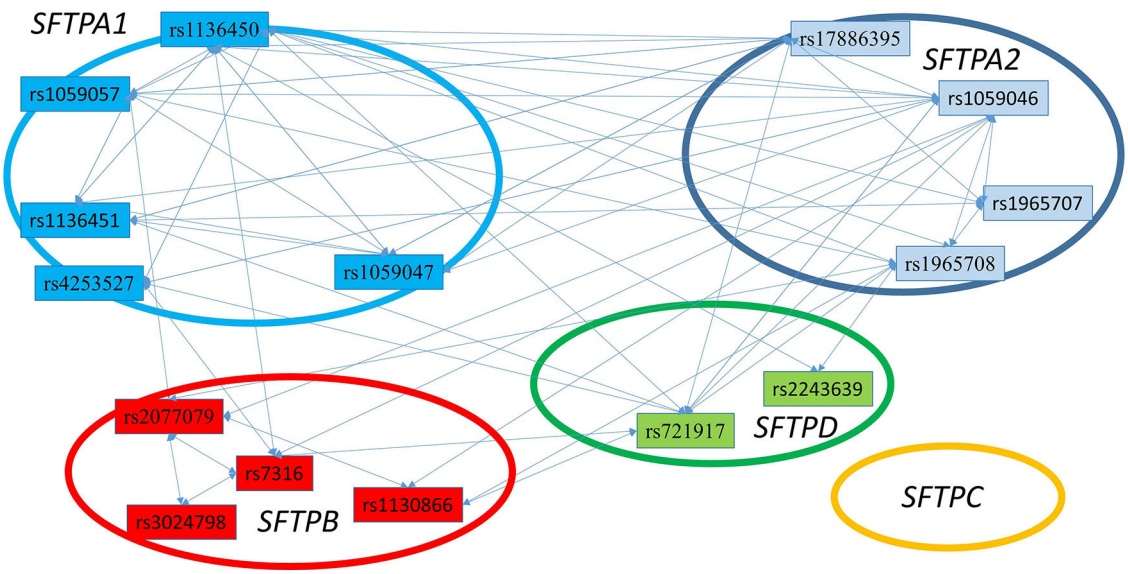


FIGURE 4 | The three SNP interactions associated with RDS susceptibility. The figure depicts intergenic and intragenic interactions of *SFTPA1*, *SFTPA2*, *SFTPD*, and *SFTPB* genes. No three SNP interactions were observed that involved *SFTPC* SNPs. There are a total of 28 three SNP interactions. All interactions (but one) involved *SFTPA1* and/or *SFTPA2*.

in RDS susceptibility in the current study may not be surprising. Infection is a common complication of RDS and prematurity, and therefore the alleles of this SNP may differentially affect disease susceptibility.

The rs17886395 (C/G) is located in the collagen-like domain of *SFTPA2* and changes the encoded amino acid Pro/Ala at codon 91 (41). It has been shown that proline normally stabilizes collagen triple helices due to conformational restrictions of the pyrrolidine ring and the presence of tertiary amides, while alanine substitutions tend to destabilize the triple helix (66). Thus, the G allele/*G*CT encoding alanine may destabilize the structure and explain the risk susceptibility.

Association of *SFTP* SNPs With RDS in a Two-SNP Model

We observed an association of the intragenic interaction between two SNPs (rs1136450 and rs4253527) of the *SFTPA1* with RDS susceptibility in the two-SNP model. The susceptibility of RDS changes based on the effect of rs4253527 in that interaction, i.e., dominant and additive effect of rs4253527 is associated with increased and decreased risk of RDS, respectively (**Figure 1**). This indicates that an additive or a dominant effect of the same SNP may change the susceptibility of an individual to a particular disease based on interactions with other SNPs. The rs4253527 (C/T) is located within the carbohydrate recognition domain (CRD) of the *SFTPA1* and changes the amino acid arginine (CGG) to tryptophan (TGG) at amino acid 219. This change may differentially affect innate immune processes under various conditions, including oxidative stress, because tryptophan is more sensitive to oxidation than arginine (56, 67). *SFTPA1* variants that differ in CRD at rs4253527 have been shown to differ in their ability to enhance phagocytosis (68) and cytokine production (69). Moreover, the CRD of surfactant proteins A and D are known to mediate binding to infectious agents such as *Pneumocystis carinii* (70, 71) and therefore the susceptibility to RDS may be interconnected with response to infection. The rs1136450 (C/G) SNP has a leucine (CTC) to valine (CTC) substitution at amino acid 50 and together with rs4253527, may impact protein function, but direct experimental evidence is lacking. Moreover, *SFTPA1* has been shown to more efficiently affect surfactant reorganization (than *SFTPA2*) in the alveolar space and inhibit surfactant inactivation by serum proteins (25). However, considering the complexity of *SFTPA* variants and their potential contribution to health and disease status, it is conceivable that the activity of a gene product in a given microenvironment, such as that in prematurity, is altered, and this may variably affect the health of the individual.

There were no significant interactions observed between SNPs of the hydrophilic and hydrophobic SPs. In contrast, previous observations have shown an association of *SFTPB* and *SFTPA1* and/or *SFTPA2* with increased risk of neonatal RDS in case-control studies (32, 34, 36). These apparent contrasting findings could be due to differences in the patient population, sample size, and/or statistical approaches used in previously reported studies and the present study.

Association of *SFTP* SNPs With RDS in a Three-SNP Model

This study, to our knowledge, is the first to show that interactions among three SNPs of the SP genes and their epistatic effect associate with RDS susceptibility. The majority of prior studies have at most reported interactions between two SNPs of the SP genes. The three SNP models in the present study showed that the highest number of intergenic and intragenic interactions involved *SFTPA1* and *SFTPA2,* indicating perhaps the importance of these genes in RDS.

An SFTPA1 SNP Is Involved in the Highest Number of the Three-SNP Interactions

The SNP rs1136450 with a dominant effect had the highest number of interactions ($n = 9$), and these were associated with either risk or protection for RDS. The rs1136450 (C/G) results in an amino acid change, Leu/Val (CTC/GTC) at codon 50 (39, 41). This SNP is located in the N-terminal collagen region and the change in amino acid may affect the binding to receptors such as calreticulin/CD91 on phagocytes (72–74). The G allele (valine) of this SNP is associated with risk of interstitial pulmonary fibrosis (IPF) in a Mexican study group (49). On the other hand, the same allele was protective in community-acquired pneumonia (63). In prior studies, this allele has been associated with risk for RDS in Finnish, whites, and blacks (29, 30); however, this was not seen in a Korean study group (75). The current study showed that this SNP had a risk or protective effect based on interactions with SNPs of other *SFTP* genes. The various interactions may change the qualitative and/or quantitative function of *SFTPA,* and this could explain the variable outcome.

SFTPA2 SNPs Are Involved in the Three-SNP Interactions

The rs1059046 SNP of *SFTPA2* was also found to have a high number of interactions ($n = 8$), and all of the interactions with a dominant effect were shown to be protective for RDS. This SNP changes the amino acid Asn/Thr at codon 9 (AAC/ACC). This amino acid is part of the signal peptide and may affect the processing of SP-A2. The A allele of this SNP of *SFTPA2* was also noted to have a protective role in community-acquired pneumonia (63). Of note, prior studies have shown the A allele, either in its homozygous or heterozygous form to be associated with risk for the respiratory syncytial virus (RSV) (61, 76) as well as influenza (77). The rs17886395 of *SFTPA2*, which was described in detail above, was also noted to have a high number of three SNP interactions ($n = 7$), five of them had a dominant effect with a protective role and the remaining two (dominant or additive) were associated with risk in RDS. These together highlight the complexity of SNP interactions and their important effect on disease susceptibility.

SFTPB SNPs Are Involved in the Three-SNP Interactions

There was one significant intragenic interaction (rs2077079, rs3024798, and rs7316). Each SNP exhibited a dominant effect and this interaction was associated with decreased risk of RDS. The rs2077079 (C/A) is located 10 nt downstream of the TATAA box, 5′ regulatory region and may affect gene transcription. The rs3024798 (A/C) is located at the splice sequence of intron 2-exon 3 and may affect splicing. The rs7316 (A/G) is located in the

3′UTR, at 4 nt upstream of the TAATAAA polyadenylation signal and may affect polyadenylation (78). The location of these SNPs indicates that these may affect the processing and/or regulation of SP-B. Whether any of these mechanisms are negatively affected in RDS remains to be determined. However, each of these three SNPs has been previously shown to associate with various lung diseases (29, 57, 79, 80). The A allele of rs2077079 is associated with risk of RDS in blacks, whereas the A allele of rs3024798 is associated with protection of RDS (29). The A allele of rs7316 is associated with risk of RDS (79) and acute lung injury in African-American children (80). However, the dominant effect of rs7316 is associated with mild CF (57). It is interesting that these SNPs by themselves have been associated with risk or protection of RDS; however, the present study highlights the importance of SNP interactions, as these could mediate a differential epistatic effect compared to individual SNPs and that this may have a significant effect on the actual health/disease outcome of an individual under certain conditions.

SFTPC SNPs Were Not Involved in the Three-SNP Interactions

None of the SNPs were identified in the three SNP model, even though single *SFTPC* SNPs have been associated with RDS (38, 81) and other pulmonary diseases such as interstitial lung disease (82). The hydrophobic *SFTPB* and *SFTPC* SNPs showed significant interactions in the two SNP model but not in the three SNP model. Furthermore, the two *SFTPC* SNPs rs1124 and rs4715 change amino acids 186 and 138, respectively. Although their effect on the functional or structural integrity of SP-C is not known, these likely affect processing of the precursor SP-C molecule rather than the mature SP-C, because these amino acids are part of the SP-C precursor and not of the mature SP-C.

SFTPD SNPs Are Involved in the Three-SNP Interactions

The *SFTPD* SNPs were involved in intergenic interactions associated with RDS susceptibility. The *SFTPD* rs721917 (C/T) SNP changes Threonine (C) to Methionine (T) at position 11 in the mature protein. The C allele of the rs721917 SNP, is associated with O- linked glycosylation of threonine leading to a partial posttranslational modification and this may alter the tendency to form multimers (83, 84). Moreover, this SNP is associated with SP-D levels, with the T allele (methionine) being correlated with increased levels (83–85). The T allele of this SNP was protective for RDS (86, 87), whereas some studies reported no association with RDS (88). The current study also supports previous observations where *SFTPA2* and *SFTPD* haplotypes were shown to be protective against RDS (42).

Although the present study has a relatively large sample size, one limitation is that the patient population differs from that of the controls in terms of age, birth weight, and sex. However, the analyses were adjusted for age and sex (birth weight was not corrected due to collinearity with gestational age). Another study limitation may be reduced generalizability as both study groups were predominantly whites. It is also possible that we have missed some significant interactions due to the use of stringent criteria such as those imposed by the FDR correction, set at 1% to avoid spurious associations. Nonetheless, the present findings need to be replicated. The SNP interactions and their association with the disease phenotype may be affected by the severity of RDS, which was not captured in this study. Around 40% of the data on important parameters such as antenatal use of steroids were missing and that may have introduced bias in the estimation of the difference between groups. The diagnosis of chronic lung disease included oxygen use at 28 days or oxygen at 36 weeks postmenstrual age. The definition for BPD has evolved over time and hence the study characteristic does not capture the current definition of BPD, consistently, as per NICHD 2019 (89).

Despite the above limitations, this study indicates a greater role of *SFTPA1* and *SFTPA2* in RDS susceptibility as they had the most interactions with SNPs of other SFTPs in the two and three-SNP models. Furthermore, the concern for infection in the setting of prematurity and chorioamnionitis sets up the *SFTPA1* and *SFTPA2* gene products, SP-A1 and SP-A2, as very important molecules for the first line of defense and regulation of various processes of the alveolar macrophage (17). Our animal studies, among others, have shown that SP-A1 and SP-A2 regulate the miRNome of the alveolar macrophage (90) and the alveolar epithelial type II cells in response to ozone exposure (91). Most importantly, these differentially affect survival in response to infection in young and old mice (92, 93) and lung function (26). Of interest, the commercially available exogenous surfactant preparations used to treat RDS, lack SP-A (94) (they only have SP-B and S-PC), but yet infection is a major comorbidity with RDS.

Furthermore, surfactant lipids and SP-A exhibit anti- and pro-inflammatory effects, respectively, on immune cells under baseline conditions, and surfactant lipids have been shown to attenuate the SP-A effect (13, 95, 96). Thus, the absence of SP-A in the exogenous surfactant preparations and the additional surfactant lipids provided by the exogenous preparation may negatively contribute to a further imbalance of pro and anti-inflammatory processes (95) in the premature lungs. With ongoing trials of SP-A peptides to treat asthma and the use of SP-A peptides to treat RSV (97–99) the present findings point to a future need to investigate SP-A as adjunct therapeutic modality for RDS as well.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: the de-identified dataset is part of the FLOROS biobank at the Penn State University, College of Medicine. Requests to access these datasets should be directed to Joanna FLoros, Jfloros@psu.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Research Board (IRB) at Penn State University, College of Medicine. Written informed consent to

participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

SA: data curation. MY, LY, and RW: formal analysis. JF: funding acquisition. BN and JF: resources. RW and JF: supervision and writing—review and editing. SA, CG, and JF: writing—original draft. All authors read and approved the final manuscript.

FUNDING

This study was supported by NIH grant R37 HL34788 to JF.

ACKNOWLEDGMENTS

The authors would like to acknowledge all the collaborators associated with the different institutions as they have been mentioned in previously published papers and Dr. R. Auten for contributing nine samples and Dr. T. Weaver and P. Ballard for contributing one specimen each.

REFERENCES

1. Avery ME, Mead J. Surface properties in relation to atelectasis and hyaline membrane disease. *AMA J Dis Children.* (1959) 97:517–23. doi: 10.1001/archpedi.1959.02070010519001
2. March of Dimes. *Infant Deaths Due to Respiratory Distress Syndrome United States | PeriStats | March Of Dimes.* (2017). Available online at: www.marchofdimes.org
3. Crowley PA. Antenatal corticosteroid therapy: a meta-analysis of the randomized trials, 1972 to 1994. *Am J Obstet Gynecol.* (1995) 173:322–35. doi: 10.1016/0002-9378(95)90222-8
4. Fanaroff AA, Stoll BJ, Wright LL, Carlo WA, Ehrenkranz RA, Stark AR, et al. Trends in neonatal morbidity and mortality for very low birthweight infants. *Am J Obstet Gynecol.* (2007) 196:147–e1. doi: 10.1016/j.ajog.2006.09.014
5. Usher RH, Allen AC, McLean FH. Risk of respiratory distress syndrome related to gestational age, route of delivery, and maternal diabetes. *Am J Obstet Gynecol.* (1971) 111:826–32. doi: 10.1016/0002-9378(71)90495-9
6. Farrell PM, Avery ME. Hyaline membrane disease. *Am Rev Respir Dis.* (1975) 111:657–88.
7. Richardson DK, Torday JS. Racial differences in predictive value of the lecithin/sphingomyelin ratio. *Am J Obstet Gynecol.* (1994) 170:1273–8. doi: 10.1016/S0002-9378(13)90449-X
8. Myrianthopoulos NC, Churchill JA, Baszynski AJ. Respiratory distress syndrome in twins. *Acta Geneticae Med Gemellologiae Twin Res.* (1971) 20:199–204. doi: 10.1017/S1120962300011628
9. Lankenau HM. A genetic and statistical study of the respiratory distress syndrome. *Eur J Pediatr.* (1976) 123:167–77. doi: 10.1007/BF00452094
10. Floros J, Kala P. Surfactant proteins: molecular genetics of neonatal pulmonary diseases. *Annu Rev Physiol.* (1998) 60:365–84. doi: 10.1146/annurev.physiol.60.1.365
11. DeMello DE, Phelps DS, Patel G, Floros J, Lagunoff D. Expression of the 35kDa and low molecular weight surfactant-associated proteins in the lungs of infants dying with respiratory distress syndrome. *Am J Pathol.* (1989) 134:1285.
12. deMello DE, Heyman S, Phelps DS, Floros J. Immunogold localization of SP-A in lungs of infants dying from respiratory distress syndrome. *Am J Pathol.* (1993) 142:1631.
13. Phelps DS. Surfactant regulation of host defense function in the lung: a question of balance. *Pediatric Pathol Mol Med.* (2001) 20:269–92. doi: 10.1080/152279501750412225
14. Kishore U, Greenhough TJ, Waters P, Shrive AK, Ghai R, Kamran MF, et al. Surfactant proteins SP-A and SP-D: structure, function and receptors. *Mol Immunol.* (2006) 43:1293–315. doi: 10.1016/j.molimm.2005.08.004
15. Floros J, Wang G, Mikerov AN. Genetic complexity of the human innate host defense molecules, surfactant protein A1 (SP-A1) and SP-A2—impact on function. *Crit Reviews™ Eukaryotic Gene Expression.* (2009) 19:125–37. doi: 10.1615/CritRevEukarGeneExpr.v19.i2.30
16. Crouch EC. Surfactant protein-D and pulmonary host defense. *Respir Res.* (2000) 1:1–16. doi: 10.1186/rr19
17. Floros J, Thorenoor N, Tsotakos N, Phelps DS. Human surfactant protein SP-A1 and SP-A2 variants differentially affect the alveolar microenvironment, surfactant structure, regulation and function of the alveolar macrophage, and animal and human survival under various conditions. *Front Immunol.* (2021) 12:2889. doi: 10.3389/fimmu.2021.681639
18. Weaver TE, Conkright JJ. Function of surfactant proteins B and C. *Annu Rev Physiol.* (2001) 63:555–78. doi: 10.1146/annurev.physiol.63.1.555
19. Cochrane CG, Revak SD. Pulmonary surfactant protein B (SP-B): structure-function relationships. *Science.* (1991) 254:566–8. doi: 10.1126/science.1948032
20. Pérez-Gil J. Structure of pulmonary surfactant membranes and films: the role of proteins and lipid–protein interactions. *Biochimica Biophysica Acta Biomembranes.* (2008) 1778:1676–95. doi: 10.1016/j.bbamem.2008.05.003
21. Cañadas O, Olmeda B, Alonso A, Pérez-Gil J. Lipid–protein and protein–protein interactions in the pulmonary surfactant system and their role in lung homeostasis. *Int J Mol Sci.* (2020) 21:3708. doi: 10.3390/ijms21103708
22. Williams MC, Hawgood S. Hamilton RL. Changes in lipid structure produced by surfactant proteins SP-A, SP-B, and SP-C. *Am J Respir Cell Mol Biol.* (1991) 5:41. doi: 10.1165/ajrcmb/5.1.41
23. Poulain FR, Allen L, Williams MC, Hamilton RL, Hawgood S. Effects of surfactant apolipoproteins on liposome structure: implications for tubular myelin formation. *Am J Physiol Lung Cell Mol Physiol.* (1992) 262:L730–9. doi: 10.1152/ajplung.1992.262.6.L730
24. Korfhagen TR, Bruno MD, Ross GF, Huelsman KM, Ikegami M, Jobe AH, et al. Altered surfactant function and structure in SP-A gene targeted mice. *Proc Nat Acad Sci USA.* (1996) 93:9594–9. doi: 10.1073/pnas.93.18.9594
25. Lopez-Rodriguez E, Pascual A, Arroyo R, Floros J, Perez-Gil J. Human pulmonary surfactant protein SP-A1 provides maximal efficiency of lung interfacial films. *Biophys J.* (2016) 111:524–36. doi: 10.1016/j.bpj.2016.06.025
26. Thorenoor N, Zhang X, Umstead TM, Halstead ES, Phelps DS, Floros J. Differential effects of innate immune variants of surfactant protein-A1 (SFTPA1) and SP-A2 (SFTPA2) in airway function after Klebsiella pneumoniae infection and sex differences. *Respir Res.* (2018) 19:1–14. doi: 10.1186/s12931-018-0723-1
27. Tsitoura MEI, Stavrou EF, Maraziotis IA, Sarafidis K, Athanassiadou A, Dimitriou G. Surfactant protein A and B gene polymorphisms and risk of respiratory distress syndrome in late-preterm neonates. *PLoS ONE.* (2016) 11:e0166516. doi: 10.1371/journal.pone.0166516
28. Somaschini M, Presi S, Ferrari M, Vergani B, Carrera P. Surfactant proteins gene variants in premature newborn infants with severe respiratory distress syndrome. *J Perinatol.* (2018) 38:337–44. doi: 10.1038/s41372-017-0018-2
29. Floros J, Fan R, Matthews A, DiAngelo S, Luo J, Nielsen H, et al. Family-based transmission disequilibrium test (TDT) and case–control association studies reveal surfactant protein A (SP-A) susceptibility alleles for respiratory distress syndrome (RDS) and possible race differences. *Clin Genet.* (2001) 60:178–87. doi: 10.1034/j.1399-0004.2001.600303.x
30. Rämet M, Haataja R, Marttila R, Floros J, Hallman M. Association between the surfactant protein A (SP-A) gene locus and respiratory-distress syndrome in the Finnish population. *Am J Human Genet.* (2000) 66:1569–79. doi: 10.1086/302906
31. Floros J, Veletza SV, Kotikalapudi P, Krizkova L, Karinch AM, Friedman C, et al. Dinucleotide repeats in the human surfactant protein-B gene and respiratory-distress syndrome. *Biochem J.* (1995) 305:583–90. doi: 10.1042/bj3050583

32. Kala P, Ten Have T, Nielsen H, Dunn M, Floros J. Association of pulmonary surfactant protein A (SP-A) gene and respiratory distress syndrome: interaction with SP-B. *Pediatr Res.* (1998) 43:169–77. doi: 10.1203/00006450-199802000-00003
33. Wambach JA, Yang P, Wegner DJ, An P, Hackett BP, Cole FS, et al. Surfactant protein-C promoter variants associated with neonatal respiratory distress syndrome reduce transcription. *Pediatr Res.* (2010) 68:216–20. doi: 10.1203/PDR.0b013e3181eb5d68
34. Marttila R, Haataja R, Guttentag S, Hallman M. Surfactant protein A and B genetic variants in respiratory distress syndrome in singletons and twins. *Am J Respir Crit Care Med.* (2003) 168:1216–22. doi: 10.1164/rccm.200304-524OC
35. Marttila R, Haataja R, Rämet M, Pokela ML, Tammela O, Hallman M. Surfactant protein A gene locus and respiratory distress syndrome in Finnish premature twin pairs. *Ann Med.* (2003) 35:344–52. doi: 10.1080/07853890310006389
36. Floros J, Fan R. Surfactant protein A and B genetic variants and respiratory distress syndrome: allele interactions. *Neonatology.* (2001) 80:22–5. doi: 10.1159/000047173
37. Hilgendorff A, Heidinger K, Bohnert A, Kleinsteiber A, König IR, Ziegler A, et al. Association of polymorphisms in the human surfactant protein-D (SFTPD) gene and postnatal pulmonary adaptation in the preterm infant. *Acta Paediatr.* (2009) 98:112–7. doi: 10.1111/j.1651-2227.2008.01014.x
38. Lahti M, Marttila R, Hallman M. Surfactant protein C gene variation in the Finnish population–association with perinatal respiratory disease. *Euro J Human Genet.* (2004) 12:312–20. doi: 10.1038/sj.ejhg.5201137
39. Silveyra P, Floros J. Genetic variant associations of human SP-A and SP-D with acute and chronic lung injury. *Front Biosci.* (2012) 17:407. doi: 10.2741/3935
40. Floros, J., and Thomas, N. (2009). Genetic variations of surfactant proteins and lung injury. Surfactant Pathogenesis and Treatment of Lung Disease, edited by Nakos G, Papathanasiou A. *Kerala, India: Research Signpost* 25–48.
41. Karinch AM, Floros J. 5'splicing and allelic variants of the human pulmonary surfactant protein A genes. *Am J Respir Cell Mol Biol.* (1995) 12:77–88. doi: 10.1165/ajrcmb.12.1.7811473
42. Thomas NJ, Fan R, Diangelo S, Hess JC, and Floros J. Haplotypes of the surfactant protein genes A and D as susceptibility factors for the development of respiratory distress syndrome. *Acta paediatrica.* (2007) 96:985–9. doi: 10.1111/j.1651-2227.2007.00319.x
43. Haataja R, Rämet M, Marttila R, Hallman M. Surfactant proteins A and B as interactive genetic determinants of neonatal respiratory distress syndrome. *Hum Mol Genet.* (2000) 9:2751–60. doi: 10.1093/hmg/9.18.2751
44. Frankel WN, Schork NJ. Who's afraid of epistasis?. *Nat Genet.* (1996) 14:371–3. doi: 10.1038/ng1296-371
45. Moore JH. The ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Hum Hered.* (2003) 56:73–82. doi: 10.1159/000073735
46. Floros J, DiAngelo S, Koptides M, Karinch AM, Rogan PK, Nielsen H, et al. Human SP-A locus: allele frequencies and linkage disequilibrium between the two surfactant protein A genes. *Am J Respir Cell Mol Biol.* (1996) 15:489–98. doi: 10.1165/ajrcmb.15.4.8879183
47. Floros J, Thomas NJ, Liu W, Papagaroufalis C, Xanthou M, Pereira S, et al. Family-based association tests suggest linkage between surfactant protein B (SP-B) (and flanking region) and respiratory distress syndrome (RDS): SP-B haplotypes and alleles from SP-B–linked loci are risk factors for RDS. *Pediatr Res.* (2006) 59:616–21. doi: 10.1203/01.pdr.0000203145.48585.2c
48. Newton ER. Chorioamnionitis and intraamniotic infection. *Clin Obstet Gynecol.* (1993). 36:795–808. doi: 10.1097/00003081-199312000-00004
49. Selman M, Lin HM, Montaño M, Jenkins AL, Estrada A, Lin Z, et al. Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis. *Hum Genet.* (2003) 113:542–50. doi: 10.1007/s00439-003-1015-4
50. Jobe AH, Bancalari E. Bronchopulmonary dysplasia. *Am J Respir Crit Care Med.* (2001) 163:1723–9. doi: 10.1164/ajrccm.163.7.2011060
51. Lin Z, Pearson C, Chinchilli V, Pietschmann SM, Luo J, Pison U, et al. Polymorphisms of human SP-A, SP-B, and SP-D genes: association of SP-B Thr131Ile with ARDS. *Clin Genet.* (2000) 58:181–91. doi: 10.1034/j.1399-0004.2000.580305.x
52. DiAngelo S, Lin Z, Wang G, Phillips S, Ramet M, Luo J, et al. Novel, non-radioactive, simple and multiplex PCR-cRFLP methods for genotyping human SP-A and SP-D marker alleles. *Dis Markers.* (1999) 15:269–81. doi: 10.1155/1999/961430
53. Wang Z, Liu T, Lin Z, Hegarty J, Koltun WA, Wu R. A general model for multilocus epistatic interactions in case-control studies. *PLoS ONE.* (2010) 5:e11384. doi: 10.1371/journal.pone.0011384
54. Taylor MB, Ehrenreich IM. Higher-order genetic interactions and their contribution to complex traits. *Trends Genet.* (2015) 31:34–40. doi: 10.1016/j.tig.2014.09.001
55. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Nat Acad Sci USA.* (2012) 109:1193–8. doi: 10.1073/pnas.1119675109
56. Wang G, Bates-Kenney SR, Tao JQ, Phelps DS, Floros J. Differences in biochemical properties and in biological function between human SP-A1 and SP-A2 variants, and the impact of ozone-induced oxidation. *Biochemistry.* (2004) 43:4227–39. doi: 10.1021/bi036023i
57. Lin Z, Thorenoor N, Wu R, DiAngelo SL, Ye M, Thomas NJ, et al. Genetic association of pulmonary surfactant protein genes, SFTPA1, SFTPA2, SFTPB, SFTPC, and SFTPD with cystic fibrosis. *Front Immunol.* (2018) 9:2256. doi: 10.3389/fimmu.2018.02256
58. Gandhi CK, Chen C, Wu R, Yang L, Thorenoor N, Thomas NJ, et al. Association of SNP–SNP interactions of surfactant protein genes with pediatric acute respiratory failure. *J Clin Med.* (2020) 9:1183. doi: 10.3390/jcm9041183
59. Gandhi CK, Chen C, Amatya S, Yang L, Fu C, Zhou S, et al. SNP and haplotype interaction models reveal association of surfactant protein gene polymorphisms with hypersensitivity pneumonitis of Mexican population. *Frontiers in medicine.* (2020) 7. doi: 10.3389/fmed.2020.588404
60. Floros J, Lin HM, García A, Salazar MA, Guo X, DiAngelo S, et al. Surfactant protein genetic marker alleles identify a subgroup of tuberculosis in a Mexican population. *J Infect Dis.* (2000) 182:1473–8. doi: 10.1086/315866
61. Lüfgren J, Rämet M, Renko M, Marttila R, Hallman M. Association between surfactant protein A gene locus and severe respiratory syncytial virus infection in infants. *J Infect Dis.* (2002) 185:283–9. doi: 10.1086/338473
62. Malik S, Greenwood CMT, Eguale T, Kifle A, Beyene J, Habte A, et al. Variants of the SFTPA1 and SFTPA2 genes and susceptibility to tuberculosis in Ethiopia. *Hum Genet.* (2006) 118:752–9. doi: 10.1007/s00439-005-0092-y
63. García-Laorden M, de Castro FR, Solé-Violán J, Rajas O, Blanquer J, Borderías L, et al. Influence of genetic variability at the surfactant proteins A and D in community-acquired pneumonia: a prospective, observational, genetic study. *Crit Care.* (2011) 15:1–12. doi: 10.1186/cc10030
64. Floros J, Wang G, Lin Z. Genetic diversity of human SP-A, a molecule with innate host defense and surfactant-related functions; characteristics, primary function, and significance. *Curr Pharmacogenomics.* (2005) 3:87–95. doi: 10.2174/1570160054022935
65. Stray-Pedersen A, Vege A, Opdal SH, Moberg S, and Rognum TO. Surfactant protein A and D gene polymorphisms and protein expression in victims of sudden infant death. *Acta paediatrica.* (2009) 98:62–8. doi: 10.1111/j.1651-2227.2008.01090.x
66. Kersteen EA, Raines RT. Contribution of tertiary amides to the conformational stability of collagen triple helices. *Biopolymers.* (2001) 59:24–8. doi: 10.1002/1097-0282(200107)59:1<24::AID-BIP1002>3.0.CO;2-N
67. Floros J, Wang G. A point of view: quantitative and qualitative imbalance in disease pathogenesis; pulmonary surfactant protein A genetic variants as a model. *Comparative Biochem Physiol Part A Mol Integrative Physiol.* (2001) 129:295–303. doi: 10.1016/S1095-6433(01)00325-7
68. Mikerov AN, Umstead TM, Gan X, Huang W, Guo X, Wang G, et al. Impact of ozone exposure on the phagocytic activity of human surfactant protein A (SP-A) and SP-A variants. *Am J Physiol Lung Cell Mol Physiol.* (2008) 294:L121–L130. doi: 10.1152/ajplung.00288.2007
69. Wang G, Umstead TM, Phelps DS, Al-Mondhiry H, Floros J. The effect of ozone exposure on the ability of human surfactant protein a variants to stimulate cytokine production. *Environ Health Perspect.* (2002) 110:79–84. doi: 10.1289/ehp.0211079
70. McCormack FX, Festa AL, Andrews RP, Linke M, Walzer PD. The carbohydrate recognition domain of surfactant protein A mediates binding to the major surface glycoprotein of Pneumocystis carinii. *Biochemistry.* (1997) 36:8092–9. doi: 10.1021/bi970313f

71. Vuk-Pavlovic Z, Standing JE, Crouch EC, Limper AH. Carbohydrate recognition domain of surfactant protein D mediates interactions with Pneumocystis carinii glycoprotein A. *Am J Respir Cell Mol Biol.* (2001) 24:475–84. doi: 10.1165/ajrcmb.24.4.3504
72. Hickling TP, Malhotra R, Sim RB. Human lung surfactant protein A exists in several different oligomeric states: oligomer size distribution varies between patient groups. *Mol Med.* (1998) 4:266–75. doi: 10.1007/BF03401923
73. Crouch EC. Collectins and pulmonary host defense. *Am J Respir Cell Mol Biol.* (1998) 19:177–201. doi: 10.1165/ajrcmb.19.2.140
74. Palaniyar N, Ikegami M, Korfhagen T, Whitsett J, McCormack FX. Domains of surfactant protein A that affect protein oligomerization, lipid structure and surface tension. *Compar Biochem Physiol Part A Mol Integrative Physiol.* (2001) 129:109–27. doi: 10.1016/S1095-6433(01)00309-9
75. Jo HS, Cho SI, Chang YH, Kim BI, Choi JH. Surfactant protein A associated with respiratory distress syndrome in Korean preterm infants: evidence of ethnic difference. *Neonatology.* (2013) 103:44–7. doi: 10.1159/000342498
76. El Saleeby CM, Li R, Somes GW, Dahmer MK, Quasney MW, DeVincenzo JP. Surfactant protein A2 polymorphisms and disease severity in a respiratory syncytial virus-infected population. *J Pediatr.* (2010) 156:409–14. doi: 10.1016/j.jpeds.2009.09.043
77. Herrera-Ramos E, López-Rodríguez M, Ruíz-Hernández JJ, Horcajada JP, Borderías L, Lerma E, et al. Surfactant protein A genetic variants associate with severe respiratory insufficiency in pandemic influenza A virus infection. *Crit Care.* (2014) 18:1–12. doi: 10.1186/cc13934
78. Lin Z, Demello DE, Batanian JR, Khammash HM, DiAngelo S, Luo J, et al. Aberrant SP-B mRNA in lung tissue of patients with congenital alveolar proteinosis (CAP). *Clin Genet.* (2000) 57:359–69. doi: 10.1034/j.1399-0004.2000.570506.x
79. Fatahi N, Niknafs N, Kalani M, Dalili H, Shariat M, Amini E, et al. Association of SP-B gene 9306 A/G polymorphism (rs7316) and risk of RDS. *J Maternal Fetal Neonatal Med.* (2018) 31:2965–70. doi: 10.1080/14767058.2017.1359829
80. Dahmer MK, O'cain P, Patwari PP, Simpson P, Li SH, Halligan N, et al. The influence of genetic variation in surfactant protein B on severe lung injury in African American children. *Crit. Care Med.* (2011) 39:1138–44. doi: 10.1097/CCM.0b013e31820a9416
81. Fatahi N, Dalili H, Kalani M, Niknafs N, Shariat M, Tavakkoly-Bazzaz J, et al. Association of SP-C gene codon 186 polymorphism (rs1124) and risk of RDS. *The J Maternal Fetal Neonatal Med.* (2017) 30:2585–9. doi: 10.1080/14767058.2016.1256994
82. Nogee LM, Dunbar AE, Wert SE, Askin F, Hamvas A, Whitsett JA. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med.* (2001) 344:573–9. doi: 10.1056/NEJM200102223440805
83. Leth-Larsen R, Garred P, Jensenius H, Meschi J, Hartshorn K, Madsen J, et al. A common polymorphism in the SFTPD gene influences assembly, function, and concentration of surfactant protein D. *J Immunol.* (2005) 174:1532–8. doi: 10.4049/jimmunol.174.3.1532
84. Sorensen GL. Surfactant protein D in respiratory and non-respiratory diseases. *Front Med.* (2018) 5:18. doi: 10.3389/fmed.2018.00018
85. Heidinger K, König IR, Bohnert A, Kleinsteiber A, Hilgendorff A, Gortner L, et al. Polymorphisms in the human surfactant protein-D (SFTPD) gene: strong evidence that serum levels of surfactant protein-D (SP-D) are genetically influenced. *Immunogenetics.* (2005) 57:1–7. doi: 10.1007/s00251-005-0775-5
86. Chang HY, Li F, Li FS, Zheng CZ, Lei YZ, Wang J. Genetic polymorphisms of SP-A, SP-B, and SP-D and risk of respiratory distress syndrome in preterm neonates. *Med Sci Monit Int Med J Exp Clin Res.* (2016) 22:5091. doi: 10.12659/MSM.898553
87. Sorensen GL, Dahl M, Tan Q, Bendixen C, Holmskov U, Husby S. Surfactant protein-D–encoding gene variant polymorphisms are linked to respiratory outcome in premature infants. *J Pediatr.* (2014) 165:683–9. doi: 10.1016/j.jpeds.2014.05.042
88. Gower WA, Nogee LM. Candidate gene analysis of the surfactant protein D gene in pediatric diffuse lung disease. *J Pediatr.* (2013) 163:1778–80. doi: 10.1016/j.jpeds.2013.06.063
89. Jensen EA, Dysart K, Gantz MG, McDonald S, Bamat NA, Keszler M, et al. The diagnosis of bronchopulmonary dysplasia in very preterm infants. An evidence-based approach. *Am J Respir Crit Care Med.* (2019) 200:751–9. doi: 10.1164/rccm.201812-2348OC
90. Noutsios GT, Thorenoor N, Zhang X, Phelps DS, Umstead TM, Durrani F, et al. SP-A2 contributes to miRNA-mediated sex differences in response to oxidative stress: pro-inflammatory, anti-apoptotic, and anti-oxidant pathways are involved. *Biol Sex Differ.* (2017) 8:1–15. doi: 10.1186/s13293-017-0158-2
91. Noutsios GT, Thorenoor N, Zhang X, Phelps DS, Umstead TM, Durrani F, et al. Major effect of oxidative stress on the male, but not female, SP-A1 type II cell miRNome. *Front Immunol.* (2019) 10:1514. doi: 10.3389/fimmu.2019.01514
92. Thorenoor N, Umstead TM, Zhang X, Phelps DS, Floros J. Survival of surfactant protein-A1 and SP-A2 transgenic mice after Klebsiella pneumoniae infection, exhibits sex-, gene-, and variant specific differences; treatment with surfactant protein improves survival. *Front Immunol.* (2018) 9:2404. doi: 10.3389/fimmu.2018.02404
93. Thorenoor N, Phelps SD, Kala P, Ravi R, Floros Phelps A, Umstead MT, et al. Impact of surfactant protein-A variants on survival in aged mice in response to Klebsiella pneumoniae infection and ozone: serendipity in action. *Microorganisms.* (2020) 8:1276. doi: 10.3390/microorganisms8091276
94. Echaide M, Autilio C, Arroyo R, Perez-Gil J. Restoring pulmonary surfactant membranes and films at the respiratory surface. *Biochimica Biophysica Acta Biomembranes.* (2017) 1859:1725–39. doi: 10.1016/j.bbamem.2017.03.015
95. Phelps D. Pulmonary surfactant modulation of host-defense function. *Appl Cardiopulmonary Pathophysiol.* (1995) 5:221–9.
96. Koptides M, Umstead TM, Floros J, Phelps DS. Surfactant protein A activates NF-kappa B in the THP-1 monocytic cell line. *Am J Physiol Lung Cell Mol Physiol.* (1997) 273:L382–8. doi: 10.1152/ajplung.1997.273.2.L382
97. Watson A, Kronqvist N, Spalluto CM, Griffiths M, Staples KJ, Wilkinson T, et al. Novel expression of a functional trimeric fragment of human SP-A with efficacy in neutralisation of RSV. *Immunobiology.* (2017) 222:111–8. doi: 10.1016/j.imbio.2016.10.015
98. Watson A, Sørensen GL, Holmskov U, Whitwell HJ, Madsen J, Clark H. Generation of novel trimeric fragments of human SP-A and SP-D after recombinant soluble expression in *E. coli*. *Immunobiology.* (2020) 225:151953. doi: 10.1016/j.imbio.2020.151953
99. Dy ABC, Tanyaratsrisakul S, Voelker DR, Ledford JG. The emerging roles of surfactant protein-A in asthma. *J Clin Cell Immunol.* (2018) 9:553. doi: 10.4172/2155-9899.1000553

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 Pediatrics

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fped.2021.727301

Neonatal Metabolic Acidosis in the Neonatal Intensive Care Unit: What Are the Genetic Causes?

Haiyan Ma[1†], Zezhong Tang[2†], Feifan Xiao[3], Long Li[4], Yangfang Li[5], Wenyan Tang[6], Liping Chen[7], Wenqing Kang[8], Yulan Lu[3], Xinran Dong[3], Guoqiang Cheng[9], Laishuan Wang[9], Wei Lu[10], Lin Yang[10], Qi Ni[3], Xiaomin Peng[3], Yao Wang[3], Yun Cao[9], Bingbing Wu[3], Wenhao Zhou[3,9], Deyi Zhuang[11], Guang Lin[1]* and Huijun Wang[3]**

[1] Department of Neonatology, Zhuhai Women and Children's Hospital, Zhuhai, China, [2] Department of Pediatrics, Peking University First Hospital, Beijing, China, [3] Center for Molecular Medicine, Children's Hospital of Fudan University, Shanghai, China, [4] Department of Neonatology, People's Hospital of Xinjiang Uygur Autonomous Region, Urumqi, China, [5] Department of Neonatology, Kunming Children's Hospital, Kunming, China, [6] Department of Neonatology, Jiangxi Maternal Hospital, Nanchang, China, [7] Department of Neonatology, Jiangxi Provincial Children's Hospital, Nanchang, China, [8] Department of Neonatology, Children's Hospital Affiliated to Zhengzhou University, Zhengzhou, China, [9] Department of Neonatology, Children's Hospital of Fudan University, Key Laboratory of Neonatal Diseases, Ministry of Health, Shanghai, China, [10] Department of Endocrinology and Inherited Metabolic Diseases, Children's Hospital of Fudan University, Shanghai, China, [11] Xiamen Key Laboratory of Neonatal Diseases, Xiamen Children's Hospital, Xiamen, China

Neonatal metabolic acidosis (NMA) is a common problem, particularly in critically ill patients in neonatal intensive care units (NICUs). Complex etiologies and atypical clinical signs make diagnosis difficult; thus, it is crucial to investigate the underlying causes of NMA rapidly and provide disorder-specific therapies. Our study aims to provide an overview of the genetic causes of NMA in patients from NICUs. We performed next-generation sequencing (NGS) on neonates with NMA from January 2016 to December 2019. Clinical features, genetic diagnoses, and their effects on clinical interventions were collected for analysis. In the 354 enrolled patients, 131 (37%) received genetic diagnoses; 95 (72.5%) of them were autosomal recessively inherited diseases. Two hundred and fifteen variants spanning 57 genes were classified as pathogenic (P) or likely pathogenic (LP) in 131 patients. The leading cause was metabolic disorders due to 35 genes found in 89 patients (68%). The other 42 NMA patients (32%) with 22 genes had malformations and renal, neuromuscular, and immune-hematological disorders. Seven genes (*MMUT*, *MMACHC*, *CHD7*, *NPHS1*, *OTC*, *IVD*, and *PHOX2B*) were noted in more than four patients, accounting for 48.9% (64/131) of the identified P/LP variants. Forty-six diagnosed patients with uncorrected NMA died or gave up. In conclusion, 37% of neonates with metabolic acidosis had genetic disorders. Next-generation sequencing should be considered when investigating the etiology of NMA in NICUs. Based on early molecular diagnoses, valuable treatment options can be provided for some genetic diseases to achieve better outcomes.

Keywords: neonatal metabolic acidosis, neonatal intensive care units, next-generation sequencing, gene, neonate

OPEN ACCESS

Edited by:
Desheng Liang,
Central South University, China

Reviewed by:
Yuan Shi,
Children's Hospital of Chongqing Medical University, China
Kejian Zhang,
Mount Sinai Genomics, United States

***Correspondence:**
Deyi Zhuang
zhuangdy526@163.com
Guang Lin
linguang_9999@126.com
Huijun Wang
huijunwang@fudan.edu.cn

†These authors share first authorship

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Pediatrics

Received: *18 June 2021*
Accepted: *20 September 2021*
Published: *18 October 2021*

Citation:
Ma H, Tang Z, Xiao F, Li L, Li Y, Tang W, Chen L, Kang W, Lu Y, Dong X, Cheng G, Wang L, Lu W, Yang L, Ni Q, Peng X, Wang Y, Cao Y, Wu B, Zhou W, Zhuang D, Lin G and Wang H (2021) Neonatal Metabolic Acidosis in the Neonatal Intensive Care Unit: What Are the Genetic Causes? Front. Pediatr. 9:727301. doi: 10.3389/fped.2021.727301

INTRODUCTION

Neonatal metabolic acidosis (NMA) is the accumulation of non-carbonic acid equivalents, which arises from excessive production or inadequate excretion of hydrogen ions or from an increased loss of bicarbonate (1). Neonatal metabolic acidosis is associated with poor clinical outcomes (2). The clinical features of NMA are atypical and are related to varying degrees of primary disease. The causes of NMA are complicated and varied, including birth asphyxia, cold stress, hypovolemia, sepsis, congenital heart disease, renal disease, and inborn errors of metabolism (1). Evaluation of detailed history, physical examination, and basic laboratory tests were insufficient to determine the cause. Administration of a fluid bolus or sodium bicarbonate is the initial management for correcting metabolic acidosis (3). These treatments do not deal with the cause but only correct the pH (3–6). Although interventions have been established in clinical practice, the impact on long-term neurological morbidity remains uncertain. With the increased awareness of the association between genotype and diseases, it has been found that infants with metabolic acidosis admitted to neonatal intensive care units (NICUs) are more likely to have genetic disorders. Thus, early genetic diagnosis is essential for neonates with NMA.

In recent years, next-generation sequencing (NGS) technology has been increasingly applied in clinical practice as a genetic diagnostic tool. Next-generation sequencing, especially exome sequencing and whole-genome sequencing, has shown advantages in facilitating an accurate diagnosis that is difficult to confirm using clinical or laboratory criteria for patients with disorders in the NICU (7, 8). Recent studies have reported enabling rapid sequencing at reduced costs and better performance in critically ill neonates (9, 10).

In this study, we analyzed the clinical medical records and genetic testing results of patients with NMA and summarized the clinical features, genetic diagnoses, and their effect on clinical intervention. Our study aims to provide a better understanding of the genetic causes of NMA that translate to more accurate management and better prognosis.

MATERIALS AND METHODS

Study Design and Participants

This was a retrospective cohort study conducted at the Clinical Genetics Laboratory of the Children's Hospital of Fudan University. Patients in NICUs were recruited retrospectively between January 1, 2016, and December 31, 2019. The inclusion criteria were as follows: (1) diagnosis of NMA based on arterial blood gas, pH $<$ 7.3, BE $<$ $-$10 mmol/L, and $PaCO_2$ 35–45 mmHg; and (2) an order for NGS. Notably, patients with NMA who did not undergo NGS were excluded. In addition, patients were excluded if they were genetically diagnosed before enrollment. The clinical features of each patient were ascertained comprehensively by a physician in addition to a thorough review of their medical records. The clinical data included sex, major clinical features, and outcomes of metabolic acidosis. Pre-test counseling was performed by physicians. This study was approved by the Medical Ethics Committee of the Children's Hospital of Fudan University (2015-130). Informed consent was obtained from the patients' parents or guardians.

Next-Generation Sequencing

Peripheral blood was collected, and genomic DNA was extracted using the TIANGEN DNA Blood Mini Kit according to the manufacturer's protocol. Sequences were generated using the Agilent ClearSeq Inherited Disease panel kit (including 2,742 genes) for clinical exome sequencing or the Agilent SureSelect XT Human All Exon V5 kit for exome sequencing. Next-generation sequencing was performed using the Illumina HiSeq X10 platform. The average on-target sequencing depth was $\times$200 for the clinical exome sequencing and $\times$120 for exome sequencing. Sequencing reads were mapped to the reference human genome (UCSC hg19) using the Burrows–Wheeler Aligner. A phenotype-scoring algorithm named PhenoPro was used for the variant filtering process (11). The detected variants were confirmed using PCR, and PCR-amplified DNA products were subjected to direct automated sequencing (3500XL Genetic Analyzer, Applied Biosystems) according to the manufacturer's specifications. *De novo* variants were detected by parental confirmation using Sanger sequencing. The pathogenicity of the variant was defined based on the American College of Medical Genetics and Genomics criteria (12). Detailed methods can be found in our previous studies (13, 14).

Statistical Analysis

Frequency count and proportions were used for categorical data. Data were analyzed using Pearson's χ^2 independence test. Statistical significance was set at a p-value $<$0.05. All statistical analyses were conducted using IBM SPSS version 20 (IBM Corp., Armonk, NY, USA).

RESULTS

Demographics of Clinical Features

From January 1, 2016, to December 31, 2019, 733 patients were clinically diagnosed with metabolic acidosis in NICUs. Three cases had genetic diagnoses, 38 neonates did not undergo NGS because their parents declined, and 338 patients did not undergo NGS for other reasons. Finally, 354 (48.3%) neonates with NGS were enrolled and classified as diagnosed and undiagnosed according to the genetic findings of our analysis (**Figure 1**). Of the enrolled patients, 186 (52.5%) were males and 168 (47.5%) were females. The major clinical manifestations (one patient may have more than one clinical feature) were neuromuscular dysfunction (49.2%), cardiopulmonary dysfunction (37.3%), hepatorenal disorder (17.2%), infection (13.6%), and malformation (12.1%) (**Table 1**).

Genetic Diagnosis

Of these 354 neonates, 131 (37%) received genetic diagnoses through NGS, while the remaining did not (n = 223, 63%). The malformation rate was higher in the diagnosed group (9.1%, p = 0.002) than in the undiagnosed group (8.0%). However,

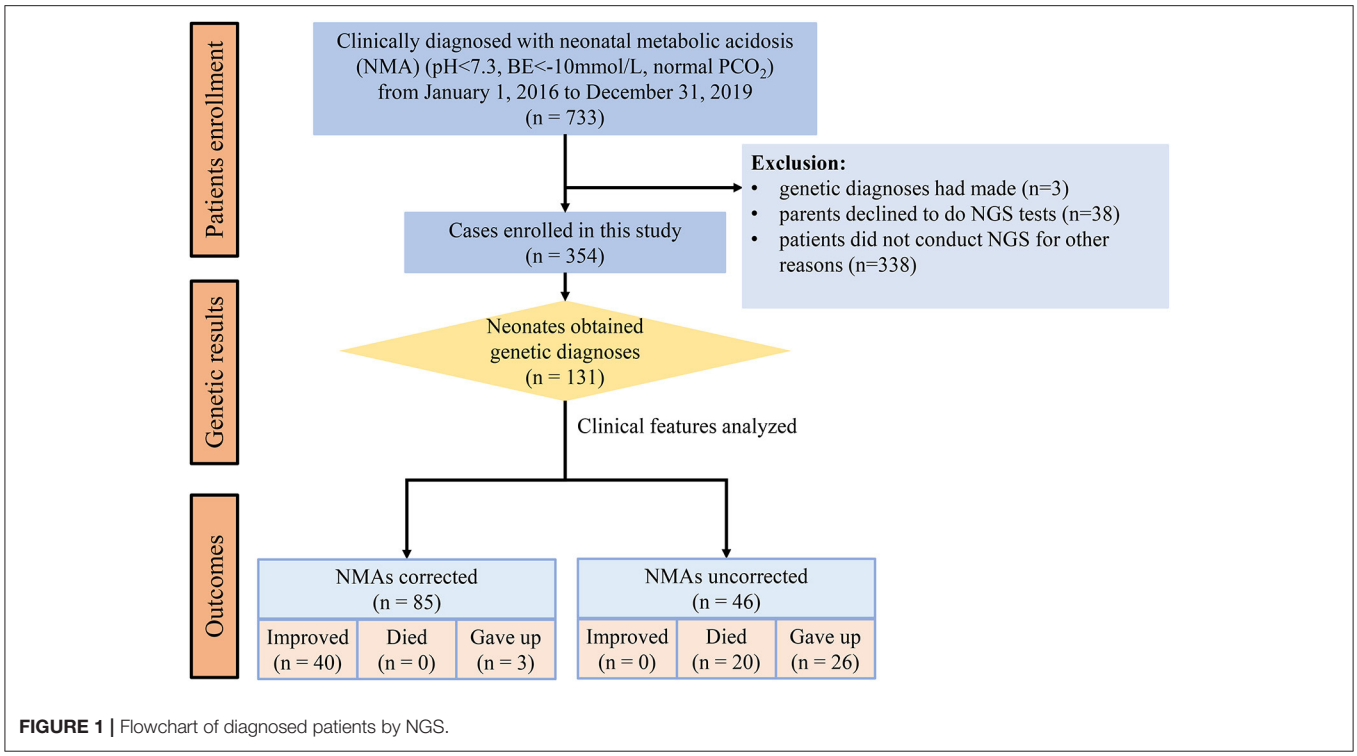


FIGURE 1 | Flowchart of diagnosed patients by NGS.

TABLE 1 | Demographic features and phenotypes of NMA patients tested with NGS.

Characteristics	Total *N* (%) (Total = 354)	Diagnosed *N* (%) (Total = 131)	Undiagnosed *N* (%) (Total = 223)	*p*-value
Sex				0.043
Male	186 (52.5)	78 (59.5)	108 (48.4)	
Female	168 (47.5)	53 (40.5)	115 (51.6)	
Major clinical features by systems*				
Neuromuscular dysfunction	174 (49.2)	56 (42.7)	118 (52.9)	0.065
Cardiorespiratory dysfunction	132 (37.3)	45 (34.3)	87 (39.0)	0.381
Hepatorenal disorder	61 (17.2)	22 (16.8)	39 (17.5)	0.867
Infection	48 (13.6)	20 (15.3)	28 (12.6)	0.472
Malformation	43 (12.1)	25 (19.1)	18 (8.1)	0.002
Outcomes of metabolic acidosis				0.63
Been corrected	224 (63.3)	85 (64.9)	139 (62.3)	
Uncorrected	130 (36.7)	46 (35.1)	84 (37.7)	
Overall outcomes				
Improved	149 (42.1)	40 (30.5)	109 (48.9)	0.001
Gave up medical support	55 (15.5)	29 (22.1)	26 (11.7)	0.009
Died with diseases	52 (14.7)	20 (15.3)	32 (14.3)	0.814

**One patient may have more than one clinical feature. p-values are obtained from a χ^2-test.*

there were no other significantly different major clinical features between the diagnosed and undiagnosed groups (**Table 1**).

In 131 patients with genetic diagnoses, 215 variants spanning 57 genes were classified as pathogenic (P) or likely pathogenic (LP). These patients were diagnosed with a monogenetic disorder, with 95 autosomal recessive cases, which included 80 compound heterozygous and 15 homozygous cases, 22 autosomal dominants with eight cases identified as *de novo*, 12 X-link recessive cases with hemizygote, and two X-link dominants with heterozygous cases.

According to the 57 phenotype–genotype-related diagnoses, diseases were classified into six categories: metabolic, renal, neuromuscular, and immune-hematological disorders; malformation; and others. Metabolic disorder was the most common disease found in 89 (67.9%) patients, followed by malformation in 22 (16.8%) patients. Renal, neuromuscular, and immune-hematological disorders were detected in nine (6.9%), six (4.6%), and three (2.3%) patients, respectively. Two other genes were detected in two patients (**Figure 2**). Seven genes with P/LP variants identified in more than four patients were *MMUT*, *MMACHC*, *CHD7*, *NPHS1*, *OTC*, *IVD*, and *PHOX2B*, accounting for 48.9% (64/131) of the patients in our cohort (**Figure 3**).

Of the metabolic disorders, 35 genes were classified into eight subgroups: amino acids, organic acids, sugar, fatty acids, carnitine, urea cycle, vitamin, and mitochondria disorders. Eleven genes were responsible for organic acid disorders, covering 53.9% (48/89) of the metabolic disorder cases. Four genes (*OTC*, *ASL*, *CPS1*, and *ASS1*) were responsible for urea

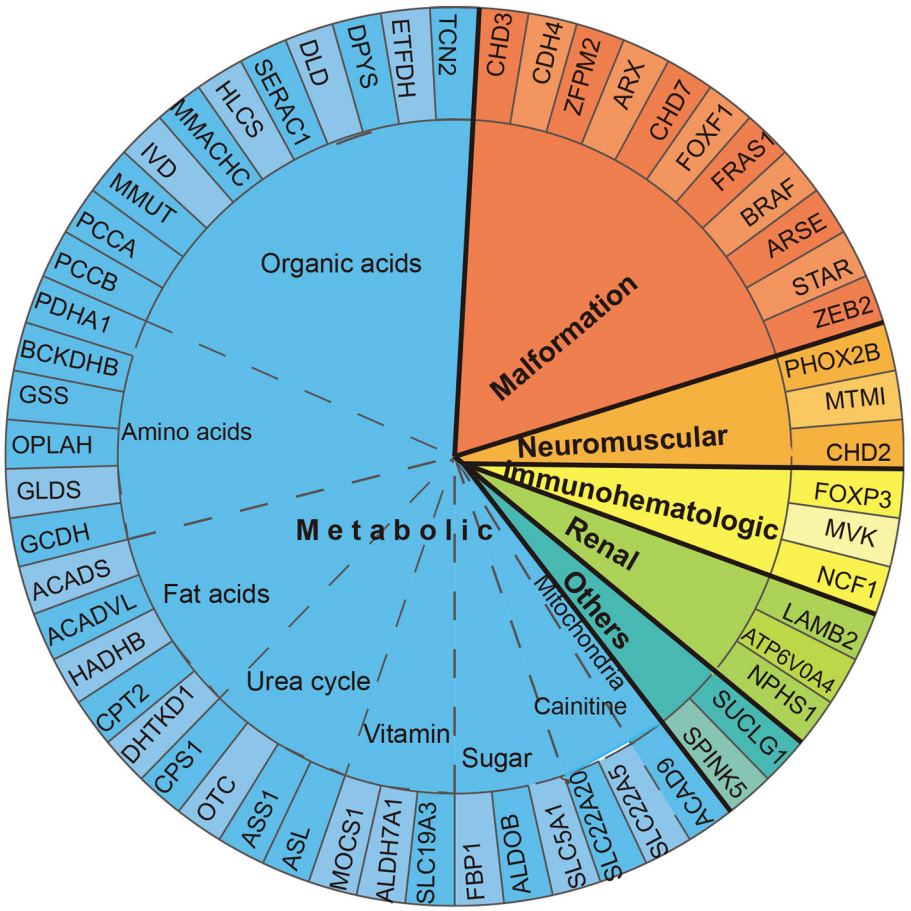


FIGURE 2 | Diagnosed gene distribution according to the disease. Fifty-seven genes are displayed by classification. Thirty-five genes related to metabolic disorder were classified into eight subgroups: organic acid, amino acid, fat acid, urea cycle, sugar, vitamin, carnitine, and mitochondria disorder.

cycle disorders, accounting for 14.6% (13/89). Fatty acid, amino acid, and carnitine disorders were detected in eight (9%), eight (9%), and five (5.6%) patients, respectively. Sugar and vitamin disorders were found in three (3.3%) patients. Mitochondria disorders were found in a patient (1.1%) (**Figure 2**).

Clinical Intervention and Outcomes

Metabolic acidosis was clinically corrected in 224 (63.3%) patients. However, metabolic acidosis was not corrected in 130 (36.7%) patients, including 46 diagnosed and 84 undiagnosed neonates. In the improved outcome cases, the rate of improvement was higher in the undiagnosed group (48.9%) than in the diagnosed group (30.5%) ($p < 0.01$). In contrast, the rate of patients who gave up medical support was higher in the diagnosed group (22.1 vs. 11.6%, $p < 0.01$) (**Table 1**). In the 131 diagnosed cases, 46 neonates with uncorrected NMAs all died or gave up. However, only three cases gave up, and no case died in the 85 neonates with corrected NMAs (**Figure 1**).

Of the 52 deaths in this study, 38.5% (20/52) received a genetic diagnosis (**Table 1**). Of these, 5 (25%) were diagnosed postmortem because their symptoms developed and progressed rapidly, resulting in mortality 24 h after admission. Through genetic diagnosis, accurate diagnoses can explain the deaths. *OTC* was the cause of death due to hyperammonemia in case NMA011 and encephalopathy and sepsis in case NMA047. Case NMA069, who carried two missense variants in *ACADVL*, died from cardiopulmonary arrest. Case NMA002 with encephalopathy and multiple organ failure was identified as a missense variant of GSS.

In addition, 40 patients with genetic diagnoses may receive precise treatment options. Similar to case NMA074, a 28-day-old male patient presented with poor feeding and lethargy. He was diagnosed with distal renal tubular acidosis after genetic identification with compound heterozygous variants in the *ATP6V0A4* gene. If the baby was not treated accurately and was followed up regularly, he would have developed severe complications, which would affect growth and development. Case NMA036, who was a 27-day-old male infant with lethargy and poor weight gain, was admitted to NICU. Blood tandem mass spectrometry showed increased tyrosine level, and the rest of the routine tests were highly suspicious for infection. Finally, methylmalonic acidemia (MMA) and homocystinuria, cblC type, were diagnosed based on genetic findings. The infant had improved upon targeted management. NMA003 was an 8-day-old baby boy who presented with vomiting, hypoglycemia,

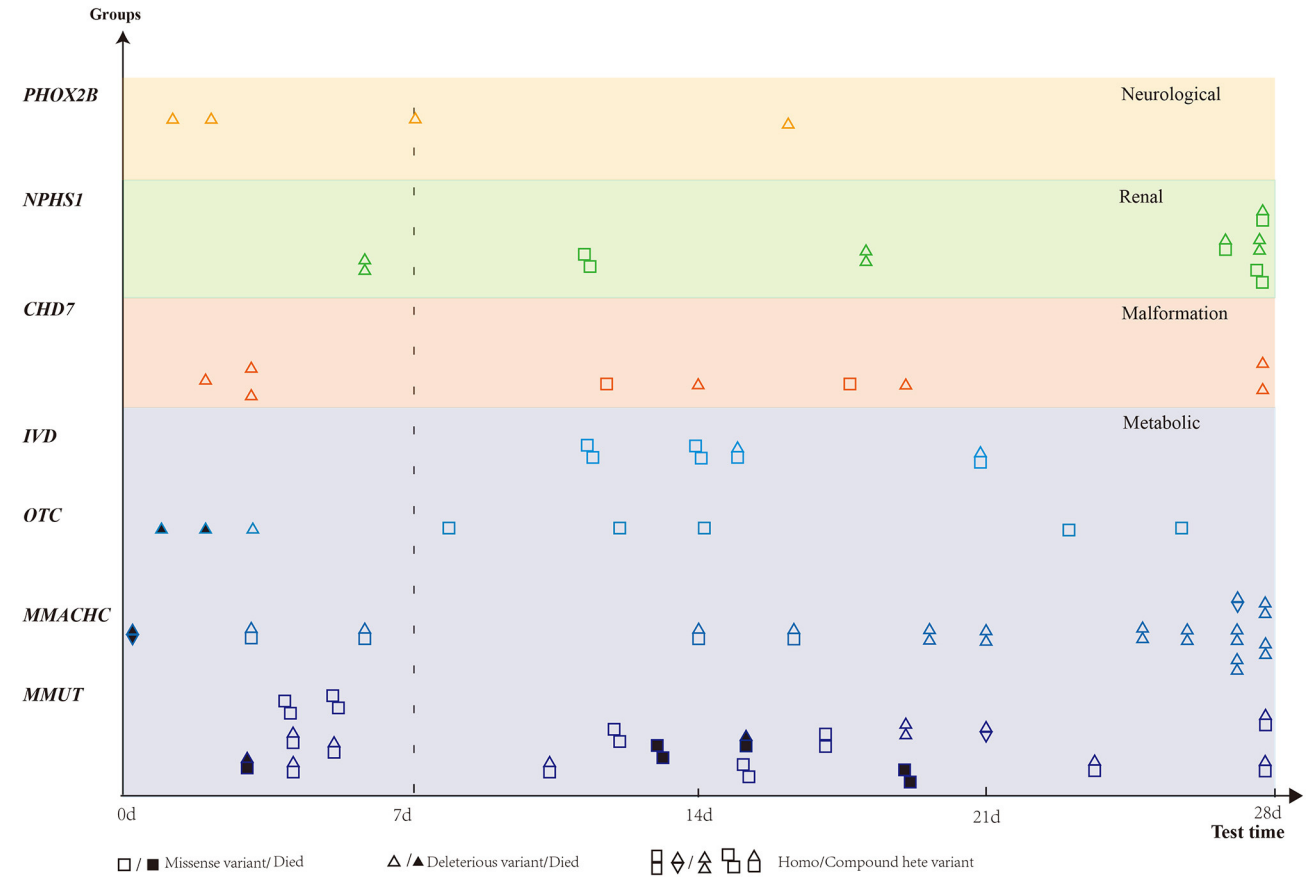


FIGURE 3 | Distribution of genes detected in more than four patients. Six genes were classified into four groups based on diseases. X-axis presents blood gas test time. Patients with missense variant are denoted by square, and deleterious variants are denoted by a triangle. Death is denoted by a solid black square or triangle.

and seizures. A high level of tyrosine was found in blood mass spectrometry, and tyrosine metabolism disorder was more likely to be diagnosed. However, a hemizygote variation (G71E) was detected in the *OTC* gene by NGS and was classified as LP. The baby was diagnosed with ornithine transcarbamylase deficiency (OTCD, MIM 311250) based on genetic findings. The patient survived on a new effective treatment option. Case NMA111, who was a 20-day-old female infant with hepatic dysfunction and suspected sepsis, had compound heterozygous variants in the gene *ALDOB*. The patient was diagnosed with hereditary fructose intolerance (HFI, MIM 229600). The disease generally develops after feeding in the neonatal period. Severity is related to the amount and duration of fructose consumption. The baby was treated immediately with a fructose intake-controlled diet, which can prevent life-threatening complications and avoid invasive procedures such as liver biopsy. The baby can grow into adulthood on a well-controlled diet.

DISCUSSION

In our study, 131 of 354 (37%) neonates with metabolic acidosis received a genetic diagnosis. In total, 57 genes were responsible for a wide range of diseases, from seemingly healthy infants to critically ill patients. Different degrees of primary disease lead to varied and atypical clinical presentations. Complex etiologies and non-specific clinical signs made diagnosis more difficult.

The leading cause of NMA was metabolic disorders, which involved 35 genes, covering 68% of genetically diagnosed patients. Only 38 (38/89, 42.9%) patients had clinical diagnoses before receiving genetic diagnoses. Organic acid disorder is the major cause of metabolic acidosis. The top two genes, *MMUT* and *MMACHC*, resulted in MMA that was related to organic acid disorder. A total of 24 (24/32, 75%) MMA patients were diagnosed with MMA using mass spectrometric tests before NGS. Next-generation sequencing can help confirm diagnoses and determine the subtype of organic acid disorder. Phenotype–genotype correlations may predict disease severity depending on mRNA stability and protein residual function (15). Variations in *MMUT* were identified as MMUT(0) phenotypes, indicating no detectable enzymatic activity (16). Variations in *MMACHC* have been detected as MMA and homocystinuria, cblC type (15). Patients who are compound heterozygotes for a missense allele appear to have a milder phenotype (15). Variation analysis was associated with responses to vitamin B12. Generally, the cblC type is almost entirely vitamin B12 responsive, the MMUT(0) type is vitamin B12 unresponsive, and other types are partly

responsive to vitamin B12 (17). *IVD* variations were found in four patients; three of them were diagnosed with isovaleric acidemia by mass spectrometric tests before NGS. Genetic findings confirmed the diagnoses and treatment. *OTC* variation results in OTCD, which accounts for approximately half of the urea cycle defects (18). As our data presented, *OTC* variations were detected in eight patients. Three of these patients were diagnosed with urea cycle defects by mass spectrometric tests before NGS. Genetic findings may help physicians understand the progress of OTCD. Six patients with missense variants that can reduce OTC enzymatic activity or stability survived OTCD (18). Patient NMA047 died 24 h after admission. He was found to carry a splicing variant, affecting mRNA processing and decreasing OTC enzyme levels (18).

Interestingly, in 42 diagnosed patients (42/131, 32%), NMA was caused by malformation and renal, neuromuscular, and immune-hematological disorders, and not by metabolic disorders. P/LP variations in *NPHS1*, which encodes nephrin, were detected in seven patients with massive proteinuria, edema, infection, and poor feeding or respiratory distress. NPHS1 is the main underlying cause of congenital nephritic syndrome, which is associated with high morbidity and mortality (19). Genetic diagnosis provides precise information on treatment and benefits patient outcomes. Pathogenic variants of *CHD7* were detected in nine neonates with infection, feeding difficulty, respiratory distress, and one or more malformations, which led to severe metabolic acidosis. These patients should be treated as if they are diagnosed with CHARGE syndrome (20). However, physicians will not give the diagnosis of CHARGE syndrome without genetic basis because criteria that focus on typical clinical phenotypes may exclude patients with a mild phenotype in the neonatal period. It has been proposed that pathogenic *CHD7* variant status is now a major criterion in CHARGE syndrome diagnoses (21). In four patients with recurrent apnea, infection, hypoxemia, and pathogenic variants, *PHOX2B* was identified. Genetic tests help physicians and families find the cause of diseases and help predict progress thereafter (22). Our study showed that mass spectrometric tests were insufficient for investigating the etiology of NMA, as opposed to NGS.

Based on early molecular diagnoses, valuable treatment options can be provided for some genetic diseases, and patients can survive or achieve better outcomes. For case NMA074, genetic diagnosis (distal renal tubular acidosis) was performed before the appearance of severe complications. With earlier adequate metabolic control, the infant can experience better growth and kidney function (23). Case NMA003, an 8-day-old male infant with abnormal mass spectrometric tests, was highly suspected to have metabolic disorder. OTCD was confirmed by genetic testing and appropriately treated. Case NMA036, a 27-day-old infant, underwent many blood tests and invasive procedures. The infant did not improve on anti-infection treatment but did improve with targeted treatment of MMA, which was diagnosed by NGS. For case NMA111, the female infant only developed clinical symptoms when she was exposed to fructose as a monosaccharide, sucrose, or sorbitol (24). The patient recovered after a fructose-controlled diet, and liver biopsy was no longer necessary. These are the most desirable results for physicians and families and help push research into new therapies.

Although there was no difference between diagnosed and undiagnosed patients, if we compare the outcomes of NMA, the genetic test can effectively identify the etiology of the disease, and these precise diagnoses enable physicians and families to understand the cause of disease and the unavoidable poor clinical outcomes (25). Genetic counseling for families provided detailed progress, helped parents identify family carriers, and facilitated better planning of reproductive choices based on the risk of familial recurrence (26). The families in our study may benefit from previous genetic diagnoses and genetic counseling.

Based on our analysis, early genetic diagnosis is essential for neonates with NMA. We proposed that NGS be performed as early as possible when patients suffer from uncorrected metabolic acidosis and combine with one of the following items: (1) abnormal blood mass spectrometry or other abnormal biochemical markers such as blood ammonia or blood lactic acid; and (2) NMA neonates accompanied with malformation. For other NMA neonates, we suggest that NGS tests such as clinical exome sequencing should be ordered when physicians think that the NMA is a genetic factor involved.

This study had two limitations that should be addressed. First, although we designed to test all NMA patients in the NICU, we had 338 cases that did not undergo NGS for different reasons. Second, as a multicenter study, follow-up information was limited, and the outcomes of some patients were missed.

CONCLUSION

In our study, we found that metabolic acidosis in approximately 37% of neonates in NICU was caused by genetic disorders. Next-generation sequencing should be considered when investigating the etiology of NMA. Based on early molecular diagnoses, valuable treatment options can be provided for some genetic diseases to achieve better outcomes.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Medical Ethics Committee of Children's Hospital of Fudan University (2015-130). 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

HW, DZ, and GL designed and supervised the overall study. FX, YLu, XD, GC, LW, WL, LY, QN, XP, and YW conducted analysis on the aggregated large cohort data. LL, YLi, WT, LC, WK, YC, BW, and WZ collected,

supervised, and reviewed the clinical data. HM and ZT wrote the original manuscript draft. HW critically reviewed and revised the manuscript for important intellectual content. All authors approved the final manuscript as submitted and agreed to be accountable for all aspects of the work.

REFERENCES

1. Walter JH. Metabolic acidosis in newborn infants. *Arch Dis Child.* (1992) 67(7 Spec No):767–9. doi: 10.1136/adc.67.7_spec_no.767
2. Kraut JA, Madias NE. Treatment of acute metabolic acidosis: a pathophysiologic approach. *Nat Rev Nephrol.* (2012) 8:589–601. doi: 10.1038/nrneph.2012.186
3. Lawn CJ, Weir FJ, McGuire W. Base administration or fluid bolus for preventing morbidity and mortality in preterm infants with metabolic acidosis. *Cochrane Database Syst Rev.* (2005) (2):CD003215. doi: 10.1002/14651858.CD003215.pub2
4. Aschner JL, Poland RL. Sodium bicarbonate: basically useless therapy. *Pediatrics.* (2008) 122:831–5. doi: 10.1542/peds.2007-2400
5. Rochwalsky U, Seitz C, Heinzmann T, Poeschl J, Koch L. [Correction of acidosis in neonatal intensive-care medicine: a national survey]. *Klin Padiatr.* (2015) 227:219–24. doi: 10.1055/s-0034-1396866
6. Collins A, Sahni R. Uses and misuses of sodium bicarbonate in the neonatal intensive care unit. *Semin Fetal Neonat Med.* (2017) 22:336–41. doi: 10.1016/j.siny.2017.07.010
7. Meng L, Pammi M, Saronwala A, Magoulas P, Ghazi AR, Vetrini F, et al. Use of exome sequencing for infants in intensive care units: ascertainment of severe single-gene disorders and effect on medical management. *JAMA Pediatr.* (2017) 171:e173438. doi: 10.1001/jamapediatrics.2017.3438
8. Smith HS, Swint JM, Lalani SR, Yamal JM, de Oliveira Otto MC, Castellanos S, et al. Clinical application of genome and exome sequencing as a diagnostic tool for pediatric patients: a scoping review of the literature. *Genet Med.* (2019) 21:3–16. doi: 10.1038/s41436-018-0024-6
9. Elliott AM, du Souich C, Lehman A, Guella I, Evans DM, Candido T, et al. RAPIDOMICS: rapid genome-wide sequencing in a neonatal intensive care unit-successes and challenges. *Eur J Pediatr.* (2019) 178:1207–18. doi: 10.1007/s00431-019-03399-4
10. Wang H, Lu Y, Dong X, Lu G, Cheng G, Qian Y, et al. Optimized trio genome sequencing (OTGS) as a first-tier genetic test in critically ill infants: practice in China. *Hum Genet.* (2020) 139:473–82. doi: 10.1007/s00439-019-02103-8
11. Li Z, Zhang F, Wang Y, Qiu Y, Wu Y, Lu Y, et al. PhenoPro: a novel toolkit for assisting in the diagnosis of Mendelian disease. *Bioinformatics.* (2019) 35:3559–66. doi: 10.1093/bioinformatics/btz100
12. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. 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.* (2015) 17:405–24. doi: 10.1038/gim.2015.30
13. Yang L, Kong Y, Dong X, Hu L, Lin Y, Chen X, et al. Clinical and genetic spectrum of a large cohort of children with epilepsy in China. *Genet Med.* (2019) 21:564–71. doi: 10.1038/s41436-018-0091-8
14. Wang H, Xiao F, Dong X, Lu Y, Cheng G, Wang L, et al. Diagnostic and clinical utility of next-generation sequencing in children born with multiple congenital anomalies in the China neonatal genomes project. *Hum Mutat.* (2021) 42:434–44. doi: 10.1002/humu.24170
15. Carrillo-Carrasco N, Chandler RJ, Venditti CP. Combined methylmalonic acidemia and homocystinuria, cblC type. I. Clinical presentations, diagnosis and management. *J Inherit Metab Dis.* (2012) 35:91–102. doi: 10.1007/s10545-011-9364-y
16. Almási T, Guey LT, Lukacs C, Csetneki K, Vokó Z, Zelei T. Systematic literature review and meta-analysis on the epidemiology of methylmalonic acidemia (MMA) with a focus on MMA caused by methylmalonyl-CoA mutase (mut) deficiency. *Orphan J Rare Dis.* (2019) 14:84. doi: 10.1186/s13023-019-1063-z
17. Zhou X, Cui Y, Han J. Methylmalonic acidemia: current status and research priorities. *Intract Rare Dis Res.* (2018) 7:73–8. doi: 10.5582/irdr.2018.01026
18. Caldovic L, Abdikarim I, Narain S, Tuchman M, Morizono H. Genotype-phenotype correlations in ornithine transcarbamylase deficiency: a mutation update. *J Genet Genomics.* (2015) 42:181–94. doi: 10.1016/j.jgg.2015.04.003
19. Sharief SN, Hefni NA, Alzahrani WA, Nazer II, Bayazeed MA, Alhasan KA, et al. Genetics of congenital and infantile nephrotic syndrome. *World J Pediatr.* (2019) 15:198–203. doi: 10.1007/s12519-018-00224-0
20. Hsu P, Ma A, Wilson M, Williams G, Curotta J, Munns CF, et al. CHARGE syndrome: a review. *J Paediatr Child Health.* (2014) 50:504–11. doi: 10.1111/jpc.12497
21. Hale CL, Niederriter AN, Green GE, Martin DM. Atypical phenotypes associated with pathogenic CHD7 variants and a proposal for broadening CHARGE syndrome clinical diagnostic criteria. *Amer J Med Genet A.* (2016) 170A:344–54. doi: 10.1002/ajmg.a.37435
22. Bishara J, Keens TG, Perez IA. The genetics of congenital central hypoventilation syndrome: clinical implications. *Appl Clin Genet.* (2018) 11:135–44. doi: 10.2147/TACG.S140629
23. Lopez-Garcia SC, Emma F, Walsh SB, Fila M, Hooman N, Zaniew M, et al. Treatment and long-term outcome in primary distal renal tubular acidosis. *Nephrol Dial Transplant.* (2019) 34:981–91. doi: 10.1093/ndt/gfy409
24. Tran C. Inborn errors of fructose metabolism. What can we learn from them? *Nutrients.* (2017). 9:356. doi: 10.3390/nu9040356
25. Bowdin SC. The clinical utility of next-generation sequencing in the neonatal intensive care unit. *CMAJ.* (2016) 188:786–7. doi: 10.1503/cmaj.160490
26. Borghesi A, Mencarelli MA, Memo L, Ferrero GB, Bartuli A, Genuardi M, et al. Intersociety policy statement on the use of whole-exome sequencing in the critically ill newborn infant. *Ital J Pediatr.* (2017) 43:100. doi: 10.1186/s13052-017-0418-0

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 in Genetics

CASE REPORT
published: 02 February 2022
doi: 10.3389/fgene.2021.752775

Renal Involvement in IPEX Syndrome With a Novel Mutation of *FOXP3*: A Case Report

Ruijuan Ke, Ying Zhu, Fang Deng and Daliang Xu**

Anhui Provincial Children's Hospital, Hefei, China

The immune dysregulation, polyendocrinopathy, enteropathy, X-linked (IPEX) syndrome is a rare genetic disease characterized by multiple immune disorders. Different mutations of the *FOXP3* gene may lead to distinct clinical manifestations. Here, we present a rare case of IPEX syndrome caused by a novel variant of *FOXP3*. Clinical manifestations include autoimmune hemolysis, bronchiectasis, diarrhea, and proteinuria but without diabetes or other endocrine disorders. The diagnosis of IPEX syndrome was confirmed by whole-exon sequencing. Supportive treatment did not ameliorate the patient's symptoms, while immunosuppressive therapy showed a promising efficacy. The patient we reported will improve the understanding of renal manifestations in IPEX syndrome.

Keywords: IPEX syndrome, FOXP3, diagnosis, treatment, whole-exon sequencing

OPEN ACCESS

Edited by:
Mingbang Wang,
Fudan University, China

Reviewed by:
Yohei Sato,
Stanford University, United States
Gesmar Rodrigues Silva Segundo,
Federal University of Uberlandia, Brazil

****Correspondence:***
Fang Deng
dengfang@ahmu.edu.cn
Daliang Xu
Xdlmy0605@163.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *03 August 2021*
Accepted: *22 December 2021*
Published: *02 February 2022*

Citation:
Ke R, Zhu Y, Deng F and Xu D (2022) Renal Involvement in IPEX Syndrome With a Novel Mutation of FOXP3*: A Case Report. Front. Genet. 12:752775. doi: 10.3389/fgene.2021.752775*

INTRODUCTION

The immunodysregulation, polyendocrinopathy, and enteropathy, X-linked (IPEX; OMIM: #304790) is a rare genetic immunologic disorder with various clinical symptoms, which is caused by mutations in the *FOXP3* gene (Gambineri et al., 2018). The onset of IPEX varies from the natal period to adulthood. Features may include refractory diarrhea, endocrine disorder, and skin disease, which are induced by the dysregulation of $CD4^+CD25^+$ regulatory T (Treg) cells (Bennett et al., 2001; Barzaghi et al., 2012). Treg cells can maintain immune tolerance to self-antigens via inhibiting activities of effective T cells. Therefore, dysregulation will lead to a multitude of clinical symptoms. Due to its X-linked recessive manner, IPEX is inherited in males, leading the boy typically die within the first 2 years if untreated (Barzaghi et al., 2018). Currently, bone marrow transplantation is the only therapeutic approach for IPEX (Ben-Skowronek, 2021).

Due to the clinical heterogeneity and low incidence of IPEX, missed and inaccurate diagnoses are common, resulting in high mortality. Herein, we report a boy who presents typical renal manifestations and carries a novel mutation in *FOXP3* with IPEX syndrome.

CASE PRESENTATION

A 12-year-old boy was admitted to our department due to proteinuria. He had a remarkable past medical history (**Figure 1C**): he was diagnosed with autoimmune hemolysis 10 years ago, chronic otitis media and bronchiectasis 3 years ago, and repeated diarrhea and alternative constipation with hair loss in recent years (**Figure 2A**). 2 weeks ago, his diarrhea aggravated to 5–6 times per day, with medium volume of dilute watery stools, with no blood and stench. The application of oral probiotics and montmorillonite powder failed to ameliorate his symptoms. Therefore, he was admitted to our hospital and transferred to our department because of a strong positive ("++++") urine protein. His

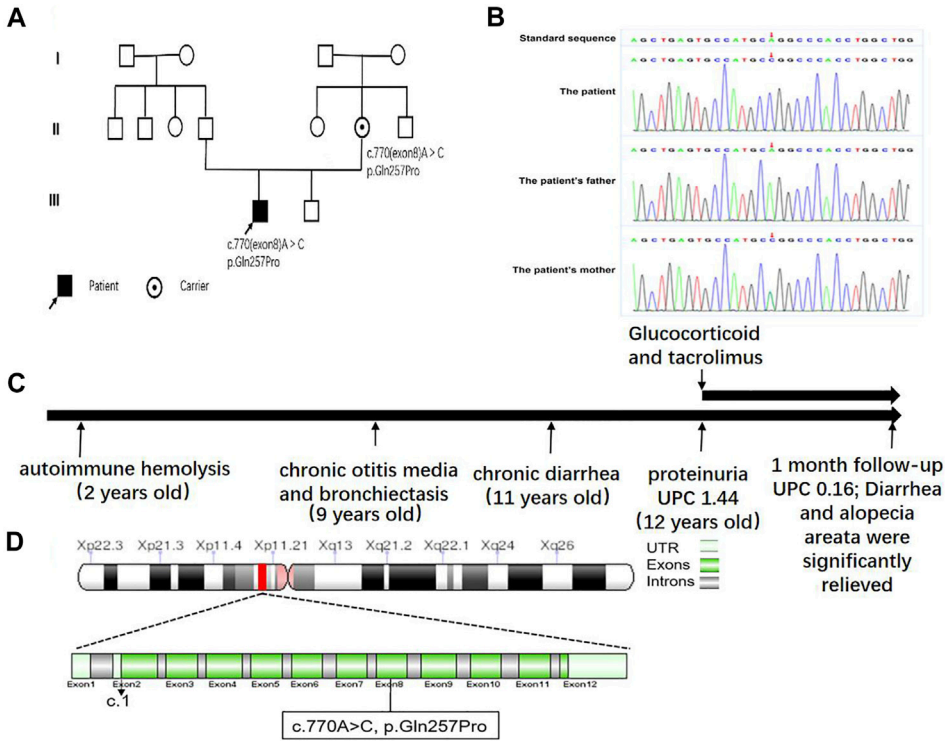


FIGURE 1 | (A) Family pedigree of the patient. **(B)** Electrophoretogram of the detected novel pathogenic variant in comparison with the wild-type FOXP3 sequence. **(C)** A schematic diagram of the disease course. UPC, urine protein–creatinine. **(D)** FOXP3 mutation analysis.

family members including his younger brother have no similar symptoms. We inquired the family history of the boy in detail and found no relatives within three generations had similar clinical manifestations. The pedigree is shown in **Figure 1A**.

Physical examination revealed that body height (141.0 cm) and weight (32.4 kg) were smaller than those of contemporary children. He had an anemic appearance with alopecia areata, purulence in the bilateral external auditory canal, hearing loss, rales and wheezing sounds in both lungs, and hyperactive bowel sounds. Lab examinations indicated that the white blood cell and $CD3^{-}CD19^{+}$ B cells were elevated, whereas the red blood cell (RBC), hemoglobin, and $CD3^{-}CD16^{+}/CD56^{+}$ NK cells were decreased (**Table 1**). Autoimmune antibodies had no positive results. Morphology examination reported several oval heteromorphic RBCs with a positive Coombs test. Erythropoietin (EPO, 42.6 IU/L) was increased, and serum transferrin was decreased (1.55 g/L). Meanwhile, the erythrocyte sedimentation rate (ESR, 68 mm/h), procalcitonin (PCT, 0.049 ng/ml), interleukin-6 (IL-6, 17.380 pg/ml), and brain natriuretic peptide (BNP, 260 pg/ml) were elevated and total albumin (28.1 g/L) was decreased. Renal function examination reported increased levels of α1-microglobulin (181.3 mg/L), β2-microglobulin (0.53 mg/L), microalbumin (1,071.3 mg/L), immunoglobulin G (68 mg/L), retinol binding protein (0.13 mg/L), N-acetyl-β-D-glucosamine (70.1 U/L), and transferrin (84 mg/L). The 24 h urine protein quantification was 289.00 mg. Fecal calprotectin (>1,800 ug/g) was notably promoted. Abdominal ultrasound and CT showed swollen and thickened ileocecal and colon walls and enlarged lymph nodes (**Figure 2A**). Thoracic high-resolution CT (HRCT) reported bronchiectasis and pneumonia mainly in the lingual lobe of the left lung and the middle lobe of the right lung (**Figure 2A**). Other examinations had no obvious abnormity.

TABLE 1 | Blood examination.

Blood index	Results	Reference range
White blood cell (x109/L)	10.21	5.2–9.1
Platelet (x109/L)	280.00	100–300
Neutrophil (x109/L)	6.73	2–7
Lymphocyte (x109/L)	2.64	0.8–4
Eosinophil (x109/L)	0.24	0.05–0.5
Red blood cell (x1,012/L)	3.29	4–4.5
Hemoglobin (g/l)	91	120–170
Reticulocyte	0.029	0.005–0.015
$CD3^{+}$ (%)	71.2	61.7–77
$CD3^{+}CD4^{+}$ (%)	36.4	25.8–41.6
$CD3^{+}CD8^{+}$ (%)	29.0	18.1–29.6
$CD3^{-}CD16^{+}/CD56^{+}$ NK (%)	1.5	10.4–19.78
$CD3^{-}CD19^{+}$ B (%)	26.1	9.02–14.1

We made the primary diagnosis as proteinuria, inflammatory bowel disease (suspected), chronic otitis media, bronchiectasis, and autoimmune hemolysis. The application of antibiotics ceftazidime and ceftizoxime did not ameliorate the diarrhea, external auditory canal purulence, proteinuria, and other symptoms. In order to

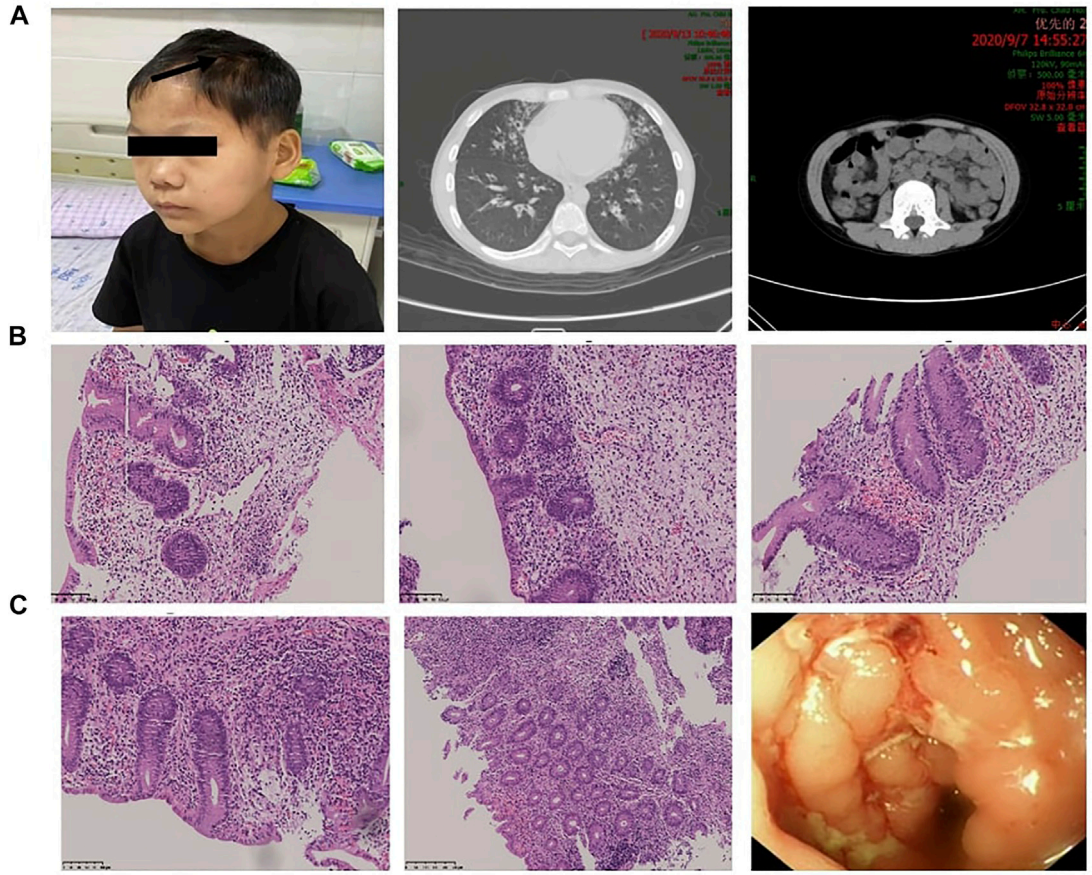


FIGURE 2 | Manifestations caused by autoimmune disorders in different organ systems. **(A)** The patient presented with alopecia areata. Chest CT showed signs of bronchiectasis, and abdominal CT showed swollen and thickened ileocecal and colon walls and enlarged lymph nodes. **(B,C)** Colonoscopic findings and pathological changes in different parts of the intestine (ileocecal junction, ascending colon, descending colon, sigmoid colon, rectum). Magnification: **(B,C)** ×200–400.

clarify the diagnosis, we performed renal biopsy under the guidance of ultrasound. Pathologic examinations including immunofluorescence (IF), electron microscope, and histological examinations were conducted, in which IF revealed IgG (+++), IgA (±), IgM (+), C3 (++), C1q (+), Fib (−), ALB (+), IgG1 (+++), IgG2 (±), IgG3 (±), IgG4 (++), PLA2R (±), and THSD7A (−) granule deposition (**Figure 3B**). Electron microscopy showed that the basement membrane of glomerular capillary loop was irregularly thickened, about 240–1,100 nm (**Figure 3C**); segmental proliferation of mesangial cells and stroma in the glomerular mesangial area; and podocyte foot process diffuse fusion (>80%). There were no obvious proliferation of parietal cells of renal capsule and no crescent formation. A large number of electron dense deposits in the basement membrane and under the epithelium of the glomerular capillary loop are accompanied by basement membrane reaction. A small amount of electron dense deposits can be seen around the transparent area, and electron dense deposits can be seen in the segmental glomerular mesangial area. Under a light microscope, the appearance of capillary loop was rigid, the segmental mesangial area was slightly widened, and there was no obvious mesangial cell proliferation; PASM staining revealed spike-like structure on the epithelial side of basement membrane of glomerular capillary loop; Masson staining reported eosinophilic deposition of subcutaneous eosinophils on the basement membrane of glomerular capillary loop (**Figure 3A**). We made the diagnosis of membranous nephropathy based on the following findings: the nail-like structure of glomerular basement membrane and the deposition of upper and subcutaneous erythrophils under the light microscope, the irregular thickening of basement membrane and the deposition of a large amount of electron dense matter under the upper and subcutaneous skin with basement membrane reaction under the electron microscope, and the fine granular deposition of IgG and C3 along the glomerular capillary loop by immunofluorescence. Considering that the concentration of PLA2R in blood is negative, we finally diagnose it as secondary membranous nephropathy.

Regarding his diarrhea, we performed colonoscopy and gastroscopy under anesthesia. Colonoscopy revealed particulate hyperplasia and submucosal hemorrhage in the terminal ileum, longitudinal ulcer in the sigmoid colon and descending colon, and erosive mucosa in the rectum (**Figures 2B,C**). Gastroscopy did not report any significant pathological changes. Biopsy indicated no obvious inflammatory changes existed in the terminal ileum, duodenum, or stomach, while chronic inflammatory cells in the

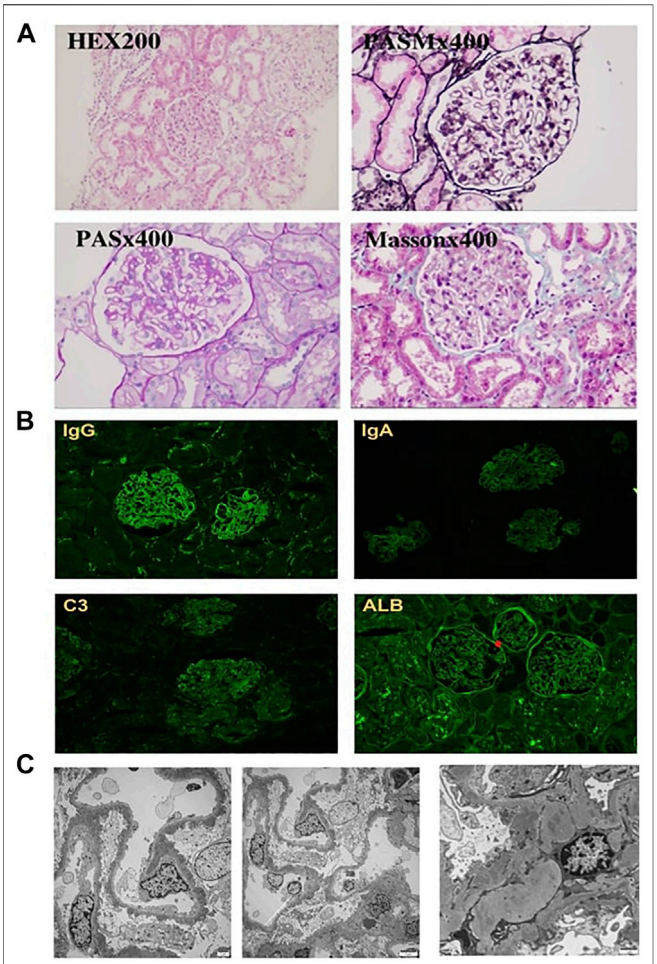


FIGURE 3 | Light and electron microscopy findings and results of immunohistochemistry studies. **(A)** HE staining did not detect any crescent, PASM staining revealed spike-like structures deposited in the basement membrane, and Masson staining reported eosinophilic deposition in the basement. **(B)** Immunoglobulin (Ig) G, Ig A, C3, and albumin deposits identified at the glomerular capillary wall by immunohistochemical staining. **(C)** Electron microscopy confirms the diagnosis of MN. Magnification: **(A,B)** ×200–400, **(C)** ×2,500.

stroma of large intestine mucosa increased, and the goblet cells decreased or disappeared. The focal crypt was distorted, a large amount of crypt inflammation and a small amount of crypt abscess were seen, and the apoptosis of crypt epithelium increased. Therefore, we made the diagnosis of inflammatory bowel disease based on these results.

However, expectant and supportive treatments failed to improve his symptoms. Since antibiotics had no significant efficacy, we intravenously injected 48 mg methylprednisolone, which slightly ameliorated his cough, while diarrhea, bloody purulent stool, and purulent auditory canal retained. In regard to autoimmune diseases such as systemic lupus erythematosus (SLE), we treated him with tacrolimus (1.0 mg at morning and 0.5 mg at evening) and 40 mg valsartan, which improved his cough, diarrhea, and proteinuria. To find the etiology, with the approval of parents, the peripheral blood of the boy and his parents was collected for whole-exon sequencing. Whole-exome sequencing yielded a mean of 10 Gb data with more than 99% coverage of the target region. Compared with the reference sequence (hg19), a missense variant, located in chr X:4 9 11 1 9 36, was detected in exon 8 of *FOXP3* (c.770A > C (p.Gln257Pro); N M _0 1 4 0 09) (**Figures 1B,D**) in both the boy and his mother. The variant was predicted to be pathogenic by all bioinformatics tools (Provean, SIFT, Polyphen2, mutationtaster, M-CAP, and REVEL); among them, Polyphen2 predicts that the pathogenicity score of the mutation at this site is 0.986, and mutationtaster predicts that the pathogenicity is disease causing. The variant has not been reported so far in ExAC, 1000 Genomes, and gnomAD. Structurally, the polypeptide surrounding the Gln257 residue is conserved across species. Sager sequencing showed the variant was not carried by his younger brother. We confirmed that the mutation segregated with IPEX within the pedigree, as illustrated in the pedigree. After discharge, the patient continued to receive the treatment of prednisone and tacrolimus and receive regular follow-up every 1 month. A recent follow-up suggested that his diarrhea, cough, alopecia areata, bilateral auditory canal pus, and hearing loss were significantly improved. Urinary protein fluctuated between "–" and "+," and hemoglobin and liver and kidney function were in the normal range. During the follow-up period, there was a soft tissue infection of the right lower limb, which recovered after guiding intravenous drip of antibiotics. The mother and the boy were satisfied with the current treatment effect and had no intention of stem cell transplantation.

DISCUSSION

In recent years, although IPEX syndrome is characterized by inflammatory bowel disease, type I diabetes, and skin diseases, the number of IPEX cases exhibiting atypical symptoms was increasing (Bacchetta et al., 2018). The onset of IPEX in most children usually occurs within a few months after birth (Luo et al., 2018), even during the fetal period (Xavier-da-Silva et al., 2015). In our case, the patient suffered from autoimmune hemolysis, bronchiectasis, chronic otitis media, alopecia, relapse diarrhea, and proteinuria. These symptoms were not the primary symptoms of IPEX syndrome. Therefore, through this case, we recognize that the consideration of IPEX syndrome is important when a child develops multiple system disorders.

IPEX syndrome was firstly reported by Powell et al. as an X-linked recessive inherited disease since 19 males were found to have IPEX syndrome in a family. As a rare immunodeficiency syndrome, the bowel disease, endocrine disorder, and skin damage are the more common clinical manifestations. As physicians of nephrology, we paid more attention to the proteinuria of the patient. Therefore, we performed renal puncture biopsy that indicated membranous nephropathy. The expression of PLA2R in biopsy tissue was weakly positive, and that in serum was negative. Besides, the patient exhibited multiple system disorders, and there was no indication of autoimmune disease. Therefore, we performed whole-exon sequencing which confirmed the diagnosis of IPEX syndrome. In 2018, a multicenter

study reported that 25% of 30 children from 25 families developed renal disease, with a median age of 3 years (1.8–9 years), in which 5 of them developed interstitial nephritis (three cases were probably linked to cyclosporine medication and one case developed chronic renal failure) (Duclaux-Loras et al., 2018a). Abir Bousset et al. (Boussetta et al., 2021) revealed that 19% of IPEX and IPEX-like syndromes would develop renal impairment, in which membranous nephropathy was the most common subtype. Other possible renal disorders included tubulointerstitial nephritis, focal tubular atrophy, minimal change disease, and irregular granular immune deposits in glomeruli and tubular basement membranes (Sheikine et al., 2015). Chuva et al. (2017) reported a case of PLA2R-positive membranous nephropathy in a child with IPEX syndrome, which indicated that PLA2R might be involved in renal impairment. However, they did not screen membranous nephropathy–related genes. In our case, the patient had isolated proteinuria, renal pathology suggested membranous nephropathy, and PLA2R was weakly positive in biopsy tissue, which might support the hypothesis of Teresa Chuv et al. that PLA2R might correlate with renal diseases. Meanwhile, the weakly positive expression of PLA2R in this case might be due to heterogeneity of the IPEX syndrome that different mutation sites caused different clinical symptoms.

A previous study reported more than 70 pathogenic variants in IPEX syndrome, in which 40% of variants located at the C-terminal forkhead (FKH) domain, 23% at the N-terminal proline-rich (PRR) domain, 14% at the leucine-zipper (LZ)-FKH loop, and 9% at the LZ domain (Barzaghi et al., 2018). It remained controversial to correlate the genetic variants with clinical manifestations. Duclaux-Loras et al. reported that clinical symptoms of IPEX syndrome had no significant association with the genetic variant in the FKH domain of FOXP3 gene (Duclaux-Loras et al., 2018b). They hypothesized that environmental exposures might modify the course of disease, leading to the varied manifestations in patients with the same mutation. In contrast, Jae HR et al. performed a systematic review that indicated the association between FOXP3 mutations and clinical manifestations: mutations in exon 1 and intron 1 of the N-terminal PRR domain were associated with infection-associated symptoms, which might be due to the mutation of a canonical LxxLL motif that could suppress the development of Th17 cells through the inhibition of retinoid-related orphan receptor-α (RORα)-mediated transcriptional activation (Du et al., 2008; Cohen et al., 2011); mutations in the LZ domain would reduce the suppressive activities mediated by Treg cells (Ziegler, 2006; Kinsey et al., 2009) and lead to Th2 cell–related diseases such as allergic rhinitis and asthma (Chatila et al., 2000; Bacchetta et al., 2006); mutations in the FKH domain would ameliorate the inhibition of transcriptional factors NFAT and AP1, which exaggerated immune responses (Wu et al., 2006); mutations in exon 11 in the FKH domain presented hematological, skin, and bowel diseases (Park et al., 2020); mutations in the ZF domain were primarily regarded to correlate with diarrhea and autoimmune hemocytopenia (Park et al., 2020). For patients with the same mutations but different manifestations, gene modification, epigenetic modification, and environmental factors were considered the contributors. In our case, we found a novel mutation in exon 8 c.770A > C (p.Gln257Pro) in the LZ domain. However, its manifestations were not consistent with the previous description (immune disorder). The primary onset of our case was autoimmune hemolysis and bronchiectasis. Diarrhea and proteinuria arose in recent years. Therefore, the correlation between mutation sites and clinical manifestations required additional investigation. In addition, the disease course in our case lasted more than 10 years, which raised the difficulty to diagnose IPEX syndrome. Our case would provide a reference for the diagnosis of IPEX syndrome that the previous history of different system impairment should be combined and taken into consideration.

Current treatment for IPEX syndrome is supportive strategies regarding clinical symptoms (Torgerson and Ochs, 2007). Early diagnosis and intervention remain to be the most important strategy to improve the prognosis. Hematopoietic stem cell transplantation presents to be the only effective approach to cure IPEX syndrome (Mazzolari et al., 2005; Rao et al., 2007). Moreover, the immunosuppressive drugs such as cyclosporine A or FK506 have been successfully used but only effective in some patients (Ferguson et al., 2000; Levy-Lahad and Wildin, 2001; Wildin et al., 2002). In our case, the patient received the treatment of glucocorticoid and tacrolimus, which successfully ameliorated the diarrhea, infection, and kidney dysfunction. Therefore, our study proposed the tailored and appropriate application of immunosuppressive drugs for the treatment of IPEX syndrome.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, and further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Anhui Provincial Children's Hospital. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

FD and DX conceived and supervised the study. RK collected the data and drafted the manuscript. YZ facilitated the collection of data. All authors reviewed and approved the final manuscript.

ACKNOWLEDGMENTS

We would like to show our gratitude for the assistance provided by Dr. Xiaoshan Tang and Chunyan Wang (Children's Hospital of Fudan University) and Dr. Yujuan Fu (Department of Pathology, Sir Run Run Shaw Hospital, Zhejiang University).

REFERENCES

Bacchetta, R., Barzaghi, F., and Roncarolo, M.-G. (2018). From IPEX Syndrome to FOXP3 Mutation: a Lesson on Immune Dysregulation. *Ann. N.Y. Acad. Sci.* 1417 (1), 5–22. doi:10.1111/nyas.13011

Bacchetta, R., Passerini, L., Gambineri, E., Dai, M., Allan, S. E., Perroni, L., et al. (2006). Defective Regulatory and Effector T Cell Functions in Patients with FOXP3 Mutations. *J. Clin. Invest.* 116 (6), 1713–1722. doi:10.1172/JCI25112

Barzaghi, F., Amaya Hernandez, L. C., Neven, B., Ricci, S., Kucuk, Z. Y., Bleesing, J. J., et al. (2018). Long-term Follow-Up of IPEX Syndrome Patients after Different Therapeutic Strategies: An International Multicenter Retrospective Study. *J. Allergy Clin. Immunol.* 141 (3), 1036–1049. doi:10.1016/j.jaci.2017.10.041

Barzaghi, F., Passerini, L., and Bacchetta, R. (2012). Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked Syndrome: a Paradigm of Immunodeficiency with Autoimmunity. *Front. Immun.* 3, 211. doi:10.3389/fimmu.2012.00211

Ben-Skowronek, I. (2021). IPEX Syndrome: Genetics and Treatment Options. *Genes* 12 (3), 323. doi:10.3390/genes12030323

Bennett, C. L., Christie, J., Ramsdell, F., Brunkow, M. E., Ferguson, P. J., Whitesell, L., et al. (2001). The Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked Syndrome (IPEX) Is Caused by Mutations of FOXP3. *Nat. Genet.* 27 (1), 20–21. doi:10.1038/83713

Boussetta, A., Jellouli, M., Isa, K., Bacherouch, H., Sherayet, K., Trabelsi, R. A., et al. (2021). Kidney Involvement in a Child with Autoimmune Disease: Answers. *Pediatr. Nephrol.* 36 (11), 3643–3644. doi:10.1007/s00467-021-05034-7

Chatila, T. A., Blaeser, F., Ho, N., Lederman, H. M., Voulgaropoulos, C., Helms, C., et al. (2000). JM2, Encoding a fork Head-Related Protein, Is Mutated in X-Linked Autoimmunity-Allergic Disregulation Syndrome. *J. Clin. Invest.* 106 (12), R75–R81. doi:10.1172/JCI11679

Chuva, T., Pfister, F., Beringer, O., Felgentreff, K., Büttner-Herold, M., and Amann, K. (2017). PLA2R-positive (Primary) Membranous Nephropathy in a Child with IPEX Syndrome. *Pediatr. Nephrol.* 32 (9), 1621–1624. doi:10.1007/s00467-017-3682-8

Cohen, C. J., Crome, S. Q., MacDonald, K. G., Dai, E. L., Mager, D. L., and Levings, M. K. (2011). Human Th1 and Th17 Cells Exhibit Epigenetic Stability at Signature Cytokine and Transcription Factor Loci. *J.I.* 187 (11), 5615–5626. doi:10.4049/jimmunol.1101058

Du, J., Huang, C., Zhou, B., and Ziegler, S. F. (2008). Isoform-Specific Inhibition of RORα-Mediated Transcriptional Activation by Human FOXP3. *J. Immunol.* 180 (7), 4785–4792. doi:10.4049/jimmunol.180.7.4785

Duclaux-Loras, R., Charbit-Henrion, F., Neven, B., Nowak, J., Collardeau-Frachon, S., Malcus, C., et al. (2018). Clinical Heterogeneity of Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked Syndrome: A French Multicenter Retrospective Study. *Clin. Transl Gastroenterol.* 9 (10), 201. doi:10.1038/s41424-018-0064-x

Duclaux-Loras, R., Charbit-Henrion, F., Neven, B., Nowak, J., Collardeau-Frachon, S., Malcus, C., et al. (2018). Clinical Heterogeneity of Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked Syndrome: A French Multicenter Retrospective Study. *Clin. Transl Gastroenterol.* 9 (10), e201. doi:10.1038/s41424-018-0064-x

Ferguson, P. J., Blanton, S. H., Saulsbury, F. T., McDuffie, M. J., Lemahieu, V., Gastier, J. M., et al. (2000). Manifestations and Linkage Analysis in X-Linked Autoimmunity-Immunodeficiency Syndrome. *Am. J. Med. Genet.* 90 (5), 390–397. doi:10.1002/(sici)1096-8628(20000228)90:5<390:aid-ajmg9>3.0.co;2-m

Gambineri, E., Ciullini Mannurita, S., Hagin, D., Vignoli, M., Anover-Sombke, S., DeBoer, S., et al. (2018). Clinical, Immunological, and Molecular Heterogeneity of 173 Patients with the Phenotype of Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked (IPEX) Syndrome. *Front. Immunol.* 9, 2411. doi:10.3389/fimmu.2018.02411

Kinsey, G. R., Sharma, R., Huang, L., Li, L., Vergis, A. L., Ye, H., et al. (2009). Regulatory T Cells Suppress Innate Immunity in Kidney Ischemia-Reperfusion Injury. *Jasn* 20 (8), 1744–1753. doi:10.1681/ASN.2008111160

Levy-Lahad, E., and Wildin, R. S. (2001). Neonatal Diabetes Mellitus, Enteropathy, Thrombocytopenia, and Endocrinopathy: Further Evidence for an X-Linked Lethal Syndrome. *J. Pediatr.* 138 (4), 577–580. doi:10.1067/mpd.2001.111502

Luo, Y., Chen, J., Fang, Y., Lou, J., and Yu, J. (2018). A Case of Metaplastic Atrophic Gastritis in Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked (IPEX) Syndrome. *BMC Pediatr.* 18 (1), 191. doi:10.1186/s12887-018-1169-9

Mazzolari, E., Forino, C., Fontana, M., D'Ippolito, C., Lanfranchi, A., Gambineri, E., et al. (2005). A New Case of IPEX Receiving Bone Marrow Transplantation. *Bone Marrow Transpl.* 35 (10), 1033–1034. doi:10.1038/sj.bmt.1704954

Park, J. H., Lee, K. H., Jeon, B., Ochs, H. D., Lee, J. S., Gee, H. Y., et al. (2020). Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked (IPEX) Syndrome: A Systematic Review. *Autoimmun. Rev.* 19 (6), 102526. doi:10.1016/j.autrev.2020.102526

Rao, A., Kamani, N., Filipovich, A., Lee, S. M., Davies, S. M., Dalal, J., et al. (2007). Successful Bone Marrow Transplantation for IPEX Syndrome after Reduced-Intensity Conditioning. *Blood* 109 (1), 383–385. doi:10.1182/blood-2006-05-025072

Sheikine, Y., Woda, C. B., Lee, P. Y., Chatila, T. A., Keles, S., Charbonnier, L.-M., et al. (2015). Renal Involvement in the Immunodysregulation, Polyendocrinopathy, Enteropathy, X-Linked (IPEX) Disorder. *Pediatr. Nephrol.* 30 (7), 1197–1202. doi:10.1007/s00467-015-3102-x

Torgerson, T. R., and Ochs, H. D. (2007). Immune Dysregulation, Polyendocrinopathy, Enteropathy, X-Linked: Forkhead Box Protein 3 Mutations and Lack of Regulatory T Cells. *J. Allergy Clin. Immunol.* 120 (4), 744–750. doi:10.1016/j.jaci.2007.08.044

Wildin, R. S., Smyk-Pearson, S., and Filipovich, A. H. (2002). Clinical and Molecular Features of the Immunodysregulation, Polyendocrinopathy, Enteropathy, X Linked (IPEX) Syndrome. *J. Med. Genet.* 39 (8), 537–545. doi:10.1136/jmg.39.8.537

Wu, Y., Borde, M., Heissmeyer, V., Feuerer, M., Lapan, A. D., Stroud, J. C., et al. (2006). FOXP3 Controls Regulatory T Cell Function through Cooperation with NFAT. *Cell* 126 (2), 375–387. doi:10.1016/j.cell.2006.05.042

Xavier-da-Silva, M. M., Moreira-Filho, C. A., Suzuki, E., Patricio, F., Coutinho, A., and Carneiro-Sampaio, M. (2015). Fetal-onset IPEX: Report of Two Families and Review of Literature. *Clin. Immunol.* 156 (2), 131–140. doi:10.1016/j.clim.2014.12.007

Ziegler, S. F. (2006). FOXP3: of Mice and Men. *Annu. Rev. Immunol.* 24, 209–226. doi:10.1146/annurev.immunol.24.021605.090547

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 Pediatrics

ORIGINAL RESEARCH
published: 07 March 2022
doi: 10.3389/fped.2022.749107

LncRNA-RMST Functions as a Transcriptional Co-regulator of SOX2 to Regulate miR-1251 in the Progression of Hirschsprung's Disease

Lingling Zhou[1†], Zhengke Zhi[2†], Pingfa Chen[3†], Chunxia Du[2], Binyu Wang[2], Xiang Fang[2], Weibing Tang[2] and Hongxing Li[2]*

[1] General Surgery Department, Children's Hospital of Wujiang District, Suzhou, China, [2] Department of Pediatric Surgery, Children's Hospital of Nanjing Medical University, Nanjing, China, [3] Intensive Care Unit, The Fourth Affiliated Hospital of Nantong University, The First People's Hospital of Yancheng, Yancheng, China

OPEN ACCESS

Edited by:
Mingbang Wang,
Fudan University, China

Reviewed by:
Seyed Reza Mohebbi,
Shahid Beheshti University of Medical Sciences, Iran
Peter Natesan Pushparaj,
King Abdulaziz University, Saudi Arabia

***Correspondence:**
Hongxing Li
hx8817@njmu.edu.cn

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Pediatrics

Received: *29 July 2021*
Accepted: *12 January 2022*
Published: *07 March 2022*

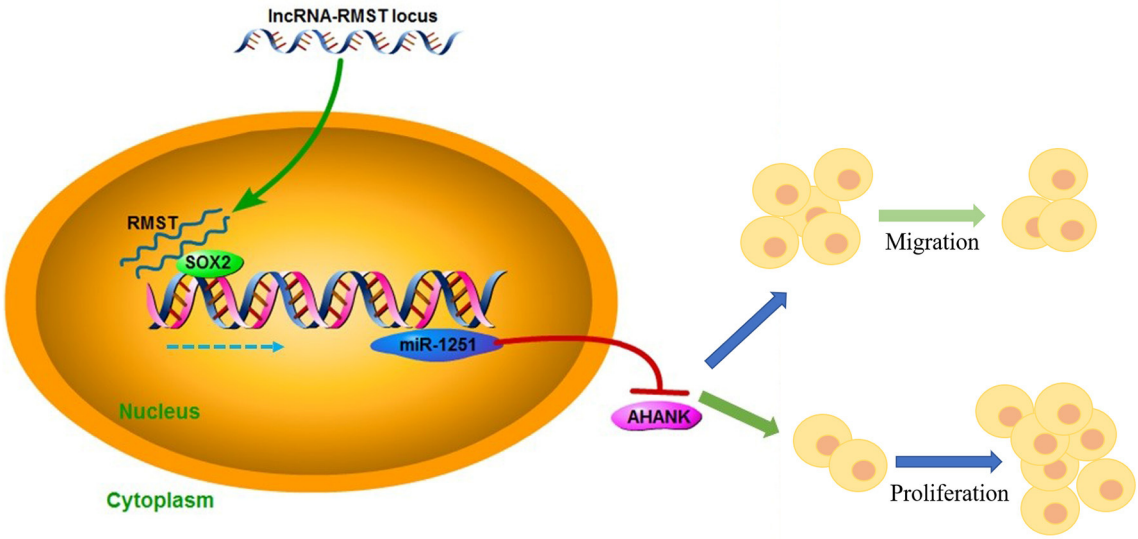
Citation:
Zhou L, Zhi Z, Chen P, Du C, Wang B, Fang X, Tang W and Li H (2022) LncRNA-RMST Functions as a Transcriptional Co-regulator of SOX2 to Regulate miR-1251 in the Progression of Hirschsprung's Disease. Front. Pediatr. 10:749107. doi: 10.3389/fped.2022.749107

Hirschsprung's disease (HSCR) is a congenital disorder characterized by the absence of enteric neural crest cells (ENCCs). LncRNA rhabdomyosarcoma 2-associated transcript (RMST) is essential for the growth and development of neuron. This study aimed to reveal the role of RMST in the pathogenesis of HSCR. The expression level of RMST, miR-1251, SOX2, and AHNAK was evaluated with qRT-PCR or western blot. CCK-8 and transwell assays were applied to detect cell proliferation and migration. CHIP and RIP assays were applied to determine the combination relationship between SOX2 and promoter region of miR-1251 or RMST and SOX2, respectively. Dual-luciferase reporter assay was performed to confirm miR-1251 targeted AHNAK. As results have shown, RMST was downregulated in the aganglionic colon of HSCR patients. The knockdown of RMST attenuated cell proliferation and migration significantly. MiR-1251, the intronic miRNA of RMST, was also low expressed in HSCR, but RMST did not alter the expression of miR-1251 directly. Furthermore, SOX2 was found to regulate the expression of miR-1251 *via* binding to the promoter region of miR-1251, and RMST strengthened this function by interacting with SOX2. Moreover, AHNAK was the target gene of miR-1251, which was co-regulated by RMST and SOX2. In conclusion, our study demonstrated that RMST functioned as a transcriptional co-regulator of SOX2 to regulate miR-1251 and resulted in the upregulation of AHNAK, leading to the occurrence of HSCR. The novel RMST/SOX2/miR-1251/AHNAK axis provided potential targets for the diagnosis and treatment of HSCR during embryonic stage.

Keywords: Hirschsprung's disease, lncRNA-RMST, miR-1251, SOX2, AHNAK

INTRODUCTION

Hirschsprung's disease (HSCR), an enteric neuropathy, is characterized by the absence of gangliocytes in the distal colon (1, 2). It is caused by the impaired migration and proliferation of enteric neural crest cells (ENCCs) during the 5th to 12th weeks of embryogenesis (3). HSCR usually occurs in about 1/5,000 neonates, while the incidence rate of females is about a quarter of



Graphical Abstract |

males (4). Current etiological studies show that HSCR is a complicated disorder involving multiple genetic factors, including RET, GDNF, GFRA1, EDNRB, and PHOX2B (5, 6). However, these genes could only explain a portion of the known cases, so further research is needed.

Long non-coding RNAs (lncRNAs) are increasingly considered to be important players in cellular biological processes, such as cell proliferation and migration, by affecting gene expression at nearly all levels (7, 8). LncRNA TPTEP1 was reported to inhibit the proliferation of non-small cell lung cancer cells through abating miR-328-5p expression (9). In renal cell carcinoma, lncRNA00312 attenuated cell proliferation and migration significantly (10). In addition, lncRNA DRAIC was found to regulate cell proliferation and migration in HSCR by affecting the miR-34a-5p/ITGA6 pathway (11). The silence of AFAP1-AS1 promoted HSCR progression by acting as an endogenous RNA to absorb miR-195 (12). Moreover, LOC100507600 is proved to participate in the development of HSCR through regulating BMI1 expression in a miR128-1-3p–dependent manner (13). However, the functions of lncRNAs in HSCR remain largely unknown.

RMST has been authenticated as a critical role during the neuronal differentiation (14, 15). Briefly, Ng et al. (16) induced differentiation of the ReN-VM neural stem cells using N2B27 medium, and 7 days later, the control group yielded $TUJ1^+$ and $MAP2^+$ neurons, while very few positive stained cells were observed in RMST knockdown group. In addition, RMST could promote the activation of microglial cells by activating TAK1-mediated NF-κB signaling (17). From the microarray analysis in our previous study, we found that RMST was downregulated in HSCR, indicating that RMST might play a key role in the progression of HSCR (18). Interestingly, miR-1251, which was transcribed from the same genomic region as RMST, was also downregulated in aganglionic segment. However, we found that RMST did not regulate miR-1251 directly. There might be other regulatory mechanisms to be uncovered.

Sex determining region Y (SRY)-box 2 (SOX2) is implicated in transcriptional regulation (19, 20). For example, SOX2 has been shown to regulate multiple malignant processes of breast cancer through the SOX2/miR-181a-5p, miR-30e-5p/TUSC3 axis (21). More importantly, SOX2 is closely related to the nervous system, and the terminal differentiation of postmitotic olfactory neurons was directly regulated by SOX2 (22). Numerous evidence has indicated that the downregulation of SOX2 attenuated cell growth and migration obviously (23, 24). Herein, SOX2 was found low expressed in HSCR and was predicted to bind to the promoter region of miR-1251 through bioinformatics analysis. Furthermore, RMST could enhance the regulation of SOX2 on downstream genes by interacting with SOX2 (25).

In this study, we demonstrated that RMST was downregulated in HSCR aganglionic colon and inhibited cell proliferation and migration by functioning as a transcriptional co-regulator of SOX2 to regulate miR-1251 in the progression of HSCR.

MATERIALS AND METHODS

Clinical Information

This study was approved by the Institutional Ethics Committee of Children's Hospital of Nanjing Medical University (approval number: 201703057), and the experiments were carried out according to approved guidelines. In total, 32 aganglionic

TABLE 1 | Clinical characteristics of study population.

Variable	Control	HSCR	*p*
Age(d, mean±SE)	103.1 ± 10.47	126.8 ± 14.31	0.19[a]
Gender			
Male	25	26	0.76[b]
Female	7	6	

[a] *Student's test.*
[b] *Two-sided chi-squared test.*

TABLE 2 | Primer sequences for quantitative RT-PCR.

Target gene	Primer sequence (5′-3′)
GAPDH	F: GCACCGTCAAGGCTGAGAAC
	R: GGATCTCGCTCCTGGAAGATG
U6	F: CTCGCTTCGGCAGCACA
	R: AACGCTTCACGAATTTGCG
RMST	F: ACTTCTGAGTGGTATGCTGCT
	R: GGATGGTGGTTTTGATGTTTC
SOX2	F: TTGCTGCCTCTTTAAGACTAGGA
	R: CTGGGGCTCAAACTTCTCTC
AHNAK	F: TACCCTTCCTAAGGCTGACATT
	R: TTGGACCCTTGAGTTTTGCAT
miR-1251	F: ACACTCCAGCTGGG ACTCTAGCTGCCAAA
	R: CTCAACTGGTGTCGTGGAGT CGGCAATTCAGTTGAG AGCGCCTT
promoter region of miR-1251	F: TGGACAAGCTGAAGATATGGACA
	R: TGACCTCGATGGCAGTGATG

colon tissues were collected from patients who accepted radical operation of HSCR (age: 126.8 ± 14.31 days; gender: male 26, female 6) in Children's Hospital of Nanjing Medical University. In total, 32 controls matched with cases on age and gender were randomly picked out from isolated patients (age: 103.1 ± 10.47 days; gender: male 25, female 7) who underwent surgery for intussusception or incarcerated strangulated inguinal hernia (without enteric nervous malformation). The clinical features are also shown in **Table 1**. Tissues were harvested and stored at −80°C immediately after surgery. All HSCR patients were diagnosed through pathological analysis. Written informed consent from all participants was obtained.

Microarray Analysis

The microarray analysis was performed as our previous study described (18). Briefly, the Agilent human lncRNA Array v3.0 (4 × 180k format) was designed for profiling human lncRNAs. GeneSpring v12.0 software (Agilent, USA) was employed for the raw data summarization, normalization, and quality control. For the selection of dysregulated lncRNAs, the threshold value of ≥2- and <2-fold change and a Benjamini–Hochberg corrected *p*-value of 0.05 were adopted. Kangchen Bio-tech (Shanghai, China) conducted this microarray analysis.

Quantitative Real-Time PCR

To isolate total RNA from tissues and cells, Trizol reagent (Invitrogen Life Technologies Co, USA) was applied. qRT-PCR was employed to detect RMST, miR-1251, and AHNAK expression level by using SYBR (Takara Bio, Japan) reactions on Light Cycler 480 (Roche, Switzerland) according to the manufacturer's protocol. GAPDH and U6 were applied as an internal control for mRNA and miRNA detection, respectively. The expression quantity was analyzed with the $2^{-\Delta\Delta CT}$ method. Primer sequences are shown in **Table 2**.

Western Blotting

RIPA lysis buffer (Beyotime, Shanghai, China) was applied to extract total proteins from colon tissues and cultured cells. BCA Protein Assay Kit (Beyotime, Shanghai, China) was used to detect protein concentration. The same amount of total proteins were isolated in 10% SDS-PAGE, subsequently transferred to PVDF membranes, and then blocked with fat-free milk for 1 h. At 4°C, primary antibodies were used for incubation overnight. Afterwards, corresponding secondary antibodies were added for 2 h of incubation at room temperature. Finally, the membranes were imaged *via* ECL and Western blot detection reagents (Thermo Fisher Scientific, MA, USA). Primary antibodies including anti-AHNAK (1:1,000, SC134252), anti-SOX2 (1:1,000, SC17320X), and anti-GAPDH (1:1,000, SC47724) were obtained from Santa Cruz (CA, USA). The corresponding secondary antibodies were obtained from Beyotime (Shanghai, China).

Chromatin Immunoprecipitation

By using ChIP Assay Kit (Thermo Fisher Scientific, Shanghai, China), ChIP was implemented in accordance with the operating instructions. First, cross-linked chromatin was sonicated into around 200- to 1,000-bp fragments. Anti-SOX2 was used to immunoprecipitate the chromatin. Goat immunoglobulin G (IgG, ab172730) was employed to be the negative control. PCR was performed using SYBR Green Mix (Takara Bio, Japan). The primer sequences are shown in **Table 2**.

Cell Culture and Transfection

SH-SY5Y and 293T cell lines were acquired from ATCC. Cells were cultured at 37°C, 5% CO_2 condition using DMEM (Hyclone, USA) culture medium containing 10% FBS, 100 U/ml penicillin, and 100 μg/ml streptomycin. The inhibitor of miR-1251 (a chemically modified RNA single strand), siRNAs of RMST, SOX2, and AHNAK, and the corresponding negative controls were synthesized by Genechem (Shanghai, China). Transfection experiments were conducted by using Lipofectamine 2000 Reagent (Invitrogen, USA).

Cell Proliferation Assay

To test the cell viability, cell counting kit-8 (CCK-8; Dojindo, Japan) was employed. After transfection, cells were cultured in 96-well-plates for 24–48 h and subsequently incubated with

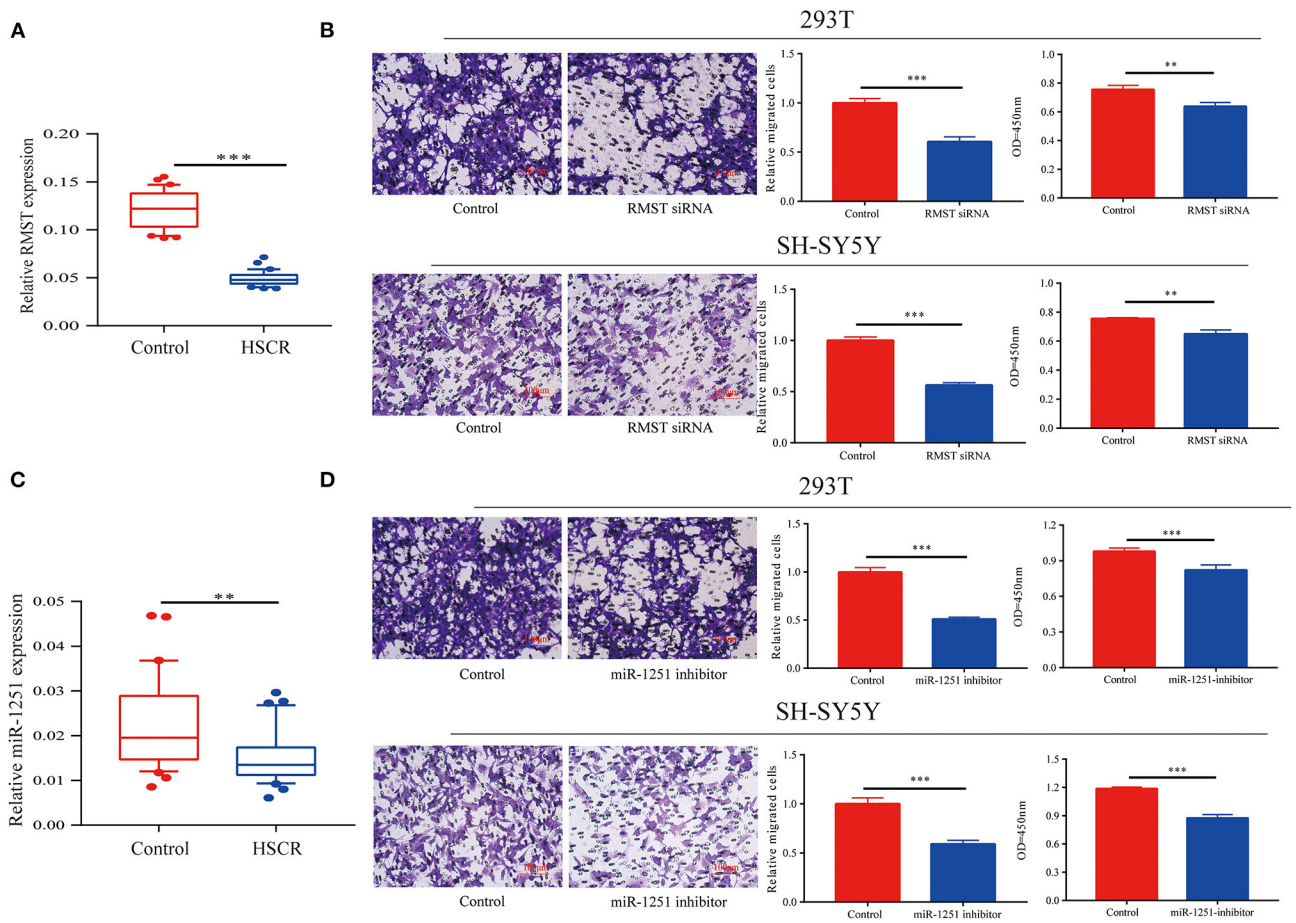


FIGURE 1 | RMST and miR-1251 were downregulated in HSCR patients. **(A)** RMST was downregulated obviously in aganglionic tracts of HSCR patients than control ones. **(B)** Transwell and CCK-8 assays showed that cell migration and proliferation was inhibited when RMST was knocked down in 293T and SH-SY5Y cells. **(C)** MiR-1251 was low expressed obviously in aganglionic colon tissues of HSCR patients compared with controls. **(D)** Cell proliferation and migration was attenuated obviously when miR-1251 was knocked down in 293T and SH-SY5Y cells. $^{**}p < 0.01$, $^{***}p < 0.001$.

CCK-8 reagent for 1–2 h. Eventually, the OD value at 450 nm was detected by the TECAN infinite M200 Multimode microplate reader (Tecan, Mechelen, Belgium). Each assay was conducted independently in triplicate.

Cell Migration Assay

Transwell chambers were placed above a 24-well-plate. After transfection around 24–48 h, cells were resuspended with serum-free medium to 1×10^6 cells/ml. About 1×10^5 cells were seeded to the upper chamber. Five hundred microliters of complete culture medium containing FBS was added to the lower chamber. Then 24–48 h later, 4% paraformaldehyde was applied to fix the lower chamber cells and then crystal violet staining solution was used to stain cells. Cells that migrated to the lower chamber were counted and imaged using an inverted microscope (×20, five fields were randomly selected for counting). All experiments were conducted in triplicate.

Dual-Luciferase Reporter Assay

pGL3-AHNAK-WT and pGL3-AHNAK-MUT were constructed by inserting the predicted 3′-UTR sequence of AHNAK binding to miR-1251 and the mutated sequence into the pGL3 promoter vector (Genechem, Shanghai, China). For reporter assay, cells were plated into 24-well-plates and transfected with 100 ng of pGL3-AHNAK-WT and pGL3-AHNAK-MUT, 50 nM miR-1251 mimics, and negative control using Lipofectamine 2000. Renilla luciferase vector pRL-SV40 (5 ng) was transfected into cells as control. Based on the obtained ratio, the activation degree of target reporter genes in different sample was compared.

Statistical Analysis

GraphPad Prism 7.0 (GraphPad Software, USA) was adopted for data analysis. Between two groups, *t*-test was applied to determine the statistically significant differences, while the comparison among multiple groups was performed *via* one-way ANOVA. All data were presented as the mean ± SEM. $P < 0.05$ was considered to be statistically significant.

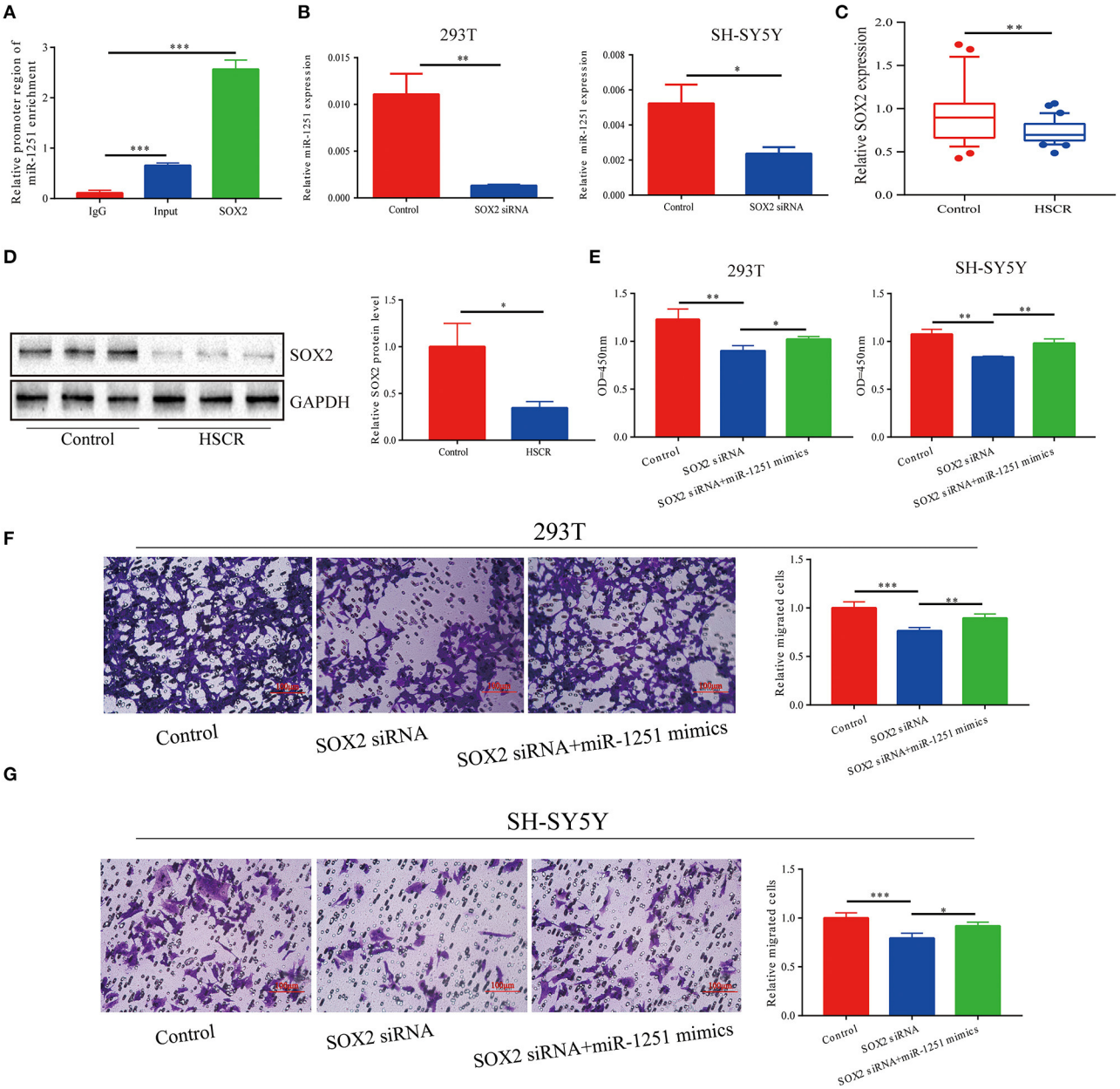


FIGURE 2 | miR-1251 was transcriptionally regulated by SOX2. **(A)** CHIP assay showed that SOX2 could bind to the promoter region of miR-1251. **(B)** After downregulating SOX2, the expression of miR-1251 was reduced. **(C,D)** SOX2 was downregulated in HSCR patients at both mRNA and protein level. **(E–G)** When 293T and SH-SY5Y cells were transfected with si-SOX2, cell migration, and proliferation was inhibited, but the upregulation of miR-1251 could partially reverse it. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

RESULTS

RMST and miR-1251 Were Downregulated in HSCR Patients

By qRT-CR, we found RMST was obviously downregulated in aganglionic colon segments compared with normal controls (**Figure 1A**). Receiver operating characteristic (ROC) curve analysis showed that RMST could serve as a molecular marker for the prognosis of HSCR (**Supplementary Figure 1G**). Transwell and CCK-8 assays showed that the knockdown of RMST inhibited both SH-SY5Y and 293T cells' proliferation and migration (**Figure 1B**). We also found the RMST intronic transcript miR-1251 was downregulated in aganglionic tracts (**Figure 1C**). When cells were transfected with miR-1251 inhibitor, the cell migration and proliferation was also attenuated significantly (**Figure 1D**).

miR-1251 Was Transcriptionally Regulated by SOX2

Since miR-1251 is transcribed from the same genomic locus as RMST, we suspected that RMST might regulate the expression level of miR-1251. However, there was no significant change on miR-1251 expression level after the downregulation of RMST, indicating RMST did not regulate miR-1251 directly (**Supplementary Figure 1B**). Moreover, SOX2 was predicted to bind with the 2-kbp upstream promoter region of miR-1251 using Promoter Scan (http://www.ncbi.nlm.nih.gov/Class/NAWBIS/Modules/DNA/dna21b.html) (**Supplementary Figure 1F**). The binding relationship was confirmed by the ChIP experiment (**Figure 2A**), and the silence of SOX2 decreased miR-1251 expression significantly (**Figure 2B**). In aganglionic colon segments, SOX2 was low expressed at both mRNA and protein levels compared with normal controls (**Figures 2C,D**). When cells were transfected with SOX2 siRNA, cell proliferation and migration was weakened obviously, while upregulating miR-1251 could reverse it partially (**Figures 2E–G**).

RMST Functioned as a Co-regulator of SOX2

The RIP assay revealed that RMST interacted with SOX2 protein, indicating RMST might function through SOX2 (**Figure 3A**). As shown in **Figure 3B**, miR-1251 was downregulated after cells were transfected with SOX2 siRNA and the expression of miR-1251 was much lower in cells co-transfected with RMST siRNA and SOX2 siRNA. CCK-8 and transwell assays have shown that the knockdown of both RMST and SOX2 attenuated cell proliferation and migration more obviously compared with just downregulating RMST or SOX2 alone. Meanwhile, the upregulation of miR-1251 partially reversed the combined suppressive effects of si-RMST and si-SOX2 (**Figures 3C,D**).

AHNAK Was the Target Gene of miR-1251

MiR-1251 was predicted to interact with the 3′-UTR region of AHNAK (**Figure 4A**). Compared with the control group, the luciferase activity was significantly decreased when cells were co-transfected with miR-1251 mimics and pGL3-AHNAK-WT plasmid (**Figure 4B**). In 293T cells, the knockdown of miR-1251 increased the expression of AHNAK (**Figure 4C**). In aganglionic tracts, AHNAK was overexpressed at both mRNA and protein levels (**Figures 4D,E**). As rescue experiment results have shown, the reduction of AHNAK could partially reverse the influence of miR-1251 inhibitor on cell migration and proliferation (**Figures 4F–H**).

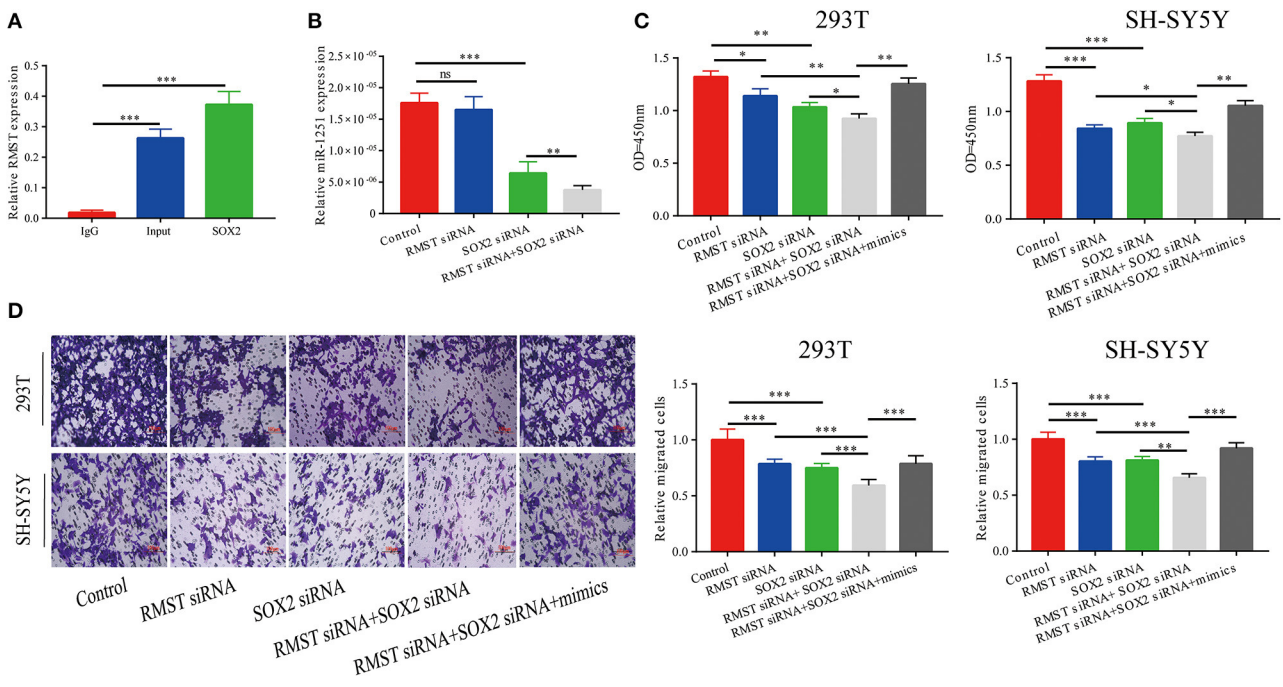


FIGURE 3 | RMST functioned as a co-regulator of SOX2. **(A)** RIP assay confirmed RMST could bind to SOX2 protein. **(B)** The expression of miR-1251 was downregulated in "SOX2 siRNA" group and was reduced more in "RMST siRNA+SOX2 siRNA" group, but was not changed significantly in "RMST siRNA" group compared with the control group. **(C,D)** CCK-8 and transwell assays revealed that when the expression of RMST and SOX2 were both knocked down, the cell proliferation and migration was more weakened than just downregulated RMST or SOX2 alone. However, the upregulation of miR-1251 could partly reverse the inhibitory effects of "RMST siRNA+SOX2 siRNA" on cell proliferation and migration. $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

RMST Functioned as a Transcriptional Co-regulator of SOX2 to Inhibit miR-1251 and Raise AHNAK Expression

Combined with the aforementioned results, we hypothesized that RMST enhanced the regulation of SOX2 on miR-1251 and then promoted the expression of AHNAK. The expression of AHNAK showed no significant difference between the RMST siRNA group and the control group. However, the expression of AHNAK was increased when SOX2 was knocked down (**Figures 5A,B**). Furthermore, the expression of AHNAK was much lower in "RMST siRNA+SOX2 siRNA" group than in "SOX2 siRNA" group, indicating that RMST upregulated the expression of AHNAK *via* acting as a SOX2 transcriptional co-regulatory factor (**Figures 5A,B**). Afterwards, the rescue experiments were set up, and we found the combined inhibitory effects of "RMST siRNA and SOX2 siRNA" on cell proliferation and migration could be partially alleviated by simultaneously downregulating the expression of AHNAK (**Figures 5C,D**).

DISCUSSION

Recently, more and more ncRNAs, especially miRNA and lncRNA, have been found to play critical roles in epigenetic regulation and take part in the occurrence and development of numerous diseases (26–28). Although there have been some reports about ncRNAs in HSCR, its mode of action and mechanism still require further studies (29–34). In this study, we detected that RMST and the intronic miR-1251 were downregulated in the aganglionic tracts of HSCR patients for the first time. The diagnostic value of RMST was also assessed *via* ROC. However, the number of samples was relatively not large enough and more population samples are needed for further validation. The knockdown of RMST inhibited cell proliferation

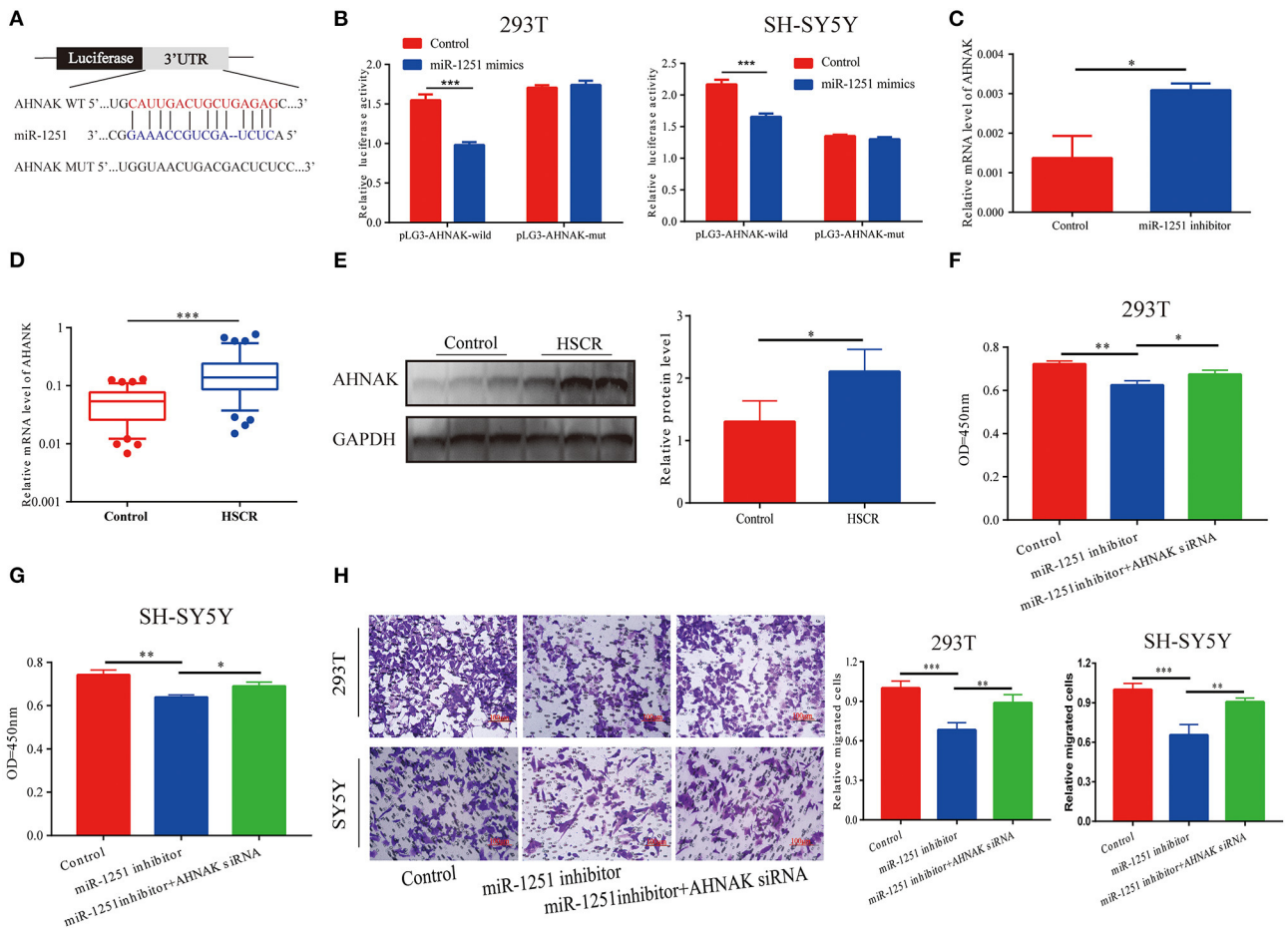


FIGURE 4 | AHNAK was the target gene of miR-1251. **(A)** The schematic diagram of binding sites between miR-1251 and AHNAK. **(B)** The luciferase activity was abated obviously when transfected with miR-1251 mimics and pLG3-AHNAK-wild compared with control; however, the luciferase activity was not changed significantly when transfected with miR-1251 mimics and pLG3-AHNAK-mut in 293T and SY5Y cells. **(C)** When miR-1251 was knocked down, the mRNA level of AHNAK was raised. **(D,E)** AHNAK was upregulated in stenotic tracts of HSCR patients at mRNA and protein level compared with control tracts. **(F–H)** The downregulation of miR-1251 inhibited cell migration and proliferation, but the knockdown of AHNAK could partially reverse it in 293T and SH-SY5Y cells. *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$.

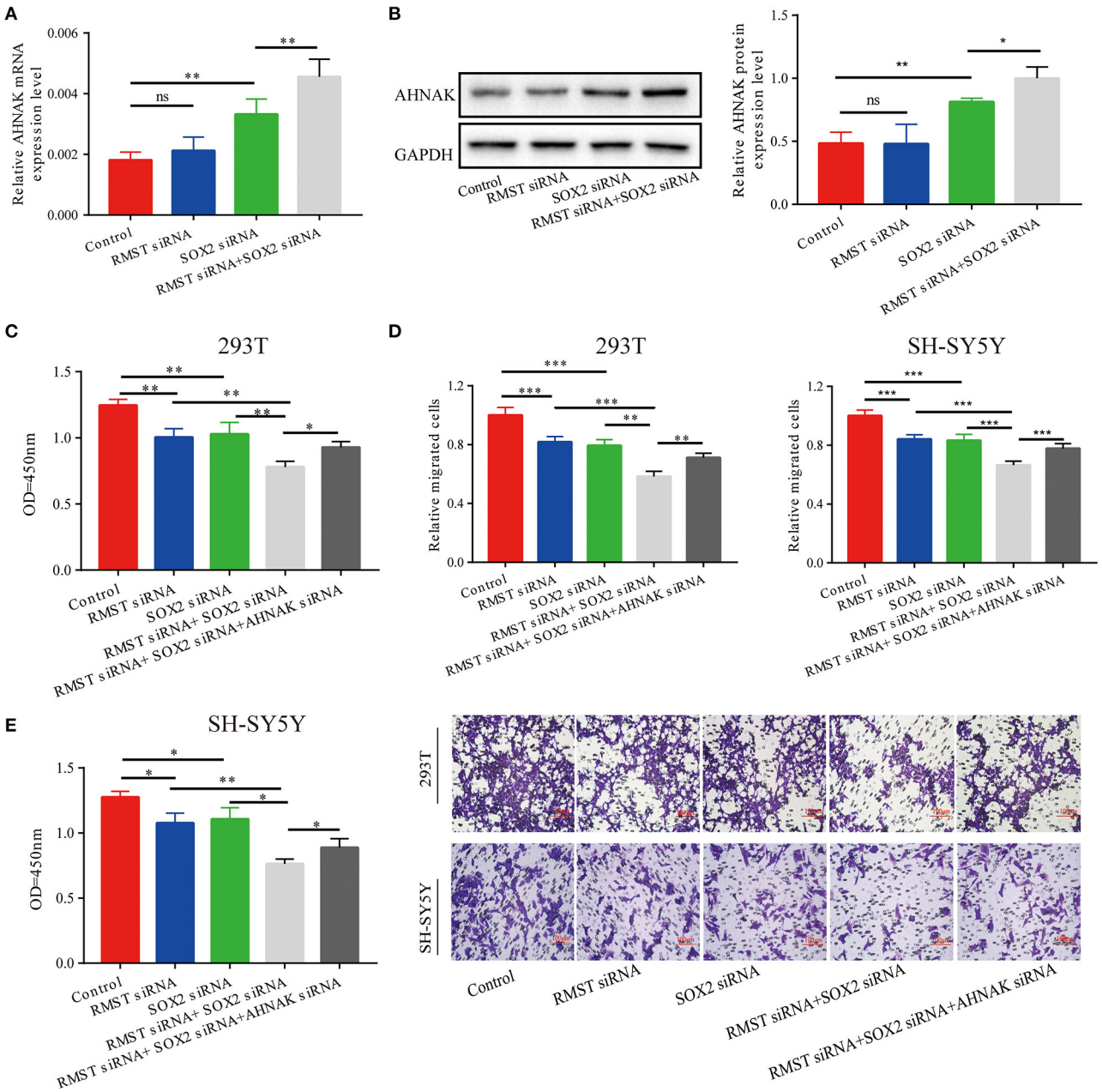


FIGURE 5 | RMST functioned as a SOX2 transcriptional co-regulator to inhibit miR-1251 and raise AHNAK expression. **(A,B)** When cells were transfected with RMST siRNA, the mRNA, and protein expression of AHNAK was not changed obviously; however, the transfection of SOX2 siRNA raised the expression of AHNAK and the downregulation of both RMST and SOX2 increased this more. **(C,D)** The cell proliferation and migration was inhibited in "RMST siRNA" or "SOX2 siRNA" groups compared with control and was inhibited more obviously in "RMST siRNA+SOX2 siRNA" group. **(E)** However, the simultaneous downregulation of AHNAK could partially alleviate the inhibitory effects caused by the knockdown of both RMST and SOX2. ns, $p \geq 0.05$, $^{*}p < 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

and migration obviously, indicating that RMST might play a key role in the progression of HSCR.

MiR-1251 was first demonstrated as a potential prognostic marker in head and neck squamous cell carcinoma (35). The role of miR-1251 in other diseases is still unclear. Herein, we found that the cell proliferation and migration was significantly inhibited by knocking down miR-1251, suggesting that miR-1251 might also be involved in the pathogenesis of HSCR through inhibiting the proliferation and migration of ENCCs. Although miR-1251 was derived from the intronic region of RMST, we found that RMST did not regulate the expression of miR-1251 directly. There might be other regulatory mechanisms.

According to the bioinformatical analysis, SOX2, a transcription factor, was predicted to bind to the promoter

region of miR-1251, and we performed ChIP assay to confirm this binding relationship. SOX family, such as SOX10, has been proved to be related to the pathogenesis of HSCR (36), but there are few reports about the role SOX2 exerts in HSCR. SOX2 has been reported to regulate the proliferation and differentiation of peripheral nerve cells in the peripheral nervous system (37). Moreover, SOX2 is closely related to the development of embryonic neural tube and neural crest cells (38–40). In the present study, we found that SOX2 was significantly downregulated in aganglionic tracts of HSCR patients. Furthermore, the downregulation of SOX2 significantly inhibited cell proliferation and migration, while the upregulation of miR-1251 could partially reverse it. These findings elucidated that SOX2 might be critical for the biological function of ENCCs by regulating miR-1251.

LncRNAs are known to bind with certain proteins to influence the function target proteins (41). Zhang et al. (42) found LINC00319 contributed to AML leukemogenesis *via* elevating SIRT6 through FUS-dependent pathway. LncRNA OCC-1 was verified to suppress cell growth through binding to and destabilizing HuR protein in colorectal cancer (43). RMST has been reported to bind to SOX2, and we also confirmed the relationship by using RIP assay in the present study, suggesting that RMST might function as a co-regulator of SOX2 in HSCR. We found that miR-1251 expression was reduced more when RMST and SOX2 were both knocked down compared with silencing SOX2 alone. Furthermore, the inhibition of cell proliferation and migration was more obviously when downregulating RMST and SOX2 than just decreasing RMST or SOX2. However, the enforced expression of miR-1251 partially reversed the combinative inhibitory effects, indicating that RMST functioned by acting as a transcriptional co-regulator of SOX2 to enhance the regulation of SOX2 on miR-1251.

MiRNAs generally exert their functions by degrading the target genes (44). We found that AHNAK, a kind of scaffold protein, was the target gene of miR-1251. The upregulation of AHNAK has been shown to impair cell proliferation and migration (45). In this study, we elucidated the upregulation of AHNAK in HSCR. Moreover, we found that AHNAK expression was not changed significantly when RMST was silenced. The knockdown of SOX2 increased the expression of AHNAK, and the downregulation of RMST strengthened this effect. Combined with the aforementioned, we revealed that RMST could regulate the expression of AHNAK and exerted its roles through RMST/SOX2/miR-1251/AHNKA axis.

Taken together, our study demonstrated that the downregulation of RMST inhibited cell proliferation and migration in HSCR. In terms of mechanism, RMST functioned as a transcriptional co-regulator of SOX2 to regulate the expression of AHNAK by strengthening the regulatory effect of SOX2 on miR-1251. The novel RMST/SOX2/miR-1251/AHNAK pathway might be helpful for the understanding of the pathogenesis of HSCR and the development of targeted therapy for HSCR.

However, this study still has some limitations. When the expression of RMST was decreased alone, the cell proliferation, and migration was also inhibited. Whether RMST has other regulatory patterns requires further investigation. In addition, *in vivo* experiments are also needed for further study.

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

This study was approved by the Institutional Ethics Committee of Children's Hospital of Nanjing Medical University (approval number: 201703057), and the experiments were conducted in accordance with the principles of the Declaration of Helsinki. All parents of patients had provided written informed consent in the study. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

HL and WT: designed the project. LZ, ZZ, PC, and BW: performed the experiments. LZ, CD, and XF: analyzed the data. ZZ and HL: wrote the paper. All authors discussed the results and commented on the article.

FUNDING

This work was supported by the Natural Science Foundation of China (NSFC 81701493) and General project of Nanjing Health and Family Planning Commission (YKK19103).

ACKNOWLEDGMENTS

We thank Dr. Jie Zhang, Xiaofeng Lv, Weiwei Jiang, Huan Chen, Wei Li, and Changgui Lu (Children's Hospital of Nanjing Medical University) for sample collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fped.2022.749107/full#supplementary-material

Supplementary Figure 1 | (A) The transfection efficiency of si-RMST in 293T and SH-SY5Y cells. **(B)** There was no significant change in miR-1251 expression in 293T and SH-SY5Y cells when transfected with si-RMST. The transfection efficiency of miR-1251 inhibitor **(C)**, si-AHNAK **(D)**, and oe-AHNAK **(E)** in 293T and SY5Y cells. **(F)** SOX2 was predicted to bind with the 2-kbp upstream promoter region of miR-1251. **(G)** The ROC analysis of RMST. ns, $p \geq 0.05$, $^{**}p < 0.01$, and $^{***}p < 0.001$.

Supplementary Figure 2 | The flowchart of bioinformatics analysis steps. The binding relationship between RMST and SOX2 was demonstrated by reference and RIP assay. Promoter Scan was applied to predict the binding relationship between SOX2 and miR-1251 promoter region. MicroRNA database was employed to predict the target gene of miR-1251.

REFERENCES

1. Sergi C. Hirschsprung's disease: historical notes and pathological diagnosis on the occasion of the 100(th) anniversary of Dr. Harald Hirschsprung's death. *World J Clin Pediatr.* (2015) 4:120–5. doi: 10.5409/wjcp.v4.i4.120
2. Jaroy EG, Acosta-Jimenez L, Hotta R, Goldstein AM, Emblem R, Klungland A, et al. "Too much guts and not enough brains": (epi)genetic mechanisms and future therapies of Hirschsprung disease - a review. *Clin Epigenet.* (2019) 11:135. doi: 10.1186/s13148-019-0718-x
3. Bergeron KF, Silversides DW, Pilon N. The developmental genetics of Hirschsprung's disease. *Clin Genet.* (2013) 83:15–22. doi: 10.1111/cge.12032
4. Wester T, Granstrom AL. Hirschsprung disease-Bowel function beyond childhood. *Semin Pediatr Surg.* (2017) 26:322–7. doi: 10.1053/j.sempedsurg.2017.09.008
5. Tam PK. Hirschsprung's disease: a bridge for science and surgery. *J Pediatr Surg.* (2016) 51:18–22. doi: 10.1016/j.jpedsurg.2015.10.021
6. Zhao J, Zhu Y, Xie X, Yao Y, Zhang J, Zhang R, et al. Pleiotropic effect of common PHOX2B variants in Hirschsprung disease and neuroblastoma. *Aging (Albany NY).* (2019) 11:1252–61. doi: 10.18632/aging.101834
7. Xu Z, Liu C, Zhao Q, Lu J, Ding X, Luo A, et al. Long non-coding RNA CCAT2 promotes oncogenesis in triple-negative breast cancer by regulating stemness of cancer cells. *Pharmacol Res.* (2020) 152:104628. doi: 10.1016/j.phrs.2020.104628
8. Shen S, Wang J, Zheng B, Tao Y, Li M, Wang Y, et al. LINC01714 enhances gemcitabine sensitivity by modulating FOXO3 phosphorylation in cholangiocarcinoma. *Mol Ther Nucleic Acids.* (2019) 19:446–57. doi: 10.1016/j.omtn.2019.11.028
9. Cao F, Wang Z, Feng Y, Zhu H, Yang M, Zhang S, et al. lncRNA TPTEP1 competitively sponges miR3285p to inhibit the proliferation of nonsmall cell lung cancer cells. *Oncol Rep.* (2020) 43:1606–18. doi: 10.3892/or.2020.7522
10. Zeng J, Li Y, Wang Y, Xie G, Feng Q, Yang Y, et al. lncRNA 00312 attenuates cell proliferation and invasion and promotes apoptosis in renal cell carcinoma *via* miR-34a-5p/ASS1 axis. *Oxid Med Cell Longev.* (2020) 2020:5737289. doi: 10.2139/ssrn.3473270
11. Sun C, Xu B, Wang L, Su Y. LncRNA DRAIC regulates cell proliferation and migration by affecting the miR-34a-5p/ITGA6 signal axis in Hirschsprung's disease. *Ups J Med Sci.* (2021) 126:e7895. doi: 10.48101/ujms.v126.7895
12. Pan W, Wu A, Yu H, Yu Q, Zheng B, Yang W, et al. Involvement of the lncRNA AFAP1-AS1/microRNA-195/E2F3 axis in proliferation and migration of enteric neural crest stem cells of Hirschsprung's disease. *Exp Physiol.* (2020) 105:1939–49. doi: 10.1113/EP088780
13. Su Y, Wen Z, Shen Q, Zhang H, Peng L, Chen G, et al. Long non-coding RNA LOC100507600 functions as a competitive endogenous RNA to regulate BMI1 expression by sponging miR128-1-3p in Hirschsprung's disease. *Cell Cycle.* (2018) 17:459–67. doi: 10.1080/15384101.2017.1403688
14. Cheng H, Sun M, Wang ZL, Wu Q, Yao J, Ren G, et al. LncRNA RMST-mediated miR-107 transcription promotes OGD-induced neuronal apoptosis *via* interacting with hnRNPK. *Neurochem Int.* (2020) 133:104644. doi: 10.1016/j.neuint.2019.104644
15. Tang W, Tang J, He J, Zhou Z, Qin Y, Qin J, et al. SLIT2/ROBO1-miR-218-1-RET/PLAG1: a new disease pathway involved in Hirschsprung's disease. *J Cell Mol Med.* (2015) 19:1197–207. doi: 10.1111/jcmm.12454
16. Ng SY, Johnson R, Stanton LW. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J.* (2012) 31:522–33. doi: 10.1038/emboj.2011.459
17. Sun XL, Wang ZL, Wu Q, Jin SQ, Yao J, Cheng H. LncRNA RMST activates TAK1-mediated NF-kappaB signaling and promotes activation of microglial cells *via* competitively binding with hnRNPK. *IUBMB Life.* (2019) 71:1785–93. doi: 10.1002/iub.2125
18. Shen Z, Du C, Zang R, Xie H, Lv W, Li H, et al. Microarray expression profiling of dysregulated long non-coding RNAs in Hirschsprung's disease reveals their potential role in molecular diagnosis. *Neurogastroenterol Motil.* (2016) 28:266–73. doi: 10.1111/nmo.12722
19. Collignon J, Sockanathan S, Hacker A, Cohen-Tannoudji M, Norris D, Rastan S, et al. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development.* (1996) 122:509–20. doi: 10.1242/dev.122.2.509
20. Schepers GE, Teasdale RD, Koopman P. Twenty pairs of sox: extent, homology, and nomenclature of the mouse and human sox transcription factor gene families. *Dev Cell.* (2002) 3:167–70. doi: 10.1016/S1534-5807(02)00223-X
21. Liu K, Xie F, Gao A, Zhang R, Zhang L, Xiao Z, et al. SOX2 regulates multiple malignant processes of breast cancer development through the SOX2/miR-181a-5p, miR-30e-5p/TUSC3 axis. *Mol Cancer.* (2017) 16:62. doi: 10.1186/s12943-017-0632-9
22. Alqadah A, Hsieh YW, Vidal B, Chang C, Hobert O, Chuang CF. Postmitotic diversification of olfactory neuron types is mediated by differential activities of the HMG-box transcription factor SOX-2. *EMBO J.* (2015) 34:2574–89. doi: 10.15252/embj.201592188
23. Schaefer T, Lengerke C. SOX2 protein biochemistry in stemness, reprogramming, and cancer: the PI3K/AKT/SOX2 axis and beyond. *Oncogene.* (2020) 39:278–92. doi: 10.1038/s41388-019-0997-x
24. Sannino G, Marchetto A, Ranft A, Jabar S, Zacherl C, Alba-Rubio R, et al. Gene expression and immunohistochemical analyses identify SOX2 as major risk factor for overall survival and relapse in Ewing sarcoma patients. *EBioMedicine.* (2019) 47:156–62. doi: 10.1016/j.ebiom.2019.08.002
25. Ng SY, Bogu GK, Soh BS, Stanton LW. The long noncoding RNA RMST interacts with SOX2 to regulate neurogenesis. *Mol Cell.* (2013) 51:349–59. doi: 10.1016/j.molcel.2013.07.017
26. Moradimotlagh A, Arefian E, Rezazadeh Valojerdi R, Ghaemi S, Jamshidi Adegani F, Soleimani M. MicroRNA-129 inhibits glioma cell growth by targeting CDK4, CDK6, and MDM2. *Mol Ther Nucleic Acids.* (2019) 19:759–64. doi: 10.1016/j.omtn.2019.11.033
27. Pan J, Fang S, Tian H, Zhou C, Zhao X, Tian H, et al. lncRNA JPX/miR-33a-5p/Twist1 axis regulates tumorigenesis and metastasis of lung cancer by activating Wnt/beta-catenin signaling. *Mol Cancer.* (2020) 19:9. doi: 10.1186/s12943-020-1133-9
28. Tai F, Gong K, Song K, He Y, Shi J. Enhanced JunD/RSK3 signalling due to loss of BRD4/FOXD3/miR-548d-3p axis determines BET inhibition resistance. *Nat Commun.* (2020) 11:258. doi: 10.1038/s41467-019-14083-4
29. Gunadi, Budi NYP, Kalim AS, Santiko W, Musthofa FD, Iskandar K, et al. Aberrant expressions of miRNA-206 target, FN1, in multifactorial Hirschsprung disease. *Orphanet J Rare Dis.* (2019) 14:5. doi: 10.1186/s13023-018-0973-5
30. Zhi Z, Zhu H, Lv X, Lu C, Li Y, Wu F, et al. IGF2-derived miR-483-3p associated with Hirschsprung's disease by targeting FHL1. *J Cell Mol Med.* (2018) 22:4913–21. doi: 10.1111/jcmm.13756
31. Wang G, Wang H, Zhang L, Guo F, Wu X, Liu Y. MiR-195-5p inhibits proliferation and invasion of nerve cells in Hirschsprung disease by targeting GFRA4. *Mol Cell Biochem.* (2021) 476:2061–73. doi: 10.1007/s11010-021-04055-y
32. Li Y, Zhou L, Lu C, Shen Q, Su Y, Zhi Z, et al. Long non-coding RNA FAL1 functions as a ceRNA to antagonize the effect of miR-637 on the down-regulation of AKT1 in Hirschsprung's disease. *Cell Prolif.* (2018) 51:e12489. doi: 10.1111/cpr.12489
33. Wen Z, Shen Q, Zhang H, Su Y, Zhu Z, Chen G, et al. Circular RNA CCDC66 targets DCX to regulate cell proliferation and migration by sponging miR-488-3p in Hirschsprung's disease. *J Cell Physiol.* (2019) 234:10576–87. doi: 10.1002/jcp.27733
34. Cai P, Li H, Huo W, Zhu H, Xu C, Zang R, et al. Aberrant expression of LncRNA-MIR31HG regulates cell migration and proliferation by affecting miR-31 and miR-31* in Hirschsprung's disease. *J Cell Biochem.* (2018) 119:8195–203. doi: 10.1002/jcb.26830
35. Hui L, Wu H, Yang N, Guo X, Jang X. Identification of prognostic microRNA candidates for head and neck squamous cell carcinoma. *Oncol Rep.* (2016) 35:3321–30. doi: 10.3892/or.2016.4698
36. Southard-Smith EM, Kos L, Pavan WJ. Sox10 mutation disrupts neural crest development in Dom Hirschsprung mouse model. *Nat Genet.* (1998) 18:60–4. doi: 10.1038/ng0198-60
37. Wakamatsu Y, Endo Y, Osumi N, Weston JA. Multiple roles of Sox2, an HMG-box transcription factor in avian neural crest development. *Dev Dyn.* (2004) 229:74–86. doi: 10.1002/dvdy.10498
38. Iida H, Furukawa Y, Teramoto M, Suzuki H, Takemoto T, Uchikawa M, et al. Sox2 gene regulation *via* the D1 enhancer in embryonic neural tube and

neural crest by the combined action of SOX2 and ZIC2. *Genes Cells.* (2020) 25:242–56. doi: 10.1111/gtc.12753
39. Cimadamore F, Fishwick K, Giusto E, Gnedeva K, Cattarossi G, Miller A, et al. Human ESC-derived neural crest model reveals a key role for SOX2 in sensory neurogenesis. *Cell Stem Cell.* (2011) 8:538–51. doi: 10.1016/j.stem.2011.03.011
40. Amador-Arjona A, Cimadamore F, Huang CT, Wright R, Lewis S, Gage FH, et al. SOX2 primes the epigenetic landscape in neural precursors enabling proper gene activation during hippocampal neurogenesis. *Proc Natl Acad Sci U S A.* (2015) 112:E1936–E45. doi: 10.1073/pnas.1421480112
41. Wang Z, Yang B, Zhang M, Guo W, Wu Z, Wang Y, et al. lncRNA epigenetic landscape analysis identifies EPIC1 as an oncogenic lncRNA that interacts with MYC and promotes cell-cycle progression in cancer. *Cancer Cell.* (2018) 33:706–20.e9. doi: 10.1016/j.ccell.2018.03.006
42. Zhang Y, Huang Z, Sheng F, Yin Z. MYC upregulated LINC00319 promotes human acute myeloid leukemia (AML) cells growth through stabilizing SIRT6. *Biochem Biophys Res Commun.* (2019) 509:314–21. doi: 10.1016/j.bbrc.2018.12.133
43. Lan Y, Xiao X, He Z, Luo Y, Wu C, Li L, et al. Long noncoding RNA OCC-1 suppresses cell growth through destabilizing HuR protein in colorectal cancer. *Nucleic Acids Res.* (2018) 46:5809–21. doi: 10.1093/nar/gky214
44. Chopra N, Wang R, Maloney B, Nho K, Beck JS, Pourshafie N,et al. MicroRNA-298 reduces levels of human amyloid-beta precursor protein (APP), beta-site APP-converting enzyme 1 (BACE1) and specific tau protein moieties. *Mol Psychiatry.* (2021) 26:5636–57. doi: 10.1038/s41380-019-0610-2
45. Chen B, Wang J, Dai D, Zhou Q, Guo X, Tian Z, et al. AHNAK suppresses tumour proliferation and invasion by targeting multiple pathways in triple-negative breast cancer. *J Exp Clin Cancer Res.* (2017) 36:65. doi: 10.1186/s13046-017-0522-4

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: 10 March 2022
doi: 10.3389/fgene.2022.821587

Novel Biallelic Variant in the *BRAT1* Gene Caused Nonprogressive Cerebellar Ataxia Syndrome

Yiming Qi[1,2†], *Xueqi Ji*[1,3†], *Hongke Ding*[1,2], *Ling Liu*[1,2], *Yan Zhang*[1,2] *and Aihua Yin*[1,2,3*]

[1]*Prenatal Diagnosis Center, Guangdong Women and Children Hospital, Guangzhou, China,* [2]*Maternal and Children Metabolic-Genetic Key Laboratory, Guangdong Women and Children Hospital, Guangzhou, China,* [3]*Clinical Medicine College, Guangzhou Medical University, Guangzhou, China*

Recessive mutations in *BRAT1* cause lethal neonatal rigidity and multifocal seizure syndrome (RMFSL), a phenotype characterized by neonatal microcephaly, hypertonia, and refractory epilepsy with premature death. Recently, attenuated disease variants have been described, suggesting that a wider clinical spectrum of *BRAT1*-associated neurodegeneration exists than was previously thought. Here, we reported a 10-year-old girl with severe intellectual disability, rigidity, ataxia or dyspraxia, and cerebellar atrophy on brain MRI; two *BRAT1* variants in the *trans* configuration [c.1014A > C (p.Pro338 =); c.706delC (p.Leu236Cysfs*5)] were detected using whole-exome sequencing. RNA-seq confirmed significantly decreased *BRAT1* transcript levels in the presence of the variant; further, it revealed an intron retention between exon 7 and exon 8 caused by the synonymous base substitute. Subsequent prenatal diagnosis for these two variants guided the parents to reproduce. We expand the phenotypic spectrum of *BRAT1*-associated disorders by first reporting the pathogenic synonymous variant of the *BRAT1* gene, resulting in clinical severity that is mild compared to the severe phenotype seen in RMFSL. Making an accurate diagnosis and prognostic evaluation of *BRAT1*-associated neurodegeneration is important for reproductive consultation and disease management.

Keywords: *BRAT1*, nonprogressive cerebellar ataxia syndrome, synonymous variant, intron retention, prenatal diagnosis

OPEN ACCESS

Edited by:
Maria Elisabetta Baldassarre, University of Bari Aldo Moro, Italy

Reviewed by:
Saud Alsahli, King Faisal Specialist Hospital and Research Centre, Saudi Arabia
Christine Tranchant, Université de Strasbourg, France

***Correspondence:**
Aihua Yin
yinaiwa@vip.126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *24 November 2021*
Accepted: *07 January 2022*
Published: *10 March 2022*

Citation:
Qi Y, Ji X, Ding H, Liu L, Zhang Y and Yin A (2022) Novel Biallelic Variant in the BRAT1 *Gene Caused Nonprogressive Cerebellar Ataxia Syndrome. Front. Genet. 13:821587. doi: 10.3389/fgene.2022.821587*

INTRODUCTION

Variations in *BRAT1* (BRCA1-associated ataxia telangiectasia mutated activator 1) are initially recognized as the cause of lethal neonatal rigidity and multifocal seizure syndrome (RMFSL; OMIM#614498), which is characterized by neonatal microcephaly, intractable seizures, hypertonia, and early demise. Subsequently, neurodevelopmental disorder with cerebellar atrophy and with or without seizures (NEDCAS; OMIM#618056) caused by biallelic *BRAT1* variants were reported and redefined the description of "lethality." Recently, a milder clinical form that manifests as nonprogressive cerebellar ataxia (NPCA) was described in some childhood-onset patients (Mahjoub et al., 2019), suggesting that a wider phenotypic spectrum of *BRAT1*-associated neurodegeneration exists than was previously thought.

Physiological functions of the disease-causing gene *BRAT1* are diverse (Fernandez-Jaen and Ouchi, 2016). It encodes a protein that interacts with the tumor suppressor gene BRCA1 at its

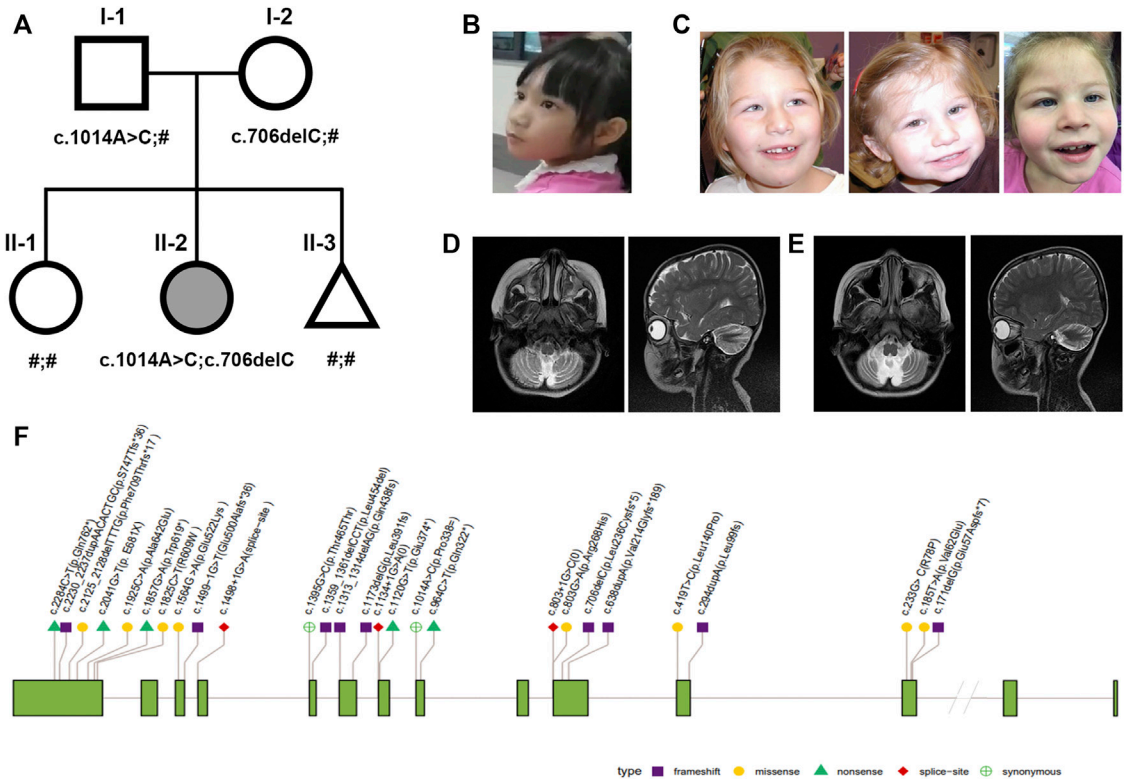


FIGURE 1 | Nonprogressive cerebellar ataxia syndrome phenotype and pedigrees. **(A)** Pedigrees of the family with NPCA profiled in this study. The proband is marked in black. **(B–C)** Facial appearance of the proband is not specific (left); typical facial dysmorphisms in reported BRAT1-associated NPCA including epicanthal folds, a high arched palate, etc [right, Ref (Srivastava et al., 2016)]. **(D–E)** Brain MRI of the proband at the age of 4 showed typical cerebellar atrophy (left), which almost showed no progress at 9 years (right). **(F)** Distribution of reported pathogenic variants in *BRAT1*. Annotations include the following: exon (grass–green square); variation types: frameshift (purple square), missense (yellow circle), nonsense (green triangle), red diamond (splice site), and synonymous (green cross circle).

C-terminus and binds to ATM1, considering a master controller of the cell cycle signaling pathways required for cellular responses to DNA damage (Ouchi and Ouchi, 2010). *BRAT1* can form a complex with an ATPase domain-containing protein, BRP1 (*BRAT1* Partner 1), and prevent transcriptional silencing at methylated genomic regions (Zhang et al., 2016). It may also be involved in cell growth and apoptosis (Straussberg et al., 2015). *BRAT1* deficiency secondary to biallelic *BRAT1* mutations may increase the glucose metabolism, reduce the mitochondrial reactive oxygen species (ROS) concentration, deteriorate cell growth and migration, and induce neuronal atrophy (Puffenberger et al., 2012; Saunders et al., 2012; So and Ouchi, 2014; Wolf et al., 2015). The complexity and extensiveness of the *BRAT1* gene function are the biological basis for the huge phenotypic heterogeneity. However, the exact mechanism by which variations in *BRAT1* trigger neurodegeneration and to what extent a defect in ATM function contributes to this disease are unknown.

In this study, we first reported the clinical course of a proband with NPCA caused by novel compound heterozygous *BRAT1* variants, which include a negligible pathogenic synonymous variant. Functional studies confirmed the effect of them on transcription. The result provided a theoretical basis and guidance for this family in reproductive genetic counseling and prenatal diagnosis. Furthermore, we summarized all published cases with *BRAT1* variation, which provide insight into the clinical–genetic correlation and the pathophysiology of the disease.

Clinical Report

A 10-year-old girl who presented severe intellectual disability and poor motor ability was transferred to our genetics center for consultation due to the family's reproductive plan.

The patient was the second child of nonconsanguineous parents of Chinese and had a healthy adolescent sister (pedigree in **Figure 1A**). There were unremarkable findings during her prenatal, perinatal, and neonatal courses. Physical examination at birth revealed a normal height, normal weight, and normal head 101 circumference (51 cm; 48th percentile).

However, global developmental delay was presented initially a few months later. She developed head control at 6 months and could not sit until 15 months. At 2.5 years, she could only babble, make consonant sounds, and communicate her needs by crying and gazing. After an individualized neurorehabilitation therapy, she was able to sit briefly and pull to stand at 3 years. One year later, she could stand independently for a few minutes and walk with a walker. At 5 years, she could say 5–10 word phrases that

were dysarthric, identify people who were in constant contact with her, and follow simple commands. Meanwhile, behavior with episodes of impulsivity and irritability began to appear since then. She attended special schooling for rehabilitation training but with poor performance until last re-evaluation and was considered to have severe intellectual disability with a mental age of less than 3 years.

Until now, she could stand alone and walk slowly with limited support in a rigid- and broad-based gait with flexed arm posture (**Supplementary Videos S1,2**). She also presented slight dysmetria, which interferes in fine motor skills during the performance of tasks (**Supplementary Video S3**). Although her facial expression shows relative paucity (**Figure 1B**), no evident *BRAT1*-assoicated dysmorphisms, such as epicanthal folds, high arched palate, fifth finger clinodactyly, or single palmar crease, were observed (**Figure 1C**). Her cranial nerves were intact except pendular nystagmus. Motor exam has shown mild hypertonia and resistance during extension, but tendon reflexes were normal.

Cranial magnetic resonance imaging (MRI) spectroscopy was first performed at the age of 4, which recorded cerebellar atrophy (**Figure 1D**). In the following years, she showed stable cerebellar atrophy on serial neuro-MRI (**Figure 1E**). Electroencephalography was uniformly negative.

Laboratory examination included serum tests for alpha-fetoprotein (AFP), inborn errors of the metabolism, amino acids, very long-chain fatty acids, acylcarnitine and carnitine profiles, infectious work-up, TORCH (toxoplasmosis, rubella, cytomegalovirus, and herpes simplex virus) tests, urine organic acid analyses, and conventional indicators in cerebrospinal fluid (CSF) analysis, which were all within normal limits. We also ruled out the possibility of serine biosynthesis defects and nonketotic hyperglycinemia by normal serine and glycine. After precluding aneuploidy and copy number variations (CNVs), we conducted a Trio WES on the family.

MATERIALS AND METHODS

Genetic Investigations

Genomic DNA was extracted using a Qiagen DNA blood mini kit (Qiagen GmbH, Hilden, Germany). Library preparation and target enrichment were performed using a SureSelectXT Clinical Research Exome kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's specifications. Then, Trio WES was performed using 2 × 150 bp in the paired end mode of the NextSeq 500 platform (Illumina, San Diego, CA) to obtain an average coverage of above 110x, with 97.6% of target bases covered at least 10x. Sequence quality analysis and filtering of mapped target sequences were performed with the 'varbank' exome and genome analysis pipeline v.2.1 as described previously (Vetro et al., 2020). Analysis of genetic results was based on the genomic variation database (http://dgv.tcag.ca/dgv/app/home), DECIPHER database (https://decipher.sanger.ac.uk/), and OMIM database (http://www.ncbi.nlm.nih.gov/omim). The found variants were further verified by Sanger sequencing in fetuses and parents.

RNA Sequence

Total RNA was extracted from peripheral blood samples using Qiagen blood RNA extraction kit I (QIAGEN, United States); the procedures and standards were performed according to the manual. After complete property control of RNA and quality control of RNA concentration and purity are qualified, 1 μg of RNA is aspirated for mRNA library construction. The mRNA library was constructed using a TIANSeq Fast RNA Library Kit (Illumina, United States) according to the manufacturer's instructions, where mRNA was purified and enriched from 1 μg of the total RNA samples and then fragmented about 250 bp, and the index adapter was added. Finally, using a high-fidelity enzyme amplifies the library. After quality control, the libraries were sequenced on an Illumina HiSeq 4,000 platform.

Chorionic Villus Sampling

Fetal samples were collected by chorionic villus sampling at 13 weeks of gestation. The procedure was performed using the transabdominal approach. Under aseptic conditions, an 18 or 20 gauge spinal needle was inserted into the placenta under continuous ultrasound guidance. A 20 cc syringe containing the collection media is attached to the end of the needle once the stylet is removed. Negative pressure is created, and the needle is moved up and down through the placenta, collecting the tissue (Jones and Montero, 2021).

RESULTS

Genetic Findings

Whole-exome sequencing analysis identified compound heterozygous mutations in the *BRAT1* gene, c.706delC (p.Leu236Cysfs*5), and a synonymous variant, c.1014A > C (p.Pro338 =), which were confirmed by sanger sequencing and segregated with the disorder in her family (**Figure 2A**).

Neither both the variants were reported in public databases nor their functional impact was examined. The deletion variant c.706delC (p.Leu236Cysfs*5) in exon 5 was predicted to lead to a frameshift and to cause loss of the full-length protein (821 amino acids) due to truncation after the first 236 residues. It is predicted to be pathogenic (VarSome, https://varsome.com/ and ClinVar https://www.ncbi.nlm.nih.gov/clinvar/).

The other variant, c.1014A > C (p.Pro338 =) in exon 7/8, is not conserved among species, and the amino acid pro at the position 338 of *BRAT1* protein is not changed as well; however, varSEAK analysis predicts that the mutation may cause the classic splicing site c.1015 + 1 to be skipped (**Figure 2B**). Thus, RNA-seq was exerted to identify potential splicing defects associated with the variants in the NPCA case.

BRAT1 expression levels in RNA-seq showed a trend toward lower expression in heterozygous parents and the compound heterozygous proband and confirmed significantly decreased *BRAT1* transcript levels in the presence of the variant (**Figure 2C**), consistent with NMD of the mutant transcript. Meanwhile, *BRAT1* expression was higher in the synonymous variants' father than in the frameshift variants' mother, although the difference was not statistically significant (**Figure 2C**).

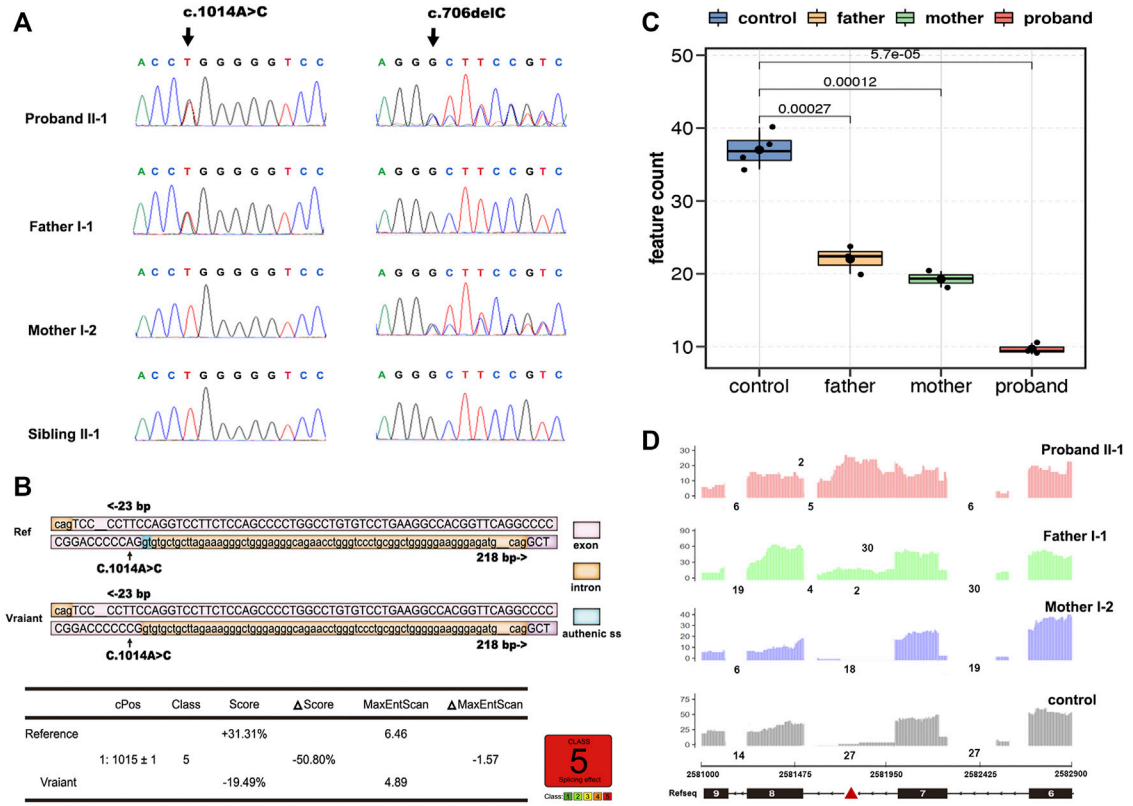


FIGURE 2 | Overview of genetic testing results of BRAT1. Sanger sequencing of BRAT1 variants. **(A)** Chromatograms demonstrated the compound heterozygous status of c.1014A > C and c.706delC in the BRAT1 gene in the proband, the proband's father, and the proband's mother. The position of the variant or the corresponding wild-type nucleotide is labeled with black arrows. **(B)** Software predictions strongly suggest that the substitution of c.1014A > C might cause aberrant splicing. **(C)** Groups were significantly different from each other (controls vs. proband, $p < 5.7*10^{-5}$; controls vs. father, $p < 2.7*10^{-4}$; controls vs. mother, $p < 1.2*10^{-4}$; and two-tailed unpaired *t*-test). **(D)** RNA-seq read coverage shows aberrant retention of intron 7 of BRAT1 in the proband and in the heterozygous (c.1014A > C) father relative to wild-type controls.

Furthermore, RNA-seq analyses identified significantly increased retention of intron 7 of *BRAT1* in the proband and the heterozygous father relative to the heterozygous mother and wild-type control samples (**Figure 2D**). It was presumed to be the effect of NM_152743.3 c.1014A > C (p.Pro338 =) mutation. Altogether, these results suggest that both the variants participate in the pathogenesis of NPCA.

Prenatal Diagnosis for Reproduction and Follow-Up

The mother of the proband underwent prenatal diagnosis at 12 weeks of gestation during her third pregnancy using chorionic villus sampling. Fortunately, Sanger sequencing showed that the fetus inherited neither of the above sites (**Figure 1A**). Prenatal and neonatal courses of the fetus were uncomplicated. A male neonate 2,560 g in weight and 51 cm high was delivered at the term, without microcephaly (HC 348 mm). In the following years, his head circumference has grown consistently within the 95th centile. Other growth parameters, strength, reflexes, and sensation were normal until the last visit at 18 months.

DISCUSSION

A total of 30 patients with clinical manifestations ranging from RMFSL to NPCA had been identified as homozygous or compound heterozygous variants in *BRAT1* (**Table 1**, **Figure 1F**). In *BRAT1*-related disorders, cerebellar hypoplasia seemed not congenital (Wolf et al., 2015; Srivastava et al., 2016; Celik et al., 2017). Clinical classifications depend on the rate of progression of atrophy after birth determined. Nearly 75% of them present with severe RMFSL, with almost all accompanied by multifocal or refractory epilepsy or even intrauterine jerks (Celik et al., 2017). Epilepsy occurs in 41.6% of NEDCAS/NPCA patients, mostly before 3 years of age. It had been reported that a few NPCA cases, which mapped to the SCA15 locus on chromosome 3pter, might develop spasticity or focal dystonia with increasing age (Dudding et al., 2004). However, in NPCA caused by BRAT1, including the case here, later-onset spasticity is extremely rare on long-term follow-up.

Divergence is much huger for the clinical–genetic correlation. It has been mentioned that the phenotypic spectrum of BRAT1-associated disorders is associated with the domain, localization,

TABLE 1 | Clinical features of individuals with previously described BRAT1 variants besides the present study.

NO.	Gender	Variant	Effect	Pedigree	Reported condition	Associated clinical phenotypes	Electroencephalogram	Brain MRI	Ref
1	M	c.185T > A	p.Val62Glu	Sibling wih NO.2	24 years' old, graduated from college	Mild intellectual disability, ataxia, motor development delay, language delay, gaze-evoked nystagmus	Not offered	9 years: stable isolated cerebellar atrophy	Mahjoub et al. (2019)
		c.185T > A						16 years: stable isolated cerebellar atrophy	
2	M	c.185T > A	p.Val62Glu	Sibling wih NO.1	7 years' old	Mild intellectual disability, ataxic, motor development delay, dysarthria, gaze-evoked nystagmus	Not offered	3 years: stable isolated cerebellar atrophy	Mahjoub et al. (2019)
		c.185T > A						5 years: stable isolated cerebellar atrophy	
3	F	c.638dupA	p.Val214Glyfs[a]189	Sibling wih NO.4	10 years' old	Hypotonia, microcephaly, dysmetria and truncal titubation, ataxic, intellectual disability, ataxia, and cerebellar atrophy, head circumference, global developmental delay, bilatera l5th finger clinodactyly	Not offered	Prominent cerebellar interfolial spaces, which remained unchanged from 2, 3 years	Srivastava et al. (2016)
		c.803+1G > C	Splice site						
4	F	c.638dupA	p.Val214Glyfs[a]189	Sibling wih NO.3	6 years' old	Hypotonic, dysmetria, microcephaly, dysarthric speech and pendular nystagmus, global developmental delay, activity-induced tremor, bilateral 5th finger clinodactyly	Normal	Progressive enlargement of the cerebellar interfolial spaces, cerebellar atrophy	Srivastava et al. (2016)
		c.803+1G > C	Splice site						
5	F	c.294dupA	p.Leu99Thrfs[a]92	Sporadic	6 years' old	Seizures, hypertonia, microcephalic, generalized axial and peripheral hypertonia and hyper-reflexia, motor development delay, language delay	Mainly left-sided temporo-occipital epileptiform discharges and absence of a posterior dominant rhythm	3 mo: decreased myelination and thin corpus callosum	Mundy et al. (2016)
		c.1925C > A	p.Ala642Glu					3 years: right temporal lobe encephalomalacia and cerebellar and vermis hypoplasia	

(Continued on following page)

TABLE 1 | (*Continued*) Clinical features of individuals with previously described BRAT1 variants besides the present study.

NO.	Gender	Variant	Effect	Pedigree	Reported condition	Associated clinical phenotypes	Electroencephalogram	Brain MRI	Ref
6	M	c.1564G > A c.638dupA	p.Glu522Lys p.Val214Glyfs[a]189	Sporadic	4.5 years' old	Microcephaly, hypertonia, progressive encephalopathy never presented seizures	Normal	19 and 48 mo: moderate progressive cerebellar atrophy	Fernández-Jaén et al. (2016)
7	F	c.638dupA c.419T > C	p.Val214Glyfs[a]189 p.Leu140Pro	Sporadic	4 years 4 mo old	Right esotropia, mild optic nerve hypoplasia, with decreased visual acuity bilaterally, moderate appendicular rigidity, dyspraxia, global developmental delay, bilateral5th finger clinodactyly	Showed frequent 3–4 Hz generalized spike and wave complexes (without clinical correlate)	5 mo: normal 21 mo and 4 years 3 mo: enlargement of the cerebellar interfolial spaces compatible with cerebellar atrophy and mildly delayed myelination	Srivastava et al. (2016)
8	F	c.1857G > A c.2125_2128delTTTG	p.Trp619[a] p.Phe709Thrfs[a]17	Sibling wih NO.15	4 years and 4 mo	Drug-resistan seizures, microcephaly, developmental delay	Multifocal epileptiform activity	Not offered	Smith et al. (2016)
9	F	c.294dupA c.1825C > T	p.Leu99Thrfs[a]92	Sporadic	3 years and 8 mo	Seizures, microcephaly, difficulty swallowing, visual impairment, nystagmus, ataxia, and frequent episodes of autonomic dysregulation axial hypotonia, appendicular hypertonia, global developmental delay, motor development delay	Normal	3.5 years: progressive cerebellar and brainstem atrophy	Hanes et al. (2015)
10	F	c.294dupA c.803G > A	p.Arg609Trp p.Leu99Thrfs[a]92	Sporadic	20 mo of age	Febrile seizures, hypertonia, nystagmus, esotropia, arrested head growth P10, motor development delay, developmentally delayed	Not offered	Not offered	Oatts et al. (2017)

(Continued on following page)

TABLE 1 | (*Continued*) Clinical features of individuals with previously described BRAT1 variants besides the present study.

NO.	Gender	Variant	Effect	Pedigree	Reported condition	Associated clinical phenotypes	Electroencephalogram	Brain MRI	Ref
11	M	c.171delG	p.Arg268His p.Glu57Aspfs[a]7	Sporadic	15 mo old	Seizures, hypertonia, microcephaly, axial hypotonia and symmetric hypertonia, intermittent asymptomatic bradycardia and hypothermia, nonepileptic apnea, chronic lungdiseas, dry skin	Episodes of focal electrographic status epilepticus	One d of life: normal structures but subtle nonspecific foci of the cerebral white matter	Srivastava et al. (2016)
		c.419T > C	p.Leu140Pro					4.5 mo: mild global cerebral volume loss with prominence of the sulci and secondary enlargement of the lateral ventricle, normal cerebellar structures	
12	M	c.638_639insA	p.Val214Glyfs[a]189	Sporadic	Died at the age of 5 years and 9 mo due to respiratory insufficiency	Early onset epileptic encephalopathy postnatal, microcephaly, apnea, feeding problems, bradycardia, global developmental delay, maldescensus testis, left-sided club foot, and left-sided pes adductus	Focal continuous spike discharges in the right more than in the left occipital region	Thin corpus callosum, dilated internal and external cerebrospinal fluid spaces, and delayed myelination	Horn et al. (2016)
13	F	c.1134+1G > A c.1498+1G > A	Splice site Splice site	Sporadic	Died at 4 years 3 mo	Microcephaly, hypertonia, focal, multifocal motor seizures with clonic features, apnea, eye deviation to either side, clustering on awakening and drowsing, epileptic spasms, and tonic seizures	Multifocal epileptiform dischargesIctal: migrating focal seizures; seizures arising from right central region, vertex, left central, left occipital, right temporal, and left temporal region; epileptic spasms and periodic spasms, hypsarrhythmia	One m 12 d: very small hemosiderin deposition within lateral ventricles and subarachnoid spaces from previous IVH7.5 m: prominent ventricles and extra-axial CSF spaces with associated white matter volume loss, nonspecific abnormal white matter signal	Scheffer et al. (2020)

(Continued on following page)

TABLE 1 | (*Continued*) Clinical features of individuals with previously described BRAT1 variants besides the present study.

NO.	Gender	Variant	Effect	Pedigree	Reported condition	Associated clinical phenotypes	Electroencephalogram	Brain MRI	Ref
14	F	c.1498+1G > A c.638dupA	p.Val214Glyfs[a]189	Sibling[a] with NO. 25	Died at the age of 17 mo because of respiratory failure	Epileptic seizures (eye blinking and myoclonus left hand) and hypertonia, microcephaly	Continuous abnormal background pattern and multifocal seizure activity	Two mo: normal 12 mo: severe generalized atrophy, hardly any myelination	Wolf et al. (2015)
15	M	c.638dupA c.1857G > A	p.Trp619[a]	Sibling with NO.8	Died at 15 mo of age	Drug-resistan seizures, microcephaly, and developmental delay	Multifocal epileptiform activity	13 mo: global cerebral and cerebellar atrophy	Smith et al. (2016)
16	F	c.2125_2128delTTTG c.1359_1361delCCT	p.Phe709Thrfs[a]17 p.Leu454del	Sporadic	Died at 14 mo	Microcephaly, hypertonia, focal motor clonic seizures, migrating between hemispheres	Multifocal epileptiform discharges, discontinuous backgroundIctal: migrating focal seizures from one region to another, most frequent onset from the right posterior quadrant, other onsets in the left posterior region and left frontocentral region	Three d: small right occipital subdural hemorrhage18 d: hemorrhage resolved	Scheffer et al. (2020)
17	M	c.1395G > C c.1313_1314delAG	p.Thr465Thr p.Gln438fs	Sibling[a]	Died at the age of 12 mo	Polymorphic seizures and hypertonia, microcephaly	Generalized and focal sharp and spike waves	The myelination pattern was appropriate for the patient's age, subarachnoid space was slightly widened	Szymańska et al. (2018)
18	F	c.1313_1314delAG c.964C > T	p.Gln322[a]	Sibling with NO.27	Died at 10 mo	Microcephaly, hypertonia, focal clonic seizures with apnea, tachycardia	Multifocal epileptiform discharges, discontinuous background intermittentlyIctal: migrating focal seizures; central, right occipital spread to the left occipital region, left temporal spread to the left hemisphere then the right hemisphere, bi-occipital onset	Two d: mild thinning of the corpus callosum 2 m 10 d: mild thinning of the corpus callosum, increasing prominence of CSF spaces, likely ex vacuo dilatation	Scheffer et al. (2020)
19	M	c.2284C > T c.2230_2237dupAACACTGC	p.Gln762[a] p.S747Tfs[a]36	Sporadic	Died at the age of 10 mo	Drug-resistant seizures, hypertonia, microcephaly	4–6 Hz theta background activity, bilateral frontotemporal sharp waves and 8–10 Hz alpha waves during clinical seizures	Initial: normal, 3 mo: cerebral and cerebellar atrophy and thinning of the corpus callosum	Celik et al. (2017)

(Continued on following page)

TABLE 1 | (*Continued*) Clinical features of individuals with previously described BRAT1 variants besides the present study.

NO.	Gender	Variant	Effect	Pedigree	Reported condition	Associated clinical phenotypes	Electroencephalogram	Brain MRI	Ref
20	M	c.2230_2237dupAACACTGC c.1499-1G > T	p.Glu500Alafs[a]36	Sporadic	Died at the age of 7.5 mo	Seizures (myoclonic, tonic and clonic migrating focal), hypotonia, micrognathia, microcephaly, down-slanted palpebral fissures, myoclonic jerks, apnea, and bradycardia	Generalized epileptiform activity, migrating focal epileptiform activity, and background deceleration	Atrophic corpus callosum, hypomyelinisation, brainstem, and cerebellar vermis hypoplasia	Colak et al. (2020)
21	M	c.1499-1G > T c.233G > C	p.Arg78Pro	sporadic	died at the age of 7 mo due to respiratory infection and malnutrition	myoclonic seizures, paroxysmal convulsions, hypertonia, hyperactive deep tendon reflexes, small and asymmetrical frontal bones, overlapping cranial sutures, recurrent respiratory tract infections, dysphagia	Initial: more sharp wave discharges in the left forehead-parietal region than in the right forehead-parietal region 2mo: focal sharp wave discharges and spike and slow-wave complexes in the left forehead-temporal region	brain magnetic resonance imaging indicated that the bilateral frontal and temporal subarachnoid space was widened, and the corpus callosum was thin	Li et al. (2021)
22	M	c.233G > C c.1173delG	p.Leu391fs	Sibling with NO.23	Died at the age of 6 mo due to cardiac arrest	Myoclonic seizures, hypertonia and contractures, arrested head growth, inability to swallow, and bouts of apnea-bradycardia, cardiac arrest	Bilateral epileptic activity with bilateral discharges	Normal	Straussberg et al. (2015)
23	F	c.1173delG c.1173delG	p.Leu391fs	Sibling with NO.22	Died at the age of 5 mo due to cardiac arrest	Myoclonic seizures, hypertonia and contractures, arrested head growth, inability to swallow, and bouts of apnea-bradycardia, cardiac arrest	Sharp waves and bilateral spikes predominantly over the right hemisphere	Normal	Straussberg et al. (2015)
24	F	c.1173delG c.1395G > C	p.Thr465Thr	Sporadic	Died at 10 weeks of age	Progressive encephalopathy with refractory seizures, hypertonia, episodic apnea, microcephaly, dysmorphic features	Diffuse encephalopathy, with frequent ictal activity from multiple cortical areas	Mild thinning of the corpus callous and delayed myelination	Van Ommeren et al. (2018)

(Continued on following page)

TABLE 1 | (*Continued*) Clinical features of individuals with previously described BRAT1 variants besides the present study.

NO.	Gender	Variant	Effect	Pedigree	Reported condition	Associated clinical phenotypes	Electroencephalogram	Brain MRI	Ref
25	M	c.1395G > C c.638dupA c.638dupA	p.Val214Glyfs[a]189	Sibling[a] with NO.14	Died at the age of 2 mo due to severe necrotizing enterocolitis grade III	Epileptic seizures (loss of consciousness, tonic posturing, myoclonus), hypertonia, microcephaly, mild hypotonia	Burst-suppression pattern, with long suppressions (10–15 s), multifocal negative sharp wavers	Not offered	Wolf et al. (2015)
26	M	c.1120G > T	p.Glu374[a]	Sporadic	Died at 2 mo	Microcephaly, hypertonia, myoclonic seizures, focal clonic seizures migrating between hemispheres, excessive startle from day 1	Multifocal epileptiform discharges, discontinuous backgroundIctal: myoclonic seizures, clonic seizures, facial clonic movements, with migration from the right posterior occipital region to the left posterior region	17 d: small subacute subdural hemorrhage along the tentorium with left parietal bone cephalhematoma	Scheffer et al. (2020)
27	F	c.1120G > T c.964C > T	p.Gln322[a]	Sibling with NO.18	Died at 34 d	Focal motor seizures, microcephaly, hypertonia	Multifocal epileptiform discharges, discontinuous backgroundIctal: focal seizure migrating from the left central region to the right hemisphere	3 d: asymmetric T2 signal in deep posterior parietal white matter bilaterally, small subdural hemorrhages in the posterior parietal region and posterior fossa, focal area of subarachnoid/pial hemorrhage in the posterior fossa adjacent to the tentorium on the right side 3 w: poor opercularization of Sylvian fissure in the frontotemporal region, hemorrhages unchanged	Scheffer et al. (2020)
28	M	c.2284C > T c.2041G > T c.2041G > T	p.Gln762[a] p. E681X	Sporadic	Died at 6 d old	Intractable focal seizures, microcephaly, rigidity, apnea, and congenital heart disease	Not offered	Not offered	Pourahmadiyan et al. (2021)

EEG = electroencephalogram.
MRI = serial magnetic resonance imaging of the brain.
d = day.
mo = month.
[a]*In "Pedigree" = sibling with similar symptoms has died without exome sequencing.*

type, and zygosity of the identified variant (13) (Dudding et al., 2004). According to **Table 1**, the variant type rather than the variant domain is closely related to the phenotype severity. For example, homozygous variants R78P and p. V62E, both located on the apoptosis-related N-terminal CIDE (cell death-inducing DFF-45-like effector) (Lugovskoy et al., 1999; Choi et al., 2017), caused typical RMFSL (Li et al., 2021) and mild NPCA (Lugovskoy et al., 1999), respectively. Complex mechanisms such as the mutation affecting one or more as unidentified activities of this protein (Kurosaki and Maquat, 2016) or associated with BRAT1-related pathways might be involved. It may modulate the severity rather than simply disrupt the mitochondrial function and the ATM kinase activity. As to the variant type, individuals with BRAT1 biallelic null variants usually lead to severe symptoms and are mostly fatal at the early stage. However, in biallelic BRAT1 missense variants, phenotypic variability is huge, ranging from RMFSL to NPCA. Interestingly, we noticed that in individuals with biallelic BRAT1 gene null variants, females always exhibited a milder phenotype than males (Mundy et al., 2016; Srivastava et al., 2016), even for siblings who carried the same variants. The special phenotypic divergence cannot be explained simply by variable penetrance or genetic backdrop heterogeneity. It is hypothesized that other sub-equivalent genes located on the X chromosome have some effects of the "female protective model" in neurodevelopmental disorders (Jacquemont et al., 2014). The mechanism is yet to be further confirmed.

The variant p. Leu236Cysfs*5 resulted truncation after the first 236 residues, and half of the protein was most likely to lose the function. The trans variant c.1014A > C translates to p. Pro338Pro, which has not been functionally silent. It generated an aberrant transcript with intron 7 retention by affecting the splicing accuracy. Intron retaining can reduce gene expression at the post-transcriptional level and thereby impose an additional level of gene regulation, such as the degradation of mRNA transcripts *via* nonsense-mediated decay (NMD) and the regulation of nuclear mRNA export (Low et al., 2015; Schmitz et al., 2017). Wild-type BRAT1 was diffusely expressed in the cytoplasm and the nucleus (Li et al., 2021). As-retained intron transcripts accumulate in the nucleus (Boutz et al., 2015), which might also reduce the amount of cytoplasmic BRAT1 available for the downstream events.

The poor efficiency of mRNA transcription, rather than completely abolished BRAT1 protein in a truncating genetic backdrop, made the BRAT1 protein synthesis in the early life insufficient to maintain the good development of the cerebellum, which may underlie the etiology of the mild, nonprogress phenotypic form of *BRAT1*-related neurodevelopmental disorders. c.1014A > C is the first pathogenic synonymous variant identified in the BRAT1 gene, which is associated with autosomal recessive NPCA. Before this, there was a synonymous substitution at c.1395G > C (p.Thr465 =), which had been reported in the severe phenotypic form of RMFSL (24). Although the mechanism has not been clarified, it conjectured that there exist multiple exon skipping, mRNA degradation, and complete deletion of BRAT1 protein. Novel synonymous variants in *BRAT1* should never be ignored as silent sound in diagnosis, which occasionally shed light on the underlying pathogenesis of the disease.

CONCLUSION

Our results not only broaden the mutation/phenotype spectrum of *BRAT1* but also contribute to comprehend possible pathogenic mechanisms of *BRAT1*. It is beneficial to specific genetic counseling and timely perinatal management.

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

Written informed consent was obtained from the individual(s)' and minor(s)' legal guardian/next to kin for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors have materially participated in the study and manuscript preparation. YQ carried out all the data analyses, participated in the design of the work, and wrote the draft; XJ, HD, and YZ collected all clinical data; LL participated in molecular genetic test work; AY designed the work and revised the manuscript. All authors critically reviewed the manuscript and provided final approval for submission. All authors agree to be accountable for all aspects of the work, ensuring the accuracy and integrity of the publication.

FUNDING

The study was financially supported by the Guangdong Medical Research Foundation (A2020257) and the Science and Technology Planning Project of Guangdong Province (2021A1414020008). This funding supported us in completing the experiments, sequencing, data analysis, and publication.

ACKNOWLEDGMENTS

We would like to thank all the members in the family of the patient for their interest and cooperation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.821587/full#supplementary-material

REFERENCES

Boutz, P. L., Bhutkar, A., and Sharp, P. A. (2015). Detained Introns Are a Novel, Widespread Class of post-transcriptionally Spliced Introns. *Genes Dev.* 29 (1), 63–80. doi:10.1101/gad.247361.114

Celik, Y., Okuyaz, C., Arslankoylu, A. E., and Ceylaner, S. (2017). Lethal Neonatal Rigidity and Multifocal Seizure Syndrome with a New Mutation in BRAT1. *Epilepsy Behav. Case Rep.* 8, 31–32. doi:10.1016/j.ebcr.2017.05.003

Choi, J. Y., Qiao, Q., Hong, S.-H., Kim, C. M., Jeong, J.-H., Kim, Y.-G., et al. (2017). CIDE Domains Form Functionally Important Higher-Order Assemblies for DNA Fragmentation. *Proc. Natl. Acad. Sci. USA* 114 (28), 7361–7366. doi:10.1073/pnas.1705949114

Colak, F. K., Guleray, N., Azapagasi, E., Yazıcı, M. U., Aksoy, E., and Ceylan, N. (2020). An Intronic Variant in BRAT1 Creates a Cryptic Splice Site, Causing Epileptic Encephalopathy without Prominent Rigidity. *Acta Neurol. Belg.* 120 (6), 1425–1432. doi:10.1007/s13760-020-01513-0

Dudding, T. E., Friend, K., Schofield, P. W., Lee, S., Wilkinson, I. A., and Richards, R. I. (2004). Autosomal Dominant Congenital Non-progressive Ataxia Overlaps with the SCA15 Locus. *Neurology* 63 (12), 2288–2292. doi:10.1212/01.wnl.0000147299.80872.d1

Fernandez-Jaen, A., and Ouchi, T. (2016). BRAT1 in Brain Function. *Neurooncol. Open Access* 1, 1–8.

Fernández-Jaén, A., Álvarez, S., Young So, E., Ouchi, T., Jiménez de la Peña, M., Duat, A., et al. (2016). Mutations in BRAT1 Cause Autosomal Recessive Progressive Encephalopathy: Report of a Spanish Patient. *Eur. J. Paediatric Neurol.* 20 (3), 421–425. doi:10.1016/j.ejpn.2016.02.009

Hanes, I., Kozenko, M., and Callen, D. J. A. (2015). Lethal Neonatal Rigidity and Multifocal Seizure Syndrome-A Misnamed Disorder? *Pediatr. Neurol.* 53 (6), 535–540. doi:10.1016/j.pediatrneurol.2015.09.002

Horn, D., Weschke, B., Knierim, E., Fischer-Zirnsak, B., Stenzel, W., Schuelke, M., et al. (2016). BRAT1mutations Are Associated with Infantile Epileptic Encephalopathy, Mitochondrial Dysfunction, and Survival into Childhood. *Am. J. Med. Genet.* 170 (9), 2274–2281. doi:10.1002/ajmg.a.37798

Jacquemont, S., Coe, B. P., Hersch, M., Duyzend, M. H., Krumm, N., Bergmann, S., et al. (2014). A Higher Mutational burden in Females Supports a "female Protective Model" in Neurodevelopmental Disorders. *Am. J. Hum. Genet.* 94 (3), 415–425. doi:10.1016/j.ajhg.2014.02.001

Jones, T. M., and Montero, F. J. (2021). "Chorionic Villus Sampling," in *StatPearls [Internet]* (Treasure Island (FL): StatPearls Publishing).

Kurosaki, T., and Maquat, L. E. (2016). Nonsense-mediated mRNA Decay in Humans at a Glance. *J. Cell Sci* 129 (3), 461–467. doi:10.1242/jcs.181008

Li, W., Wu, S., Xu, H., Zhao, X., Pan, Y., Huang, H., et al. (2021). Novel Variant in BRAT1 with the Lethal Neonatal Rigidity and Multifocal Seizure Syndrome. *Pediatr. Res.* 2021, 1. doi:10.1038/s41390-021-01468-9

Low, L.-H., Chow, Y.-L., Li, Y., Goh, C.-P., Putz, U., Silke, J., et al. (2015). Nedd4 Family Interacting Protein 1 (Ndfip1) Is Required for Ubiquitination and Nuclear Trafficking of BRCA1-Associated ATM Activator 1 (BRAT1) during the DNA Damage Response. *J. Biol. Chem.* 290 (11), 7141–7150. doi:10.1074/jbc.M114.613687

Lugovskoy, A. A., Zhou, P., Chou, J. J., McCarty, J. S., Li, P., and Wagner, G. (1999). Solution Structure of the CIDE-N Domain of CIDE-B and a Model for CIDE-N/CIDE-N Interactions in the DNA Fragmentation Pathway of Apoptosis. *Cell* 99 (7), 747–755. doi:10.1016/s0092-8674(00)81672-4

Mahjoub, A., Cihlarova, Z., Tétreault, M., MacNeil, L., Sondheimer, N., Caldecott, K. W., et al. (2019). Homozygous Pathogenic Variant in BRAT1 Associated with Nonprogressive Cerebellar Ataxia. *Neurol. Genet.* 5 (5), e359. doi:10.1212/NXG.0000000000000359

Mundy, S. A., Krock, B. L., Mao, R., and Shen, J. J. (2016). BRAT1-related Disease-Identification of a Patient without Early Lethality. *Am. J. Med. Genet.* 170 (3), 699–702. doi:10.1002/ajmg.a.37434

Oatts, J. T., Duncan, J. L., Hoyt, C. S., Slavotinek, A. M., and Moore, A. T. (2017). Inner Retinal Dystrophy in a Patient with Biallelic Sequence Variants in BRAT1. *Ophthalmic Genet.* 38 (6), 559–561. doi:10.1080/13816810.2017.1290118

Ouchi, M., and Ouchi, T. (2010). Regulation of ATM/DNA-PKcs Phosphorylation by BRCA1-Associated BAAT1. *Genes & Cancer* 1 (12), 1211–1214. doi:10.1177/1947601911404222

Pourahmadiyan, A., Heidari, M., Shojaaldini Ardakani, H., Noorian, S., and Savad, S. (2021). A Novel Pathogenic Variant of BRAT1 Gene Causes Rigidity and Multifocal Seizure Syndrome, Lethal Neonatal. *Int. J. Neurosci.* 131 (9), 875–878. doi:10.1080/00207454.2020.1759589

Puffenberger, E. G., Jinks, R. N., Sougnez, C., Cibulskis, K., Willert, R. A., Achilly, N. P., et al. (2012). Genetic Mapping and Exome Sequencing Identify Variants Associated with Five Novel Diseases. *PLoS One* 7 (1), e28936. doi:10.1371/journal.pone.0028936

Saunders, C. J., Miller, N. A., Soden, S. E., Dinwiddie, D. L., Noll, A., Alnadi, N. A., et al. (2012). Rapid Whole-Genome Sequencing for Genetic Disease Diagnosis in Neonatal Intensive Care Units. *Sci. Transl. Med.* 4 (154), 154ra135. doi:10.1126/scitranslmed.3004041

Scheffer, I. E., Boysen, K. E., Schneider, A. L., Myers, C. T., Mehaffey, M. G., Rochtus, A. M., et al. (2020). BRAT 1 Encephalopathy: a Recessive Cause of Epilepsy of Infancy with Migrating Focal Seizures. *Dev. Med. Child. Neurol.* 62 (9), 1096–1099. doi:10.1111/dmcn.14428

Schmitz, U., Pinello, N., Jia, F., Alasmari, S., Ritchie, W., Keightley, M.-C., et al. (2017). Intron Retention Enhances Gene Regulatory Complexity in Vertebrates. *Genome Biol.* 18 (1), 216. doi:10.1186/s13059-017-1339-3

Smith, N. J., Lipsett, J., Dibbens, L. M., and Heron, S. E. (2016). BRAT1-associated Neurodegeneration: Intra-familial Phenotypic Differences in Siblings. *Am. J. Med. Genet.* 170 (11), 3033–3038. doi:10.1002/ajmg.a.37853

So, E. Y., and Ouchi, T. (2014). BRAT1 Deficiency Causes Increased Glucose Metabolism and Mitochondrial Malfunction. *BMC Cancer* 14, 548. doi:10.1186/1471-2407-14-548

Srivastava, S., Olson, H. E., Cohen, J. S., Gubbels, C. S., Lincoln, S., Davis, B. T., et al. (2016). BRAT1mutations Present with a Spectrum of Clinical Severity. *Am. J. Med. Genet.* 170 (9), 2265–2273. doi:10.1002/ajmg.a.37783

Straussberg, R., Ganelin-Cohen, E., Goldberg-Stern, H., Tzur, S., Behar, D. M., Smirin-Yosef, P., et al. (2015). Lethal Neonatal Rigidity and Multifocal Seizure Syndrome - Report of Another Family with a BRAT1 Mutation. *Eur. J. Paediatric Neurol.* 19 (2), 240–242. doi:10.1016/j.ejpn.2014.11.004

Szymańska, K., Laure-Kamionowska, M., Szczałuba, K., Koppolu, A., Furmanek, M., Kuśmierska, K., et al. (2018). Clinico-pathological Correlation in Case of BRAT1 Mutation. *fn* 56 (4), 362–371. doi:10.5114/fn.2018.80870

Van Ommeren, R. H., Gao, A. F., Blaser, S. I., Chitayat, D. A., and Hazrati, L.-N. (2018). BRAT1Mutation: The First Reported Case of Chinese Origin and Review of the Literature. *J. Neuropathol. Exp. Neurol.* 77 (12), 1071–1078. doi:10.1093/jnen/nly093

Vetro, A., Pisano, T., Chiaro, S., Procopio, E., Guerra, A., Parrini, E., et al. (2020). Early Infantile Epileptic-Dyskinetic Encephalopathy Due to Biallelic PIGP Mutations. *Neurol. Genet.* 6 (1), e387. doi:10.1212/NXG.0000000000000387

Wolf, N., van Weissenbruch, M., Stam, C., Weiss, J., Waisfisz, Q., Kevelam, S., et al. (2015). Early-Onset Severe Encephalopathy with Epilepsy: The BRAT1 Gene Should Be Added to the List of Causes. *Neuropediatrics* 46 (6), 392–400. doi:10.1055/s-0035-1564791

Zhang, C.-J., Hou, X.-M., Tan, L.-M., Shao, C.-R., Huang, H.-W., Li, Y.-Q., et al. (2016). The Arabidopsis Acetylated Histone-Binding Protein BRAT1 Forms a Complex with BRP1 and Prevents Transcriptional Silencing. *Nat. Commun.* 7, 11715. doi:10.1038/ncomms11715

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

CASE REPORT
published: 14 March 2022
doi: 10.3389/fgene.2022.843931

Case Report: Preimplantation Genetic Testing for Meckel Syndrome Induced by Novel Compound Heterozygous Mutations of *MKS1*

Tingting Lin[1,2,3], *Yongyi Ma*[4], *Danni Zhou*[1,2,3], *Liwei Sun*[1,2,3], *Ke Chen*[1,2,3], *Yezhou Xiang*[1,2,3], *Keya Tong*[1,2,3], *Chaoli Jia*[1,2,3], *Kean Jiang*[1,2,3], *Dongyun Liu*[1,2,3]* and *Guoning Huang*[1,2,3]*

[1]*Chongqing Key Laboratory of Human Embryo Engineering, Chongqing, China,* [2]*Chongqing Clinical Research Center for Reproductive Medicine, Chongqing, China,* [3]*Reproductive and Genetic Institute, Chongqing Health Center for Women and Children, Chongqing, China,* [4]*The Southwest Hospital of Army Medical University, Chongqing, China*

Meckel syndrome (MKS), also known as the Meckel–Gruber syndrome, is a severe pleiotropic autosomal recessive developmental disorder caused by dysfunction of the primary cilia during early embryogenesis. The diagnostic criteria are based on clinical variability and genetic heterogeneity. Mutations in the *MKS1* gene constitute approximately 7% of all MKS cases. Herein, we present a non-consanguineous couple with three abnormal pregnancies as the fetuses showed MKS-related phenotypes of the central nervous system malformation and postaxial polydactyly. Whole-exome sequencing identified two novel heterozygous mutations of *MKS1*: c.350C>A and c.1408-14A>G. The nonsense mutation c.350C>A produced a premature stop codon and induced the truncation of the MKS1 protein (p.S117*). Reverse-transcription polymerase chain reaction (RT-PCR) showed that c.1408-14A>G skipped exon 16 and encoded the mutant MKS1 p.E471Lfs*92. Functional studies showed that these two mutations disrupted the B9–C2 domain of the MKS1 protein and attenuated the interactions with B9D2, the essential component of the ciliary transition zone. The couple finally got a healthy baby through preimplantation genetic testing for monogenic disorder (PGT-M) with haplotype linkage analysis. Thus, this study expanded the mutation spectrum of *MKS1* and elucidated the genetic heterogeneity of *MKS1* in clinical cases.

Keywords: *MKS1* gene, Meckel syndrome, PGT-M, intron mutation, exon skipping variant

OPEN ACCESS

Edited by:
Wenhao Zhou,
Fudan University, China

Reviewed by:
Katarzyna Szymanska,
University of Leeds, United Kingdom
Hao Hu,
Guangzhou Medical University, China

***Correspondence:**
Dongyun Liu
dongyunliu@163.com
Guoning Huang
gnhuang217@sina.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *27 December 2021*
Accepted: *02 February 2022*
Published: *14 March 2022*

Citation:
Lin T, Ma Y, Zhou D, Sun L, Chen K, Xiang Y, Tong K, Jia C, Jiang K, Liu D and Huang G (2022) Case Report: Preimplantation Genetic Testing for Meckel Syndrome Induced by Novel Compound Heterozygous Mutations of MKS1. Front. Genet. 13:843931. doi: 10.3389/fgene.2022.843931

INTRODUCTION

Cilia are microtubule-based organelles that extend from the surface of most eukaryotic cells. Defects in this organelle cause a series of disorders known as ciliopathies (Mitchison and Valente, 2017; Reiter and Leroux, 2017; Andreu-Cervera et al., 2021; Luo et al., 2021). Meckel syndrome (MKS, MIM 249000) is a rare and lethal autosomal recessive ciliopathy with highly variable phenotypes, extreme genetic heterogeneity, and complex allelism with other related ciliopathies, such as Joubert syndrome (JBTS, MIM 213300) (Salonen et al., 1984a; Parisi, 2019). MKS is mainly characterized by central nervous system malformation (most commonly occipital encephalocele), cystic kidney dysplasia, fibrotic changes of the liver, and postaxial polydactyly (Logan et al., 2011; Hartill et al., 2017). Globally, the incidence rate of MKS has been estimated at

1/140,000–1/13,250 in live births, and a high prevalence was observed in Finland and Belgium (Salonen et al., 1984b; Auber et al., 2007). The MKS-affected fetuses usually die *in utero* or shortly after birth.

Genetic studies have identified several genes related to MKS, such as *MKS1*, *TMEM216*, *TMEM67*, *CEP290*, *RPGRIP1L*, *CC2D2A*, *NPHP3*, *TCTN2*, *B9D1*, *B9D2*, *TMEM231*, *KIF14*, and *TMEM107*, and most of them encode proteins concentrated to the ciliary transition zone (TZ) (Bergmann et al., 2016; Dean et al., 2016; Wu et al., 2020). The TZ is characterized by Y-shaped structures spanning from the axoneme to the ciliary membrane that functions as a barrier between the cilia components and the cytoplasmic group to regulate the material transport and signal transmission of the cilia (Garcia-Gonzalo et al., 2011; Anvarian et al., 2019). Systematic genetic studies have grouped the known TZ proteins into three functional modules: MKS module, NPHP module, and CEP290 module (Goncalves and Pelletier, 2017). The B9 domain-containing proteins, including MKS1, B9D1, and B9D2, function as soluble MKS-module components and are associated with normal cilia biogenesis and ciliary diffusion (Bialas et al., 2009; Gogendeau et al., 2020; Okazaki et al., 2020). Previous studies have identified >80 pathogenic *MKS1* mutations that contribute to approximately 7% of all reported MKS cases (Hartill et al., 2017).

Preimplantation genetic testing (PGT) is an invasive prenatal diagnosis that involves the biopsy of a single or few cells from *in vitro* fertilized embryos and testing of the biopsied samples for genetic aberrations, followed by the selective transfer of unaffected embryos under specific conditions (De Rycke and Berckmoes, 2020). Clinically, PGT is available for monogenic disorder (PGT-M), wherein the disease-causing locus has been identified unequivocally. In the present study, we identified two novel *MKS1* mutations, c.350C>A and c.1408-14A>G, in a couple with three times of abnormal pregnancies. The genetic analysis and functional study showed the pathogenicity of these two mutation sites. Finally, assisted reproductive technology (ART) combined with PGT-M helped the couple get a healthy baby. These findings extended the spectrum of *MKS1* mutations in MKS and manifested the role of PGT in blocking single-gene diseases.

METHODS AND MATERIALS

Subjects and Ethical Approval

The non-consanguineous couple first visited the Institute of Reproduction and Genetics, Chongqing Health Center for Women and Children (Chongqing, China), three times due to abnormal pregnancies and to consult for PGT. Pedigree data were obtained from the couple and their parents. Clinical assessments, including ultrasonic examination and assisted reproductive technology (ART), were performed in the related clinical departments. This study was approved by the Ethics Committee of the Chongqing Health Center for Women and Children. Informed consent was obtained from the couple.

Whole-Exome Sequencing and Variants Analysis

The exomes were captured using the Agilent SureSelect Human All Exon V6 Kit (Agilent Technologies Inc., CA, United States) and sequenced on an Illumina NovaSeq 6000 platform (Illumina Inc., CA, United States). The clean reads derived from targeted sequencing were filtered and aligned to the human reference genome (GRCh37/hg19) using the Burrows–Wheeler Aligner (BWA) (Li and Durbin, 2009). Single-nucleotide variants (SNVs) and InDels were called using Genome Analysis Toolkit (GATK), annotated with Ensembl Variant Effect Predictor (McLaren et al., 2016), and filtered using multiple databases, including NCBI dbSNP, HapMap, 1,000 human genome dataset, and gnomAD. Finally, all the variants were annotated according to the guidelines of the American College of Medical Genetics and Genomics (ACMG) (Richards et al., 2015), and the variants from known causative genes of MKS were analyzed. The Human Gene Mutation Database (HGMD) and VarSome were used to screen the mutations reported previously. The variants were validated by Sanger sequencing.

Validation of Mutations

Sanger sequencing was used to validate *MKS1* mutations in the aborted fetuses (II:1 and II:3) and the couple (I:1 and I:2) (**Figure 1A**) The following primers were used: *MKS1*-exon4-forward: 5′-TTCTTGGTTCCCCTGCCATTC-3′, *MKS1*-exon4-reverse: 5′-CTCACCACCTGTAGACTGTGC-3’; *MKS1*-intron15-forward: 5′-CTGTGTCATTGCTGGGGAGTC-3′, *MKS1*-intron15-reverse: 5′-CCAGCCACATGGTTACGG-3′. The products were purified on 2% agarose gels, sequenced with ABI 3500 (Thermo Fisher, MA, United States), and analyzed using Chromas 2.6.5 (Technelysium Pvt. Ltd., United States).

Amino Acid Conservation and Protein Sequences

The amino acid sequence and name of mutant MKS1 proteins were analyzed using Name Checker (https://mutalyzer.nl/name-checker). The MKS1 amino acid sequences from different species were compared using the ClustalW software and analyzed using ESPript 3.0 (https://espript.ibcp.fr/ESPript). Wild-type and mutant MKS1 proteins were modeled using Illustrator for Biological Sequences (IBS, http://ibs.biocuckoo.org/online.php).

Intron Mutation Analysis and Reverse-Transcription Polymerase Chain Reaction

The mutations at the 3′-terminus of intron 15 were found in the gnomAD database (http://gnomad-sg.org/). The effect of these mutations was analyzed with varSEAK (https://varseak.bio/) and SpliceAI (https://spliceailookup.broadinstitute.org/). Total RNA was extracted from blood samples using the QIAamp RNA blood mini kit (Qiagen, Germany), according to the manufacturer's instructions. PrimeScript Reverse-Transcription Polymerase Chain Reaction (RT-PCR) kit (Takara, Japan) was used for

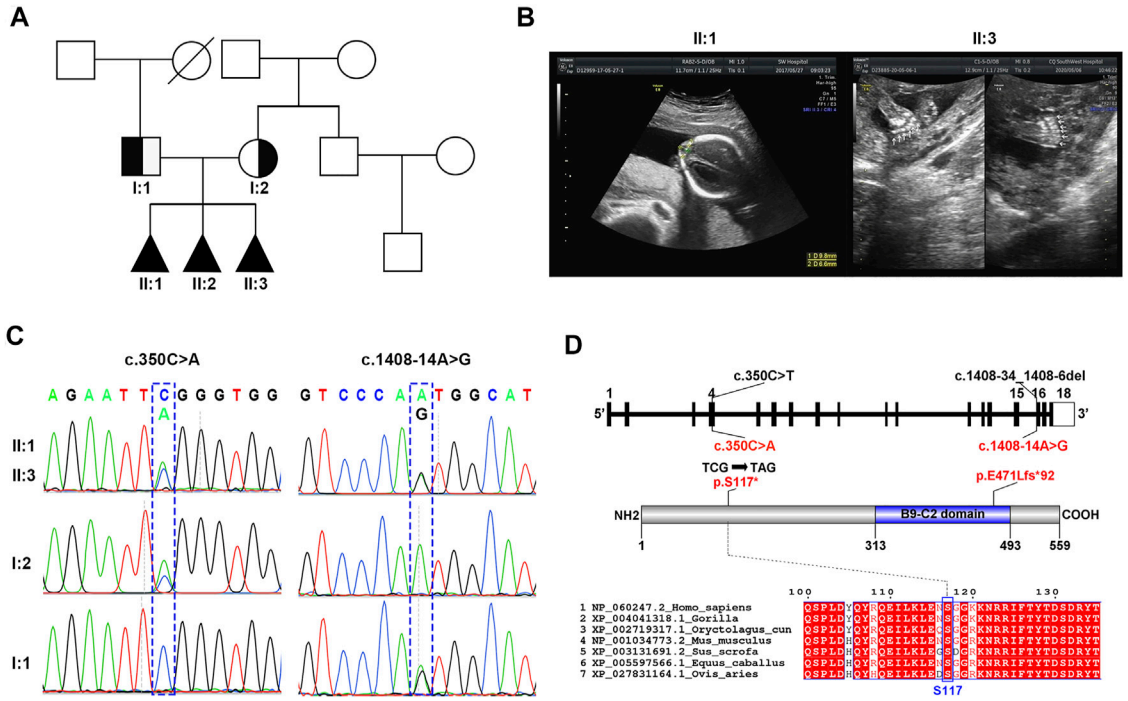


FIGURE 1 | Identification of novel *MKS1* mutations in a Chinese family with MKS. **(A)** The pedigree of this family showed a history of abnormal pregnancy. **(B)** Ultrasonographic images of fetuses II:1 and II:3 showed the occipital encephalocele and postaxial polydactyly, respectively. **(C)** Sanger sequencing of *MKS1* showed the inheritance pattern of mutant sites between the couple and the three fetuses. **(D)** Schematic of *MKS1* gene and MKS1 protein. The mutation sites with related amino acid change were labeled. The B9 domain is labeled as the B9-C2 domain.

RT-PCR. The forward primer (5′-GGCTGAGCTGAGGAGGTTTT-3′) used for cDNA amplification was located at exon 15, while the reverse primer (5′-CTTCCAGACGGTCCAACACA-3′) was located at exon 17. Then, the products were purified on 2% agarose gels and analyzed on ABI 3500.

qPCR

The primers used for real-time fluorescence quantitative PCR (qPCR) were as follows: forward 5′-CTCCGAGTCCACCTGCAAAGAATC-3′ and reverse 5′-CTCCTCCTCTTCGTCTTCCTCTGG-3′ for *MKS1* exons 2 and 3; forward 5′-GGATCCTTCAAGGGGGAACG-3′ and reverse 5′-CATGAAGGCCCTGGACTGCT-3′ for *MKS1* exon 16; forward 5′-ATGCAGAATCCACGCCAGTACAAG-3′ and reverse 5′-TCAGTCGCTCCAGGTCTTCACG-3′ for *RPS18* as the control. The expression of *MKS1* was evaluated using EvaGreen SuperMix (Bio-Rad, United States) on a CFX96 apparatus (Bio-Rad, United States) and analyzed using the $2^{-\Delta\Delta Ct}$ method by normalizing to that of *RPS18*.

Plasmid Construction

The full-length coding sequence (CDS) of *MKS1* (NM_017777.4) was amplified by RT-PCR from wild-type, c.350C>A heterozygous, and c.1408-14A>G heterozygous subjects and subcloned into the pcDNA3.1-Myc B vectors (Invitrogen, USA). B9D2 was inserted into the pCMV-HA plasmid. All plasmid sequences were validated by Sanger sequencing.

Cell Culture, Transfection, and Western Blot Analysis

HEK293T cells were provided by Stem Cell Bank, Chinese Academy of Sciences (Shanghai, China), and grown at 37°C in the presence of 5% CO_2 in DMEM (HyClone, United States) supplemented with 10% fetal bovine serum (HyClone, United States). 293T cells were seeded in six-well plates (Corning, United States), and 2.5 μg wild-type or mutant MKS1 vector was transfected using Lipofectamine 3000 reagent (Thermo Fisher, United States). The cells were lysed with cell lysis buffer for western blot and immunoprecipitation (IP) (Beyotime, China) after transfection for different times (12, 24, 36, and 48 h), and 20 μg protein was analyzed by western blot. The anti-Myc antibody (AM926, 1:1,000) was purchased from Beyotime. Actin antibody (Beyotime, AA128, 1:1,000) served as an internal control.

Co-IP

For co-IP, MKS1 (1.5 μg) and B9D2 (1.0 μg) expression vectors were co-transfected into HEK293T cells. After 36 h post-transfection, HEK293T cells were rinsed with ice-cold phosphate-buffered saline (PBS) and lysed with IP lysis buffer (Beyotime, China) supplemented with a protease inhibitor cocktail. After 20 min, cell lysates were cleared by centrifugation at 14,000 × *g*, 4°C for 5 min. The supernatant was used for the co-IP assay by shaking with BeyoMag™ anti-

Myc magnetic beads (Beyotime, China) at 4°C for 4–6 h. After three washes, protein-Myc bead complex was eluted with IP buffer containing 150 μg/ml 3× Myc peptides (Beyotime, China) for 2 h. Then, the elution products were subjected to western blotting. The following antibodies (1:1,000, Beyotime, China) were used: anti-Myc, anti-HA, and goat anti-mouse HRP.

Controlled Ovarian Stimulation

The COS was conducted using a gonadotropin-releasing hormone (GnRH) antagonist protocol based on the ovarian reserve of the I:2 subject. It was initiated on day 2 of the cycle with a dose of 250 IU recombinant follicle-stimulating hormone (rFSH, Puregon, Organon, Netherlands). GnRH antagonist (0.25 mg; Cetrotide, Merck Serono, Switzerland) was given on cycle day 8. Human chorionic gonadotropin (hCG, Merck Serono, Switzerland) was administered as a trigger on cycle day 10, and transvaginal oocyte retrieval was performed after 36 h. Consequently, 13 oocytes were obtained, and five blastocysts were biopsied after intracytoplasmic sperm injection (ICSI).

PGT-M Procedure

Whole-genome amplification of each embryo biopsy sample was performed using the MALBAC WGA kit (Yikon Genomics, China), following the manufacturer's instructions. A total of 60 single-nucleotide polymorphism (SNP) markers linked to the mutation alleles were selected for linkage analysis. The mutation site and SNPs were amplified using specific primer pairs; the amplification products were pooled with the MALBAC WGA products and sequenced. The chromosomal copy number and the mutation site and SNPs were analyzed, as published previously (Huang et al., 2015).

RESULTS

Identification of Novel *MKS1* Mutations in MKS-Related Family

As shown by the family genetic map (**Figure 1A**), the non-consanguineous couple (I:1 and I:2) suffered from abnormal pregnancy three times, while the family presented no related medical history. Ultrasonographic images showed that the three aborted fetuses (II:1, II:2, and II:3) had clinical features of MKS such as occipital encephalocele, cerebellar vermis agenesis, and postaxial polydactyly (**Figure 1B**; **Supplementary Table S1**). Renal/hepatic involvement was not observed by ultrasonography in these fetuses. WES of the proband II:3 identified two novel *MKS1* (NM_017777.4) variants, c.350C>A and c.1408-14A>G (**Figure 1C**). Moreover, analysis of *MKS1* with Sanger sequencing in I:1, I:2, II:1, and II:3 showed that these two compound heterozygous mutations were inherited from their parents (**Figure 1C**).

The maternally inherited variant, c.350C>A, is a novel nonsense mutation causing the premature stop of MKS1 translation at the conserved Ser117, thereby encoding the mutant MKS1 p.S117*, while the paternally inherited variant, c.1408-14A>G, is located at intron 15 (**Figure 1D**). *MKS1* c.350C>A is not recorded in the human disease-related databases (gnomAD, ClinVar, and HGMD), and the truncated MKS1 protein would be a loss of function without the C-terminal B9-C2 domain. *MKS1* c.1408-14A>G is known as rs1194131222 with a rare frequency (0.000007228) in the gnomAD database, and no clinical case has yet been reported. Thus, according to the ACMG guidelines, the *MKS1* c.350C>A mutation is classified as "pathogenic" (PVS1: very strong pathogenicity, PS4: strong pathogenicity, PM2: moderate pathogenicity, and PP1: supporting pathogenicity), and the *MKS1* c.1408-14A>G mutation is classified as "uncertain significance" (PM3+PP1+PP4).

c.1408-14A>G Induced the Skip of Exon 16 in *MKS1* mRNA Splicing

As previously reported, *MKS1* c.1408-34_1408-6del29bp [AGAAACCTGAGGCTGTCCCAATGGCATGC], the Finnish major mutation, affected the *MKS1* mRNA splicing with the skip of exon 16 and induced frameshift of the MKS1 protein, resulting in MKS in a homozygous pattern (Kyttälä et al., 2006; Auber et al., 2007). Aberrant splicing was reported as a crucial mutational mechanism in *MKS1*-induced Meckel–Gruber syndrome (Frank et al., 2007). To date, 22 variants have been identified at the 3′-terminus of intron 15, while two variants (c.1408-1G>A and c.1408-34_1408-6del29bp) were predicted to induce the skip of exon 16, as assessed by varSEAK and SpliceAI (**Figure 2A**; **Supplementary Table S2**). The analysis showed an uncertain significance of c.1408-14A>G on *MKS1* mRNA splicing.

To further investigate the effect of c.1408-14A>G, RT-PCR was conducted with primers specific to exons 15 and 17. The results showed that there was one band >200 bp in I:2, and there were two bands in I:1 (one >200 bp and one <200 bp) (**Figure 2B**). Sanger sequencing of the two bands demonstrated a direct connection between exons 15 and 17 in the shorter one with the skip of exon 16 (**Figure 2C**), implying that the effect of c.1408-14A>G was like c.1408-34_1408-6del29bp on *MKS1* mRNA splicing. Using the *in silico* prediction software Name-Checker, the mutant CDS was predicted to encode the MKS1 p.E471Lfs*92 protein (**Supplementary data S1**). BLAST and alignment with MKS1 proteins across evolution manifested the partial dysfunction of the highly conserved B9-C2 domain (313–493 aa) in MKS1 p.E471Lfs*92 (**Figure 2D**). Herein, we updated the clinical significance of *MKS1* c.1408-14A>G as "pathogenic."

Mutations Disrupted the Function of MKS1 Protein

Previous studies have shown that the nonsense codons in all internal exons could trigger a nonsense-mediated mRNA decay (NMD) process (Nagy and Maquat, 1998). The c.350C>A variant resulted in a premature stop codon in exon 4, which probably triggered the degradation of mutant *MKS1* mRNA through the NMD pathway, while the c.1408-14A>G variant induced the skip of exon 16 and generated a new stop codon in the 3′-untranslated region (3′-UTR) (**Figure 3A**). The results of qPCR with specific

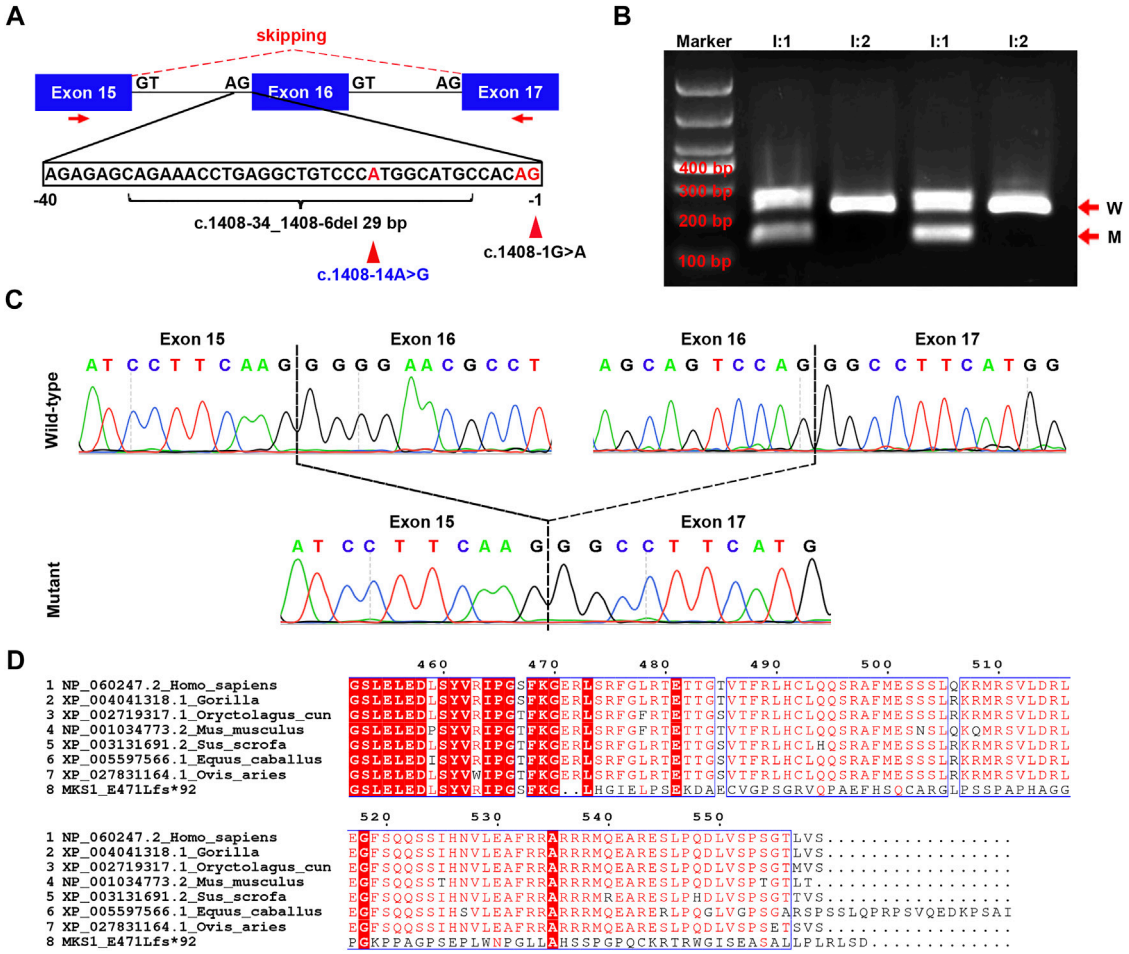


FIGURE 2 | Confirmation for the effect of c.1408-14A>G on *MKS1* mRNA splicing. **(A)** Schematic of the three exons (exons 15, 16, and 17), the partial sequence of intron 15, and the location of three variants (c.1408-1G>A, c.1408-14A>G, and c.1408-34_1408-6del29bp). The red arrows indicate the localization of primers used in RT-PCR. **(B)** Image of agarose gel electrophoresis of the RT-PCR products from the couple (I:1 and I:2). The red arrows show the wild-type (W) and mutant (M) product. **(C)** Sanger sequence of the RT-PCR products. The upper panel shows the wild-type *MKS1* sequence with two dashed lines labeling the boundary of exon 15/exon 16 and exon 16/exon 17. In the mutant sequence, exon 15 was ligated directly with exon 17, manifesting the ship of exon 16 during mRNA splicing. **(D)** Alignment results of human wild-type and mutant MKS1 proteins with other species (gorilla, pig, horse, sheep, rabbit, and mouse). Only the C-terminal sequence is displayed.

primers for exons 2–3 and 16 respectively showed that the expression level of *MKS1* gene in an I:2 heterozygous subject was equivalent to the control individual (**Figure 3B**), implying the absence of NMD process for the c.350C>A variant. Moreover, the *MKS1* level detected for exon 16 was downregulated in the c.1408-14A>G heterozygous individual (I:1), confirming the skip of exon 16 during the *MKS1* mRNA splicing process.

The C-terminal B9-C2 domain (313–493 aa) of MKS1 protein is conserved across evolution and essential for the predominant interaction between MKS1, B9D1, and B9D2, which is essential for cilial function (Dowdle et al., 2011; Romani et al., 2014; Okazaki et al., 2020). The truncated protein MKS1 p.S117* produced by the c.350C>A variant was predicted to lose its function completely, while the elongated protein MKS1 p.E471Lfs*92 maintained partial B9-C2 domain (313–470 aa) with the frameshift of the C-terminus (471–561 aa). We transfected 293T cells with Myc-tagged wild-type or mutant MKS1 expression plasmids to investigate the effect of these two variants on the B9 domain function. qPCR with specific primers for exons 2–3 confirmed the transcription of wild-type and mutant *MKS1* in the transfected 293T cells (**Figure 3C**). However, immunoblots with anti-Myc antibody demonstrated the expression of the wild-type and the frameshifted MKS1 proteins, while the MKS1 p.S117* was undetectable (**Figure 3D**). Co-IP assays with B9D2 protein showed that the wild-type MKS1 but not the MKS1 p.E471Lfs*92 interacted with B9D2, implying the dysfunction of the B9-C2 domain of MKS1 p.E471Lfs*92 (**Figure 3E**). Thus, c.350C>A and c.1408-14A>G variants disrupted the function of MKS1 and were pathogenic for fetal development.

PGT for the *MKS1* Variants

To avoid the occurrence of abnormal pregnancy, the non-consanguineous couple (I:1 and I:2) chose the *in vitro* assisted

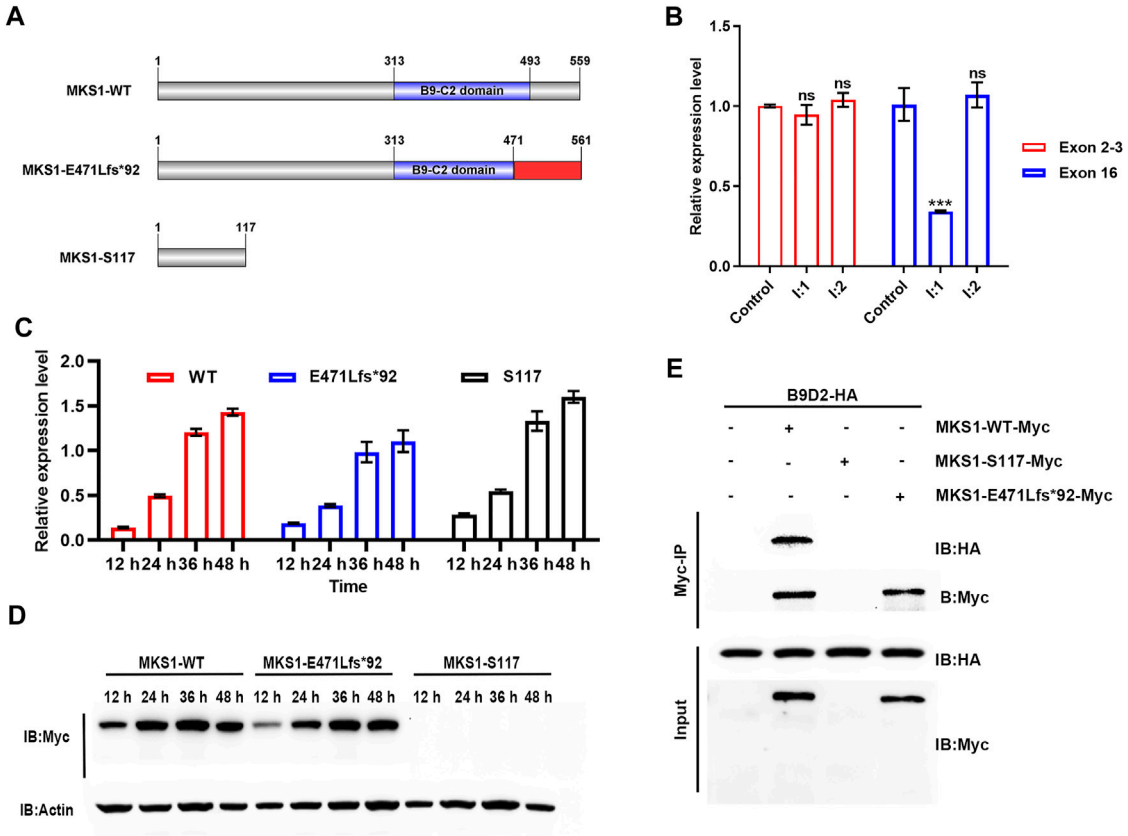


FIGURE 3 | MKS1 mutants lose the interaction with B9D2 protein. **(A)** Schematic of the MKS1 proteins (wild-type, E471Lfs*92, S117). The red rectangle indicates the frameshifted sequence of MKS1 p.E471Lfs*92. **(B)** Relative *MKS1* mRNA levels of the individuals (I:1 and I:2) compared to the healthy control. **(C)** Relative *MKS1* mRNA levels of 293T cells transfected with Myc-tagged MKS1 CDS shown in **(A)** for 12, 24, 36, and 48 h. Data from three independent experiments were used for quantification. Error bars represent mean ± SD. Statistical significance was determined by unpaired Student's *t*-test. ns: nonsense. ***$p < 0.001$. **(D)** Immunoblot analysis of 293T cells shown in **(C)**, implying the expression of wild-type and frameshifted MKS1 with the absence of MKS1 p.S117*. **(E)** Immunoprecipitation of Myc-tagged MKS1 variants with B9D2. The pull-down of B9D2 with the Myc bead demonstrated the function of the B9-C2 domain of MKS1 proteins.

reproductive technology combined with PGT-M for the *MKS1* variants. Clinically, the I:2 individual was consecutively treated with an antagonist for ovulation induction, and 11 mature oocytes at the metaphase II (MII) stage were retrieved through the laparoscopic ovarian puncture method (**Figure 4A**). After intracytoplasmic sperm injection (ICSI), seven zygotes developed into transferable embryos, and five blastocysts were biopsied for amplification with multiple annealing and looping-based cycles (**Figure 4B**).

The linkage analysis with SNP array for haplotype showed that two embryos were normal, and the other three were affected (**Figure 4C**). Finally, one healthy embryo was transferred, and luteal phase supports were administered routinely. Serum β-hCG levels were measured at 14 days after frozen embryo transplantation (FET). The presence of a gestational sac and fetal heartbeat detected by ultrasound at 5 weeks after FET was evidence of clinical pregnancy. Sanger sequencing was performed on the amniotic fluid sample collected by amniocentesis at 18 weeks of gestation; no *MKS1* mutation was found.

DISCUSSION

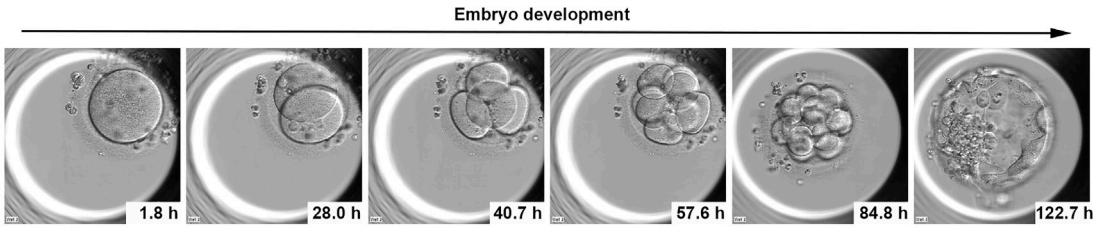
In this study, we reported a Chinese MKS-related family with two novel *MKS1* mutations displaying occipital encephalocele, cerebellar vermis agenesis, and postaxial polydactyly. No renal/hepatic involvement was observed by ultrasonography. The nonsense mutation, c.350C>A, induced premature termination of MKS1 translation but did not trigger the degradation of mutant mRNA by NMD. The c.1408-14A>G variant was in intron 15 and resulted in the skip of exon 16 during the *MKS1* mRNA splice, thereby coding an elongated MKS1 protein (p.E471Lfs*92). *In vitro* functional analysis with 293T cells showed the instability of MKS1 p.S117* and the disruption of B9-C2 domain in MKS1 p.E471Lfs*92. Finally, the non-consanguineous couple was assisted with PGT-M for pregnancy with a healthy baby without *MKS1* mutations.

The *MKS1*-related genotype–phenotype correlation was proposed as follows: two null alleles of *MKS1* result in MKS; one null allele and one non-truncating allele that leaves the B9-C2 domain intact result in JBTS; two non-truncating alleles result in Bardet–Biedl syndrome (BBS, MIM 615990) (Bader et al., 2016; Luo et al., 2020). Previous studies have identified the compound heterozygous mutations of

A

Cycle No.	No. of oocytes retrieved	No. of MII oocytes (%)	Insemination method	Total fertilization rate (%)	No. of transferable embryos	No. of embryos biopsied	Normal embryos	Affected embryos	No. of embryos transferred	Clinical outcome
1	13	11/13 (84.6)	ICSI	8/11 (72.7)	7/8	5/7	2/5	3/5	1	Pregnant

B



C

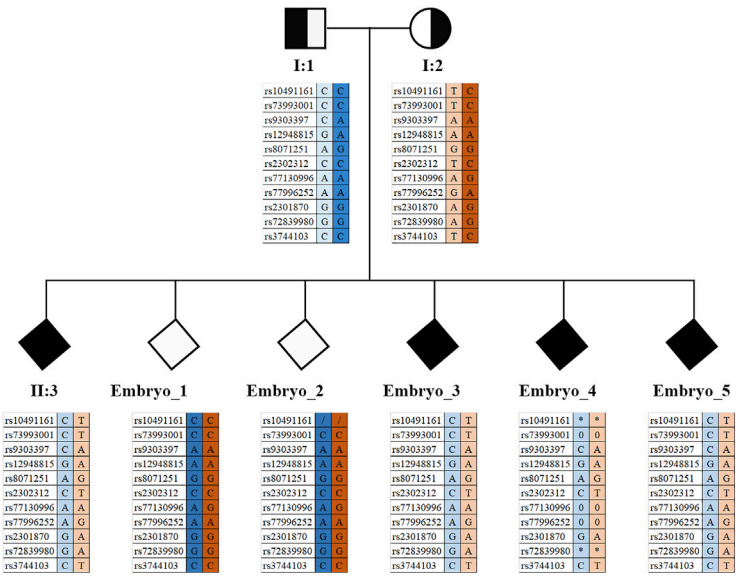


FIGURE 4 | PGT for the two *MKS1* variants. **(A)** Statistical data of the ART and PGT-M cycles of the MKS-related couple. **(B)** The images of *in vitro* early embryonic development. **(C)** Results of haplotype linkage analysis. Based on the genotype of the SNP locus of the aborted fetus II:3, embryos carrying the *MKS1* gene mutations were deduced: unknown gender without *MKS1* mutations; unknown gender with *MKS1* mutations.

MKS1 (p.R158* and p.E471Lfs*92), which disrupted the intracellular localization of MKS1 and induced defects in cilium length and the number of patient fibroblasts (Slaats et al., 2016). In the present study, the two variants (p.S117* and p.E471Lfs*92) identified in the MKS-related fetuses localized near the reported mutations and functioned through the comparable genotype–phenotype regulation model.

According to Mendel's law of inheritance, the incidence of autosomal recessive diseases in the offspring is 25%. Moreover, the frequency of genetic mutations is variable among the populations in different regions. In the assisted reproductive process, high attention is focused on preventing genetic diseases, especially autosomal recessive diseases. As a well-established alternative to invasive prenatal diagnosis, PGT for monogenic disorder (PGT-M) has evolved into an effective clinical method for MKS-related families.

In summary, we identified two novel variants, expanding the mutation spectrum of *MKS1*. Our findings further implicated that the clinical significance of *MKS1* variants needs an in-depth investigation. PGT and extended carrier screening are effective tools for genetic disease blocking in clinical applications. Together, these findings would be beneficial for the MKS patients and their families.

DATA AVAILABILITY STATEMENT

The data presented in the study are deposited in the Genome Sequence Archive for Human in China National Genomics Data Center repository, accession number HRA001710, that are publicly accessible at https://ngdc.cncb.ac.cn/gsa-human.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of the Chongqing Health Center for Women and Children. Informed consent was obtained from the proband/participants.

AUTHOR CONTRIBUTIONS

TL, DL, and GH conceived and designed the study. LS, KT, KC, and KJ carried out the experiments. YM, YX, and CJ provided the clinical samples. DZ conducted the ART cycle. TL wrote the manuscript. DL and GH critically commented on and edited the manuscript. All authors read and approved the final version of the manuscript.

FUNDING

This study was supported by grants from the Chongqing Natural Science Foundation (CSTC2021JCYJ-MSXMX0722).

ACKNOWLEDGMENTS

We thank the laboratory of *in vitro* fertilization of Chongqing Clinical Research Center for Reproductive Medicine.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.843931/full#supplementary-material

REFERENCES

Andreu-Cervera, A., Catala, M., and Schneider-Maunoury, S. (2021). Cilia, Ciliopathies and Hedgehog-Related Forebrain Developmental Disorders. *Neurobiol. Dis.* 150, 105236. doi:10.1016/j.nbd.2020.105236

Anvarian, Z., Mykytyn, K., Mukhopadhyay, S., Pedersen, L. B., and Christensen, S. T. (2019). Cellular Signalling by Primary Cilia in Development, Organ Function and Disease. *Nat. Rev. Nephrol.* 15 (4), 199–219. doi:10.1038/s41581-019-0116-9

Auber, B., Burfeind, P., Herold, S., Schoner, K., Simson, G., Rauskolb, R., et al. (2007). A Disease Causing Deletion of 29 Base Pairs in Intron 15 in the MKS1 Gene Is Highly Associated with the Campomelic Variant of the Meckel-Gruber Syndrome. *Clin. Genet.* 72 (5), 454–459. doi:10.1111/j.1399-0004.2007.00880.x

Bader, I., Decker, E., Mayr, J. A., Lunzer, V., Koch, J., Boltshauser, E., et al. (2016). MKS1 Mutations Cause Joubert Syndrome with Agenesis of the Corpus Callosum. *Eur. J. Med. Genet.* 59 (8), 386–391. doi:10.1016/j.ejmg.2016.06.007

Bergmann, C., Frank, V., and Salonen, R. (2016). Clinical Utility Gene Card for: Meckel Syndrome - Update 2016. *Eur. J. Hum. Genet.* 24 (8), 3. doi:10.1038/ejhg.2016.33

Bialas, N. J., Inglis, P. N., Li, C., Robinson, J. F., Parker, J. D., Healey, M. P., et al. (2009). Functional Interactions between the Ciliopathy-Associated Meckel Syndrome 1 (MKS1) Protein and Two Novel MKS1-Related (MKSR) Proteins. *J. Cel Sci.* 122 (Pt 5), 611–624. doi:10.1242/jcs.028621

De Rycke, M., and Berckmoes, V. (2020). Preimplantation Genetic Testing for Monogenic Disorders. *Genes (Basel)* 11 (8), 871. doi:10.3390/genes11080871

Dean, S., Moreira-Leite, F., Varga, V., and Gull, K. (2016). Cilium Transition Zone Proteome Reveals Compartmentalization and Differential Dynamics of Ciliopathy Complexes. *Proc. Natl. Acad. Sci. USA* 113 (35), E5135–E5143. doi:10.1073/pnas.1604258113

Dowdle, W. E., Robinson, J. F., Kneist, A., Sirerol-Piquer, M. S., Frints, S. G. M., Corbit, K. C., et al. (2011). Disruption of a Ciliary B9 Protein Complex Causes Meckel Syndrome. *Am. J. Hum. Genet.* 89 (1), 94–110. doi:10.1016/j.ajhg.2011.06.003

Frank, V., Ortiz Brüchle, N., Mager, S., Frints, S. G. M., Bohring, A., du Bois, G., et al. (2007). Aberrant Splicing Is a Common Mutational Mechanism inMKS1, a Key Player in Meckel-Gruber Syndrome. *Hum. Mutat.* 28 (6), 638–639. doi:10.1002/humu.9496

Garcia-Gonzalo, F. R., Corbit, K. C., Sirerol-Piquer, M. S., Ramaswami, G., Otto, E. A., Noriega, T. R., et al. (2011). A Transition Zone Complex Regulates Mammalian Ciliogenesis and Ciliary Membrane Composition. *Nat. Genet.* 43 (8), 776–784. doi:10.1038/ng.891

Gogendeau, D., Lemullois, M., Le Borgne, P., Castelli, M., Aubusson-Fleury, A., Arnaiz, O., et al. (2020). MKS-NPHP Module Proteins Control Ciliary Shedding at the Transition Zone. *Plos Biol.* 18 (3), e3000640. doi:10.1371/journal.pbio.3000640

Gonçalves, J., and Pelletier, L. (2017). The Ciliary Transition Zone: Finding the Pieces and Assembling the Gate. *Mol. Cell* 40 (4), 243–253. doi:10.14348/molcells.2017.0054

Hartill, V., Szymanska, K., Sharif, S. M., Wheway, G., and Johnson, C. A. (2017). Meckel-Gruber Syndrome: An Update on Diagnosis, Clinical Management, and Research Advances. *Front. Pediatr.* 5, 244. doi:10.3389/fped.2017.00244

Huang, L., Ma, F., Chapman, A., Lu, S., and Xie, X. S. (2015). Single-Cell Whole-Genome Amplification and Sequencing: Methodology and Applications. *Annu. Rev. Genom. Hum. Genet.* 16, 79–102. doi:10.1146/annurev-genom-090413-025352

Kyttälä, M., Tallila, J., Salonen, R., Kopra, O., Kohlschmidt, N., Paavola-Sakki, P., et al. (2006). MKS1, Encoding a Component of the Flagellar Apparatus Basal Body Proteome, Is Mutated in Meckel Syndrome. *Nat. Genet.* 38 (2), 155–157. doi:10.1038/ng1714

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

Logan, C. V., Abdel-Hamed, Z., and Johnson, C. A. (2011). Molecular Genetics and Pathogenic Mechanisms for the Severe Ciliopathies: Insights into Neurodevelopment and Pathogenesis of Neural Tube Defects. *Mol. Neurobiol.* 43 (1), 12–26. doi:10.1007/s12035-010-8154-0

Luo, M., He, R., Lin, Z., Shen, Y., Zhang, G., Cao, Z., et al. (2020). Novel Compound Heterozygous Variants in MKS1 Leading to Joubert Syndrome. *Front. Genet.* 11, 576235. doi:10.3389/fgene.2020.576235

Luo, M., Lin, Z., Zhu, T., Jin, M., Meng, D., He, R., et al. (2021). Disrupted Intraflagellar Transport Due to IFT74 Variants Causes Joubert Syndrome. *Genet. Med.* 23, 1041. doi:10.1038/s41436-021-01106-z

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

Mitchison, H. M., and Valente, E. M. (2017). Motile and Non-motile Cilia in Human Pathology: from Function to Phenotypes. *J. Pathol.* 241 (2), 294–309. doi:10.1002/path.4843

Nagy, E., and Maquat, L. E. (1998). A Rule for Termination-Codon Position within Intron-Containing Genes: when Nonsense Affects RNA Abundance. *Trends Biochem. Sci.* 23 (6), 198–199. doi:10.1016/s0968-0004(98)01208-0

Okazaki, M., Kobayashi, T., Chiba, S., Takei, R., Liang, L., Nakayama, K., et al. (2020). Formation of the B9-Domain Protein Complex MKS1-B9D2-B9D1 Is Essential as a Diffusion Barrier for Ciliary Membrane Proteins. *MBoC* 31 (20), 2259–2268. doi:10.1091/mbc.e20-03-0208

Parisi, M. A. (2019). The Molecular Genetics of Joubert Syndrome and Related Ciliopathies: The Challenges of Genetic and Phenotypic Heterogeneity. *Transl. Sci. Rare Dis.* 4 (1-2), 25–49. doi:10.3233/TRD-190041

Reiter, J. F., and Leroux, M. R. (2017). Genes and Molecular Pathways Underpinning Ciliopathies. *Nat. Rev. Mol. Cel Biol.* 18 (9), 533–547. doi:10.1038/nrm.2017.60

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

Romani, M., Micalizzi, A., Kraoua, I., Dotti, M. T., Cavallin, M., Sztriha, L., et al. (2014). Mutations in B9D1 and MKS1 Cause Mild Joubert Syndrome: Expanding the Genetic Overlap with the Lethal Ciliopathy Meckel Syndrome. *Orphanet J. Rare Dis.* 9, 72. doi:10.1186/1750-1172-9-72

Salonen, R., Opitz, J. M., and Reynolds, J. F. (1984a). The Meckel Syndrome: Clinicopathological Findings in 67 Patients. *Am. J. Med. Genet.* 18 (4), 671–689. doi:10.1002/ajmg.1320180414

Salonen, R., Norio, R., Opitz, J. M., and Reynolds, J. F. (1984b). The Meckel Syndrome in Finland: Epidemiologic and Genetic Aspects. *Am. J. Med. Genet.* 18 (4), 691–698. doi:10.1002/ajmg.1320180415

Slaats, G. G., Isabella, C. R., Kroes, H. Y., Dempsey, J. C., Gremmels, H., Monroe, G. R., et al. (2016). MKS1 Regulates Ciliary INPP5E Levels in Joubert Syndrome. *J. Med. Genet.* 53 (1), 62–72. doi:10.1136/jmedgenet-2015-103250

Wu, Z., Pang, N., Zhang, Y., Chen, H., Peng, Y., Fu, J., et al. (2020). CEP290 Is Essential for the Initiation of Ciliary Transition Zone Assembly. *Plos Biol.* 18 (12), e3001034. doi:10.1371/journal.pbio.3001034

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: 24 March 2022
doi: 10.3389/fgene.2022.815727

SNP–SNP Interactions of Surfactant Protein Genes in Persistent Respiratory Morbidity Susceptibility in Previously Healthy Children

*Chintan K. Gandhi[1], Neal J. Thomas[1], Ye Meixia[2], Debbie Spear[1], Chenqi Fu[3], Shouhao Zhou[3], Rongling Wu[3], Garrett Keim[4], Nadir Yehya[4] and Joanna Floros[1,5]**

[1]Center for Host Defense, Inflammation, and Lung Disease (CHILD) Research, Department of Pediatrics, Pennsylvania State University College of Medicine, Hershey, PA, United States, [2]Center for Computational Biology, College of Biological Sciences and Technology, Beijing Forestry University, Beijing, China, [3]Public Health Science, Pennsylvania State University College of Medicine, Hershey, PA, United States, [4]Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States, [5]Department of Obstetrics and Gynecology, Pennsylvania State University College of Medicine, Hershey, PA, United States

OPEN ACCESS

Edited by:
Maria Elisabetta Baldassarre, University of Bari Aldo Moro, Italy

Reviewed by:
Robert Fred Clark, RTI International, United States
John House, National Institute of Environmental Health Sciences (NIH), United States

***Correspondence:**
Joanna Floros
jfloros@psu.edu

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *15 November 2021*
Accepted: *21 February 2022*
Published: *24 March 2022*

Citation:
Gandhi CK, Thomas NJ, Meixia Y, Spear D, Fu C, Zhou S, Wu R, Keim G, Yehya N and Floros J (2022) SNP–SNP Interactions of Surfactant Protein Genes in Persistent Respiratory Morbidity Susceptibility in Previously Healthy Children. Front. Genet. 13:815727. doi: 10.3389/fgene.2022.815727

We studied associations of persistent respiratory morbidity (PRM) at 6 and 12 months after acute respiratory failure (ARF) in previously healthy children with single-nucleotide polymorphisms (SNPs) of surfactant protein (SP) genes. Of the 250 enrolled subjects, 155 and 127 were followed at 6 and 12 months after an ARF episode, respectively. Logistic regression analysis and SNP–SNP interaction models were used. We found that 1) in the multivariate analysis, an increased risk at 6 and 12 months was associated with rs1124_A and rs4715_A of *SFTPC*, respectively; 2) in a single SNP model, increased and decreased risks of PRM at both timepoints were associated with rs1124 of *SFTPC* and rs721917 of *SFTPD*, respectively; an increased risk at 6 months was associated with rs1130866 of *SFTPB* and rs4715 of *SFTPC*, and increased and decreased risks at 12 months were associated with rs17886395 of *SFTPA2* and rs2243639 of *SFTPD*, respectively; 3) in a two-SNP model, PRM susceptibility at both timepoints was associated with a number of intergenic interactions between SNPs of the studied SP genes. An increased risk at 12 months was associated with one intragenic (rs1965708 and rs113645 of *SFTPA2*) interaction; 4) in a three-SNP model, decreased and increased risks at 6 and 12 months, respectively, were associated with an interaction among rs1130866 of *SFTPB*, rs721917 of *SFTPD*, and rs1059046 of *SFTPA2*. A decreased risk at 6 months was associated with an interaction among the same SNPs of *SFTPB* and *SFTPD* and the rs1136450 of *SFTPA1*. The findings revealed that SNPs of all *SFTP*s appear to play a role in long-term outcomes of ARF survivors and may serve as markers for disease susceptibility.

Keywords: persistent respiratory morbidity, long-term outcomes of pediatric acute respiratory failure, SNP–SNP interaction, surfactant protein genetic variant, pediatric acute respiratory failure

INTRODUCTION

Acute respiratory failure (ARF) is a common cause of invasive mechanical ventilation need and admission to pediatric intensive care units (PICUs) in children with an incidence of 3% of total PICU admissions (Ibiebele et al., 2018; Khemani et al., 2019). Recent advances in critical care that use early lung protective strategies and improvement in supportive care have led to a gradual decrease in mortality of pediatric ARF (Matthay et al., 2017). This has shifted the focus from mortality to new morbidities in this cohort (Keim et al., 2018). Studies have shown a significant decline in the functional status of pediatric ARF survivors at discharge (Pollack et al., 2009). More specifically, persistent respiratory morbidity (PRM) occurred after 6 and 12 months of an ARF episode even in previously healthy children (Keim et al., 2020). In addition, there is a considerable heterogeneity in the progression of the disease and long-term outcomes of pediatric ARF patients (Keim et al., 2020), indicating a complex interaction between genetic and environmental factors. Nonetheless, studies of long-term sequelae of ARF in children are limited (Yehya and Thomas, 2016). To our knowledge, no studies have specifically examined the role of genetics, an important host variable, as a risk factor for PRM after an episode of ARF in previously healthy children.

Pulmonary surfactant consists of 90% lipids and 10% surfactant proteins (SPs). There are two major types of SPs in the lung; 1) The hydrophobic surfactant proteins (SP-B and -C) are responsible for reducing the surface tension and essential for normal lung function (Serrano and Perez-Gil, 2006), and 2) the hydrophilic SPs (SP-A and -D) are responsible primarily for innate immunity and host defense against infections (Wright, 2005; Kishore et al., 2006; Depicolzuane et al., 2021; Floros et al., 2021). SP-B, SP-C, and SP-D are each encoded by a single gene, *SFTPB*, *SFTPC*, and *SFTPD*, respectively, whereas SP-A is encoded by two similar genes, *SFTPA1* and *SFTPA2*, that are differentially regulated (Floros and Tsotakos, 2021) and identified with functional, structural, and other differences (Thorenoor et al., 2019; Gandhi et al., 2020b; Thorenoor et al., 2020; Xu et al., 2020; Floros et al., 2021). Several single-nucleotide polymorphisms (SNPs) have been described for each of these genes (DiAngelo et al., 1999; Wert et al., 2009; Silveyra and Floros, 2012). These SNPs are common in the general population and shown to associate with various acute and chronic pulmonary diseases, such as neonatal respiratory distress syndrome (RDS) (Kala et al., 1998; Nogee et al., 2000; Rämet et al., 2000; Floros et al., 2001), cystic fibrosis (Lin et al., 2018), acute respiratory distress syndrome (Lin et al., 2000b), chronic obstructive pulmonary disease (Seifart et al., 2002), interstitial pulmonary fibrosis (Selman et al., 2003), severity of respiratory syncytial virus (RSV) (Thomas et al., 2009), tuberculosis (TB) (Floros et al., 2000), and hypersensitivity pneumonitis (HP) (Gandhi et al., 2021). Importantly, we previously demonstrated that these SNPs are associated with pediatric ARF and its short-term outcome, pulmonary dysfunction. at discharge in the same cohort (Gandhi et al., 2020a).

Taken together, we postulated that the SPs contribute to the progression of pediatric ARF and its long-term outcome, PRM, at 6 and 12 months after the index admission for ARF. To eliminate potential confounding contribution of other chronic illnesses to long-term sequelae of pediatric ARF, we enrolled only previously healthy children for the current study. We hypothesized that multiple genetic variants of the SP genes are associated with long-term outcomes after an ARF episode through single genetic variations within a gene, and/or through intragenic (within the same gene) or intergenic (with different genes) interactions. To our knowledge, this is the first study examining the association of genetic variants in PRM after an admission for ARF in previously healthy children. Our results indicate the association of complex SNP–SNP interactions of the surfactant protein genes with PRM at 6 and 12 months, and may contribute to the pulmonary sequelae in pediatric ARF survivors.

SUBJECTS AND METHODS

Study Population

We prospectively enrolled 250 previously healthy children from 0 to 24 months of age that required invasive ventilation for an index case of ARF secondary to respiratory illness at 10 participating pediatric intensive care units (PICUs) over 5 consecutive years. This multicenter cohort has been described in detail elsewhere (Gandhi et al., 2020a; Keim et al., 2020). Briefly, previously healthy children, who met at least one of the three criteria, 1) chest radiograph with either focal or diffuse infiltrative pulmonary process, 2) radiographic evidence of air trapping, or 3) clinical exam findings of lower respiratory tract illness, were determined to have primary respiratory cause of ARF. We prospectively collected all demographic and clinical data for children with ARF.

These subjects were followed up at 6 (n = 155) and 12 months (n = 127) after the index ARF admission via telephonic interview of a designated parent about the subject's health status. Questions included the 11-item PedsQL™ asthma module health-related quality of life symptom scale (Chan et al., 2005; Greenley et al., 2008; Seid et al., 2010). The parents' responses were recorded on a scale of 0–4, where 0 = never and 4 = almost always. Parents were also asked about prescribed medications, frequency of use, and whether the child had been diagnosed with asthma, and/or had visits to the physician's office or emergency department or had been readmitted to the hospital, PICU for "breathing problems," and finally, if the child required mechanical ventilation post index admission.

Cases: children, at 6 and 12 months of discharge, who developed PRM as defined *a priori,* i.e., if the subject met one of the following criteria: 1) diagnosis of asthma, 2) use of bronchodilator in the last month, 3) use of inhaled corticosteroid, 4) representation to care for a "breathing"-related complaint, or 5) asthma module health-related quality of life symptom scale score ≥5. The cohort of the current study differs from the original ARF cohort in terms of chronicity and long-term respiratory symptoms. In other words, the initial incident is defined as ARF; however, ~45% of the ARF children continue to have breathing symptoms and get diagnosed with PRM. Thus, all children with PRM had an

episode of ARF, but not all patients with ARF developed PRM. Controls: children who did not meet predefined criteria of PRM at 6 and 12 months following an index admission to PICU for ARF.

We collected blood samples of the study participants after obtaining informed consent from a parent or legal guardian. This study was approved by the institutional review board of participating sites.

DNA Isolation and Genotype Analysis

Genomic DNAs were extracted from blood samples using QIAamp Blood kit (Qiagen, Valencia, CA, USA) as described previously (DiAngelo et al., 1999). We used the polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) method to analyze the *SFTPA1*, *SFTPA2*, *SFTPD* (DiAngelo et al., 1999; Lin et al., 2000b), *SFTPB* (Lin et al., 2000a; Lin et al., 2000b), and *SFTPC* (Selman et al., 2003) gene polymorphisms as described earlier (DiAngelo et al., 1999). The PCR primer sequences and restriction enzymes used for the current study are described elsewhere (DiAngelo et al., 1999; Gandhi et al., 2020a; Gandhi et al., 2021). A total of 14 target SNPs of surfactant protein genes *SFTPA1*, *SFTPA2*, *SFTPB*, *SFTPC*, and *SFTPD* were selected based on their associations with various acute and chronic pulmonary diseases (Lin et al., 2000b; Floros et al., 2000; Floros et al., 2001; Selman et al., 2003; Thomas et al., 2009; Silveyra and Floros, 2012; Lin et al., 2018; Gandhi et al., 2020a; Gandhi et al., 2021). These include: five SNPs from *SFTPA1*: rs1059047, rs1136450, rs1136451, rs1059057, and rs4253527; four SNPs from *SFTPA2*: rs1059046, rs17886395, rs1965707, and 1965708; one SNP from *SFTPB*: rs1130866; two SNPs from *SFTPC*: rs4715 and rs1124; and two SNPs from *SFTPD*: rs721917 and rs2243639. The details of the studied SNPs are given in **Supplementary Table S1**. The SP-A1 and SP-A2 genotypes were assigned as described (DiAngelo et al., 1999). To reduce bias in the genotype, all samples were processed together in a blinded fashion with those assigning genotypes unaware of the clinical status.

Statistical Analysis

The frequency of the alleles in the two groups were compared using the Chi-square test, or the Fisher's exact test when the expected frequency of the allele was too small (<5). Assuming no allele dose–effect, univariate logistic regression was applied to each allele or SP-A genotype to test whether the existence of a given minor allele and/or genotype distinguishes PRM from no PRM. Alleles that were significantly associated with PRM in univariate analysis (p-value < 0.1) were included in the multivariate logistic regression analysis (Floros et al., 2000; Selman et al., 2003). The univariate analysis was done for screening and selection of variables for the multivariate analysis; therefore, the relaxed p-value of less than 0.1 was used. In the multivariate analysis of PRM at 12 months, a positive bacterial culture on admission and PRM at 6 months were obliged to be included in the model due to their significant associations in the univariate analysis. Variable selection was performed using a backward elimination method with a prespecified significance level of 0.05.

Wang et al. (2010) developed a computational model for detecting additive, dominant, and epistatic effects by integrating quantitative genetic theory into a case-control design context. This model can particularly characterize high-order epistatic interactions even with the modest sample size; hence, we used this model (Wang et al., 2010) to study associations of SP gene polymorphisms with PRM at 6 and 12 months (Wang et al., 2010; Gandhi et al., 2020a; Gandhi et al., 2021). Of note, in the present study, the reference (major) and alternate (minor) alleles were assigned based on the "NCBI dbSNP database of genetic variation" using the global population (Sherry et al., 2001), and the significant findings were noted in terms of the reference to the minor allele in its homozygous or heterozygous form. The model of Wang et al. dissects the genetic effects, including the additive (a) and dominant (d) of the minor allele at a single SNP, pairwise interaction effects at two SNPs, and three-way interactions in a three-SNP model.

An example with a detailed explanation is provided below in order to understand the additive and dominant effects of each SNP in a given interaction type. Please consider the example of an SNP with three genotypes AA, Aa, and aa. To estimate its additive effect, the homozygotes (AA and aa) were compared against its heterozygote (Aa), whereas to estimate its dominant effect, the heterozygote (Aa) was compared against the average size of the two homozygotes (AA and aa). Thus, the interaction type "a1d2" in a two-SNP model [with the first SNP with three possible genotypes (AA, Aa, and aa) and the second SNP with three possible genotypes (BB, Ba, and bb)] can be interpreted as follows: two-locus genotypes with a homozygote at the first locus and heterozygote at the second locus, i.e., AABb, aaBb, perform differently than the remaining genotypes (AABB, AAbb, AaBB, AaBb, Aabb, aaBB, and aabb). According to this model, we sorted the case-control genotype observations into a 2×2 contingency table to examine the association of each of the genetic effects of individual SNPs with PRM at 6 and 12 months.

The logistic regression model was implemented to estimate the genetic effect of that particular SNP after adjusting for covariates (age, sex, race, and weight). These variables were selected based on the biological possibilities and the significant difference between groups. We used the race as a covariate to adjust for differences in allele frequencies between races. The OR with 95% confidence interval (95% CI) was estimated using the Cochran's and Mantel–Haenszel tests to assess the magnitude of the dominant/additive effect (Day and Byar, 1979). The false discovery rate (FDR) was controlled at 5% using the Benjamini–Hochberg method to account for multiple testing (Hope, 1968; Hochberg, 1995). We reported all possible SNP–SNP interactions associated with cases with p-value < 0.05 for single SNPs and two- and three-SNP interaction models.

RESULTS

Clinical Characteristics of the Study Group

Of the 250 patients enrolled in the study, follow-up questionnaires were completed for 155 patients (~61%) and 127 patients (~50%) at 6 and 12 months, respectively. Persistent respiratory morbidity was diagnosed in 66 patients (42.5%) at 6 months and in 57 patients

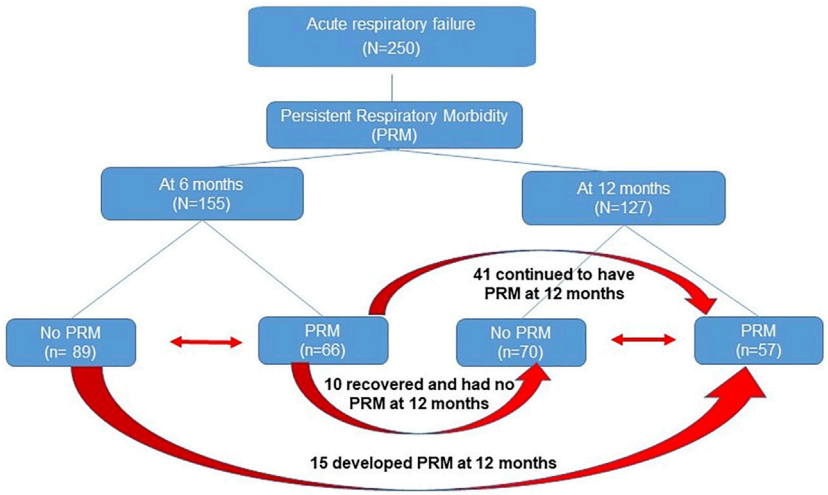


FIGURE 1 | Flow diagram of patients with persistent respiratory morbidity (PRM). Red arrow depicts transfer of patients from one group to another.

TABLE 1 | Demographics and clinical characteristics of the study group at 6 and 12 months.

Variable	At 6 months			At 12 months		
	No PRM (*n* = 89)	PRM (*n* = 66)	*p*-Value	No PRM (*n* = 70)	PRM (*n* = 57)	*p*-Value
Demographics						
Age (months)	3 ± 4.4	4.2 ± 4.8	0.123	2.8 ± 3.9	4.4 ± 5.6	0.07
Female (%)	30 (34)	24 (36)	0.734	23 (33)	21 (37)	0.6
Non-White race (%)	29 (33)	17 (26)	0.177	22 (31)	19 (33)	0.56
Hispanic (%)	14 (16)	14 (21)	0.384	12 (17)	9 (16)	0.84
Admission diagnosis (%)			0.387			0.492
RSV bronchiolitis	50 (56)	36 (54)		37 (53)	32 (56)	
Other bronchiolitis	17 (19)	10 (15)		14 (20)	11 (19)	
Other pneumonia	8 (9)	11 (17)		5 (7)	7 (12)	
Other respiratory failure	11 (12)	9 (14)		12 (17)	6 (11)	
Nonpulmonary	3 (4)	0		2 (3)	1 (2)	
Positive bacterial culture (%)	39 (44)	41 (62)	0.04	26 (37)	36 (63)	0.001
PDAD (%)	19 (21)	31 (47)	0.001	19 (27)	21 (37)	0.245
PRM at 6 months (%)	-	-	-	10 (14)	41 (72)	1.6924E–13

Note. PRM, persistent respiratory morbidity; PDAD, pulmonary dysfunction at discharge; RSV, respiratory syncytial virus.

(44.8%) at 12 months. **Figure 1** shows the flow diagram of patients with PRM. We did not observe statistically significant difference in age, sex, race, and ethnicity between groups at both timepoints as shown in **Table 1**. As shown previously in our clinical paper (Keim et al., 2020), PRM at 6 months was predictive of developing PRM at 12 months, whereas a positive respiratory bacterial culture during the index admission was predictive of developing PRM at both timepoints.

Association of Surfactant Protein Single-Nucleotide Polymorphisms With Persistent Respiratory Morbidity

Univariate and Multivariate Analyses

At 6 and 12 months, no significant differences were observed in the frequency of the studied SNPs between the two groups (PRM vs. no PRM). The frequency distribution of the majority of SNPs did not deviate from the Hardy–Weinberg equilibrium (**Supplementary Tables S2 and S3**). An increased risk of PRM at 6 months was significantly associated with rs1124 of the *SFTPC* in the univariate and multivariate logistic regression analysis, OR = 11.7 (1.9–217.9), p = 0.03 (**Supplementary Table S4**). At 12 months, significant differences (p < 0.1) were observed for the SFTPA2 marker allele (rs17886395_G), the SF*TPD* marker allele (rs721917_G), the *SFTPC* marker alleles (rs4715_A, rs1124_A), and SFTPA1 (6A3) in the univariate analysis. Of these, based on an OR <1, a decreased risk for PRM was associated with rs721917_G of the *SFTPD* and SFTPA1 (6A3), whereas an increased risk for PRM was associated with other marker alleles (OR >1) (**Table 2**). When these marker alleles were considered in the multivariate analysis, an increased risk for PRM was significantly associated with only one allele, the *SFTPC* (rs4715_A), OR = 3 (1.14–9.5), p = 0.04 (**Table 2**).

TABLE 2 | Persistent respiratory morbidity (PRM) vs. no PRM at 12 months (univariate analysis).

Gene	SNP	Chr	Position	Allele	PRM	No PRM	OR (95%CI)	*p*-Value	OR (95%CI)*	*p*-Value*
					n (%)	*n* (%)				
SFTPA2	rs17886395	10	AA91	G	30 (26)	23 (16)	2 (0.91–4.5)	0.09	2.02(0.70–6.05)	0.2
SFTPD	rs721917	10	AA11	G	57 (50)	60 (43)	0.3 (0.11–0.93)	0.04	0.4 (0.10–1.71)	0.2
SFTPC	rs4715[#]	8	AA138	A	21 (19)	38 (27)	2 (1–4.2)	0.06	3 (1.14–9.5)	0.04
SFTPC	rs1124	8	AA186	A	28 (25)	52 (37)	1.9 (1.04–3.6)	0.04	2.4 (1.03–6.2)	0.05
SFTPA1	$6A^3$	10			34 (25)	23 (21)	0.66 (0.32–1.37)	0.27	0.36 (0.11–1.05)	0.07

*Note. Chr, chromosome; AA, amino acid; n (%). number of the given allele, in parenthesis the percentage of the given allele out of the possible alleles in the particular cohort is shown. *adjusted for PRM at 6 months and positive bacterial culture. [#]remained significant in the multivariate analysis. OR, odds ratio; CI, confidence interval.*

TABLE 3 | Association of surfactant protein (SP) gene single-nucleotide polymorphisms (SNPs) with persistent respiratory morbidity (PRM) at 6 and 12 months in a single-SNP model after adjusting for covariates (age, sex, race, and weight).

SNP	Gene	Allele	Interaction type	PRM at 6 months			PRM at 12 months		
				p-Value	FDR	OR (95%CI)	*p*-Value	FDR	OR (95%CI)
rs1124	*SFTPC*	A	Additive	3.20E–03	0.02	5.8 (1.8–19.3)	0.0025	0.02	6.1 (1.9–19.8)
rs721917	*SFTPD*	G	Dominant	9.00E–04	0.01	0.5 (0.3–0.8)	3.51E–05	0.001	0.3 (0.2–0.5)
rs1130866	*SFTPB*	C	Additive	4.00E–04	0.01	3.2 (1.2–8.6)			
rs4715	*SFTPC*	A	Additive	4.30E–03	0.02	6.2 (1.4–27.4)			
rs17886395	*SFTPA2*	G	Dominant				0.001	0.02	2.0 (1.1–3.8)
rs2243639	*SFTPD*	C	Dominant				0.001	0.02	0.4 (0.2–0.8)

Note. OR, odds ratio; CI, confidence interval; FDR, false discovery rate.

Single-Single-Nucleotide Polymorphism Model

At 6 and 12 months, an increased risk of PRM was associated with rs1124, OR = 5.8 (1.8–19.3) of the *SFTPC* that exhibited an additive effect, whereas a decreased risk of PRM was associated with rs721917 of the *SFTPD* that exhibited a dominant effect, $p < 0.05$. Only at 6 months, an increased risk of PRM was associated with rs1130866 of the *SFTPB*, OR = 3.2 (1.2–8.6), and the rs4715 of the *SFTPC*, OR = 6.2 (1.4–27.4), and each exhibited an additive effect. Only at 12 months, increased and decreased risks of PRM were associated with rs17886395 of the *SFTPA2* and rs2243639 of the *SFTPD*, respectively, and exhibited a dominant effect (**Table 3**).

For SNP–SNP interaction tables, the column "interaction type" represents interactions that could be intragenic, i.e., between SNPs of an individual gene, or intergenic, i.e., between SNPs of different genes. The letter "a" is for additive and "d" is for dominant effect of that particular SNP. The number following "a" or "d" indicates the position of the corresponding SNP, for example, an interaction of the a1d2 type indicates the additive and dominant effects of SNP 1 and SNP 2, respectively, in the two-SNP model. For the three-SNP model, the a1a2d3 interaction type indicates the additive effects of SNPs 1 and 2, and the dominant effect of SNP 3.

Two-Single-Nucleotide Polymorphism Model

At 6 months, decreased risk of PRM was associated with 12 interactions of different combinations between SNPs of the studied genes in a two-SNP model, OR = 0.1–0.5 (**Table 4**). All interactions were intergenic (between SNPs of different genes). The majority of significant interactions involved the rs1130866 of the *SFTPB* ($n = 7$) and interacted with SNPs of hydrophilic SPs ($n = 6$). We observed four and two interactions between SNPs of the hydrophilic and hydrophobic SPs alone, respectively.

At 12 months, PRM was associated with a total of 29 interactions among SNPs of SP genes in a two-SNP model (**Table 5**). All but one interactions were intergenic. The one intragenic interaction (SNPs of the same gene) was between SNPs of the *SFTPA2* (rs1059046 × rs1965707, a1d2, OR = 2.9 (1.1–7.8), $p < 0.05$). Significant intergenic interactions ($n = 28$) that included the other studied genes were as follows: 15, 10, 8, and 5 interactions for each *SFTPA1*, *SFTPA2*, *SFTPB*, and *SFTPC*, respectively. The *SFTPD* SNPs had the highest number of interactions with SNPs of other SPs ($n = 18$), particularly the rs721917 of the *SFTPD* ($n = 13$).

Out of the 29 interactions, 11 were among SNPs of both hydrophilic and hydrophobic SPs, and 17 and 1 were between SNPs of the hydrophilic and hydrophobic SPs alone, respectively. A decreased risk of PRM was associated with the majority of the interactions, whereas an increased risk of PRM was associated with only 10 interactions. Of note, the susceptibility to PRM changed based on the effect of a particular SNP in a given interaction. For example, if the increased risk of PRM was associated with the interaction (rs1059046 × rs1124), and the interaction type was a1a2, OR = 5.3 (1.1–25.1), it would indicate that the additive effects of both SNPs were associated with increased risk. However, the decreased risk of PRM was associated with the same interaction, if the interaction type

TABLE 4 | Associations of SP gene SNP interactions with persistent respiratory morbidity (PRM) at 6 months in a two-SNP model after adjusting for covariates (age, sex, race, and weight).

SNP#1	Gene	SNP#2	Gene	Interaction type	*p*-Value	FDR	OR (95% CI)
rs1136451	*SFTPA1*	rs721917	*SFTPD*	a1	3.27E–04	0.0020	0.1 (0.1–0.4)
				d2	3.05E–05	0.0007	0.5 (0.2–0.9)
rs1130866	*SFTPB*	rs2243639	*SFTPD*	a1	1.55E–03	0.0050	0.2 (0.1–0.9)
				d1d2	6.68E–06	0.0006	0.5 (0.3–0.8)
rs1965707	*SFTPA2*	rs1130866	*SFTPB*	a2	2.12E–04	0.0017	0.3 (0.1–0.9)
rs1136451	*SFTPA1*	rs1130866	*SFTPB*	a2	9.22E–05	0.0011	0.2 (0.1–0.7)
rs1965707	*SFTPA2*	rs1136450	*SFTPA1*	a1d2	6.26E–05	0.0010	0.4 (0.2–0.9)
rs1136450	*SFTPA1*	rs2243639	*SFTPD*	a1d2	1.65E–03	0.0051	0.3 (0.1–0.8)
				a1a2	3.12E–02	0.0430	0.2 (0.1–0.8)
rs1130866	*SFTPB*	rs4715	*SFTPC*	a1d2	3.14E–05	0.0007	0.3 (0.1–0.6)
rs1130866	*SFTPB*	rs1124	*SFTPC*	a1d2	5.85E–06	0.0006	0.3 (0.1–0.8)
rs1130866	*SFTPB*	rs721917	*SFTPD*	a1d2	3.46E–04	0.0020	0.5 (0.2–0.9)
rs1965708	*SFTPA2*	rs721917	*SFTPD*	d1d2	5.88E–04	0.0027	0.6 (0.4–0.9)
rs1059046	*SFTPA2*	rs1130866	*SFTPB*	a2	5.74E–03	0.0131	0.2 (0.1–0.8)
rs1059046	*SFTPA2*	rs721917	*SFTPD*	d2	6.92E–03	0.0150	0.4 (0.2–0.8)

Note. Interaction type: a and d denote additive and dominant effects of the particular SNP. Numbers 1 and 2 denote effect of the particular SNP at that position. For example, the rs1965707 × rs1136450 interaction is a1d2 type indicating an additive effect of the rs1965707 and a dominant effect of the rs1136450. This interaction is associated with a decreased risk of PRM at 6 months. In some interactions, only one SNP exhibited a main effect, whereas, the other SNP remained silent but their interaction was significant. For example, the rs1136451 × rs721917 interaction shows two significant effect types a1 and d2 with the rs1136451 exhibiting a main additive effect and the rs72191 exhibiting a dominant effect.

was d1a2, OR = 0.4 (0.2–0.8), this indicates that the dominant effect of rs1059046 and the additive effect of rs1124 are associated with decreased risk. In addition, 9 out of the 12 significant interactions associated with a decreased risk of PRM at 6 months remained significant at 12 months as well.

Three-Single-Nucleotide Polymorphism Model

At 6 and 12 months, the rs1130866 of the *SFTPB*, and the rs721917 of the *SFTPD* interacted with the rs1059046 of the *SFTPA2* in a three-SNP model (**Table 6**). A decreased risk of PRM at 6 months was associated with these intergenic interactions. However, an increased risk of PRM at 12 months was associated with the same interactions. In addition, a decreased risk of PRM only at 6 months was associated with interactions among the same SNPs of the *SFTPB* and *SFTPD* (noted above) and the rs1136450 of the *SFTPA1*. Furthermore, as shown in **Table 6**, the effect size of the seven intergenic interactions. as denoted by the OR, was variable, based on the effect (additive or dominant) of the particular SNP at the particular position, OR = 0.07–0.25 (**Table 6**).

DISCUSSION

Surfactant dysfunction and dysregulated inflammation, individually or in conjunction with each other, are central to the pathophysiologic mechanisms of various pulmonary diseases, including ARF in children (Amigoni et al., 2017). Because SPs play a role in surfactant dysfunction and/or regulation of inflammatory processes/innate immunity, we hypothesized that natural genetic variants of SPs are associated with PRM at 6 and 12 months after an ARF episode. The results indicated that 1) PRM at both timepoints is associated with SNPs of all five SP genes. 2) Increased risk of PRM is associated with rs1124 of the *SFTPC* at 6 months in the univariate and multivariate analyses and in the single-SNP model, whereas increased risk of PRM at 12 months is associated with rs4715 of the *SFTPC* in the univariate and multivariate analyses. 3) At both timepoints, increased and decreased risks of PRM is associated with rs1124 of the *SFTPC* and the rs721917 of the *SFTPD*, respectively, in the single-SNP model. 4) Increased and decreased risks of PRM at 12 months is associated with rs17886395 of the SFTPA2 and rs2243639 of the *SFTPD*, respectively, in the single-SNP model. 5) PRM at 12 months is associated with one significant intragenic interaction between SNPs of the *SFTPA2* (rs1059046 × rs1965707). 6) No association between increased and decreased risks of PRM at 6 and 12 months, respectively, was observed with any of the SNP–SNP interactions in the two- and three-SNP model (6 months) or the three-SNP model (12 months). 7) In the three-SNP model, one intergenic interaction (rs1059046 × rs1130866 × rs721917) is associated with decreased and increased risk of PRM at 6 and 12 months, respectively. 8) The intergenic (rs1136450 × rs1130866 × rs721917) interaction is associated with a decreased risk of PRM at 6 months with a variable effect size.

We used two different statistical methods to study associations of SNPs of the SP genes with PRM at 6 and 12 months. The first one is the multivariate logistic regression analysis adjusting for selected covariates using backward elimination method ($p < 0.1$). The other method is the Wang's SNP–SNP interaction model, an integrated approach, which uses principles of quantitative genetics to decompose the genetic effect of a particular SNP into its underlying components (Wang et al., 2010). In this analysis, the covariates (age, sex, race, and weight) were selected based on the biological possibilities and the differences between the two groups (cases vs. controls). The marker alleles shown to associate with risk of PRM at 6 and 12 months are almost identical (based on the OR) to those observed in the univariate and multivariate analyses. These

TABLE 5 | Associations of SP gene SNPs with persistent respiratory morbidity (PRM) at 12 months in a two-SNP model after adjusting for covariates (age, sex, race, and weight).

SNP#1	Gene	SNP#2	Gene	Interaction type	*p*-Value	FDR	OR (95% CI)
rs1059046	*SFTPA2*	rs1124	*SFTPC*	d1a2	4.92E–04	0.001051	0.5 (0.3–0.9)
rs1130866	*SFTPB*	rs1124	*SFTPC*	d1a2	1.68E–05	0.000102	0.4 (0.2–0.7)
rs1136450	*SFTPA1*	rs1124	*SFTPC*	a2	8.50E–05	0.000288	0.1 (0.1–0.6)
				d1a2	3.33E–04	0.000792	0.6 (0.3–0.9)
rs1965707	*SFTPA2*	rs1124	*SFTPC*	a2	8.62E–04	0.001716	0.2 (0.1–0.6)
rs1059046	*SFTPA2*	rs1130866	*SFTPB*	a2	5.38E–03	0.007326	0.3 (0.1–0.9)
rs1136451	*SFTPA1*	rs1130866	*SFTPB*	d1d2	5.35E–06	5.83E–05	0.6 (0.4–0.9)
rs1965707	*SFTPA2*	rs1130866	*SFTPB*	a2	1.61E–05	0.000101	0.3 (0.1–0.7)
				a1d2*	4.38E–05	0.000165	1.8 (1.1–3.1)
rs1965707	*SFTPA2*	rs1136450	*SFTPA1*	a1d2	9.22E–07	1.91E–05	0.4 (0.2–0.7)
				a1*	5.17E–04	0.001092	6.6 (1.9–22.6)
				d1a2	1.22E–02	0.014944	0.5 (0.3–0.9)
rs1965708	*SFTPA2*	rs1136450	*SFTPA1*	a1d2*	1.74E–03	0.003022	1.8 (1.1–3.1)
				d1a2	5.20E–03	0.007177	0.5 (0.3–0.9)
rs1059046	***SFTPA2***	**rs1965707**	***SFTPA2***	**a1d2***	**2.31E–03**	**0.003708**	**1.9 (1.1–3.5)**
rs1124	*SFTPC*	rs2243639	*SFTPD*	d2	1.27E–03	0.002345	0.5 (0.3–0.8)
rs1130866	*SFTPB*	rs2243639	*SFTPD*	d1d2	2.38E–08	1.64E–06	0.5 (0.3–0.7)
				a1	4.38E–04	0.000965	0.2 (0.1–0.6)
				a2*	2.99E–03	0.004446	4.5 (1.2–16.4)
rs1136450	*SFTPA1*	rs2243639	*SFTPD*	a1d2	1.53E–04	0.000421	0.3 (0.2–0.6)
rs4715	*SFTPC*	rs2243639	*SFTPD*	d2	1.05E–04	0.000328	0.6 (0.4–0.9)
rs1130866	*SFTPB*	rs4715	*SFTPC*	d1a2	6.21E–04	0.001261	0.6 (0.3–0.9)
rs1059046	*SFTPA2*	rs721917	*SFTPD*	d2	1.98E–05	0.000103	0.3 (0.2–0.5)
				a1a2*	3.13E–03	0.004568	4.1 (1.3–13.6)
				a1d2*	9.75E–03	0.012155	1.9 (1.1–3.2)
rs1124	*SFTPC*	rs721917	*SFTPD*	a1	3.13E–05	0.000145	0.1 (0.1–0.8)
				d2	3.62E–04	0.000848	0.5 (0.3–0.8)
				d1d2	4.25E–05	0.000165	0.7 (0.5–0.9)
rs1130866	*SFTPB*	rs721917	*SFTPD*	a1d2	3.10E–06	4.35E–05	0.4 (0.2–0.7)
				d2	3.45E–05	0.000149	0.4 (0.2–0.8)
				d1a2	5.93E–04	0.001227	0.5 (0.3–0.9)
				d1d2	4.34E–05	0.000165	0.7 (0.5–0.9)
rs1136450	*SFTPA1*	rs721917	*SFTPD*	d2	4.24E–04	0.000963	0.4 (0.2–0.8)
				a1a2	1.49E–03	0.002732	0.1 (0.1–0.4)
rs1136451	*SFTPA1*	rs721917	*SFTPD*	a1	3.15E–06	4.35E–05	0.1 (0.1–0.3)
				d2	5.01E–09	1.01E–06	0.3 (0.2–0.6)
				a1d2	3.99E–07	1.03E–05	0.5 (0.3–0.9)
				d1d2*	6.90E–05	0.000251	1.6 (1.2–2.3)
rs1965707	*SFTPA2*	rs721917	*SFTPD*	d2	1.10E–05	7.60E–05	0.4 (0.3–0.8)
				a1d2	1.84E–05	0.000103	0.5 (0.3–0.8)
rs1965708	*SFTPA2*	rs721917	*SFTPD*	d1d2	3.50E–06	4.42E–05	0.6 (0.5–0.9)
				d2	3.69E–03	0.005302	0.5 (0.3–0.9)
				a1d2*	6.47E–03	0.008473	1.8 (1.1–3.1)
rs4715	*SFTPC*	rs721917	*SFTPD*	d2	7.89E–05	0.000277	0.5 (0.3–0.8)
				d1d2	3.14E–05	0.000145	0.7 (0.5–0.9)

Note. Interaction type: a and d denote additive and dominant effects of the particular SNP. Numbers 1 and 2 denote the effect of the particular SNP at that position. An intragenic interaction is shown in bold. Interaction type that is associated with increased risk of PRM is marked with "". In some interactions, only one SNP exhibited a main effect, whereas the other SNP remained silent, but their interaction was significant. For example, the rs1136450 × rs1124 interaction is an a2 type, indicating that in this interaction, the main additive effect of rs1124 is significant.*

observations indicate that these associations are true rather than spurious and may validate the newer two- and three-SNP–SNP interaction models.

Association of SP SNPs in the single-SNP model: In the single-SNP model, decreased risk of PRM at both timepoints was associated with rs721917 of the *SFTPD* (**Table 3**). The rs721917 results in an alteration of the codon corresponding to amino acid 11 in the mature protein, where a methionine is replaced by a threonine. The Thr11 (C allele) variant has been associated with low serum levels of SP-D (Heidinger et al., 2005) and is shown to inhibit SP-D oligomerization (Heidinger et al., 2005). Previously, this SNP (C allele) is shown to associate with an increased risk of severe RSV (Lahti et al., 2002) and TB (Floros et al., 2000). In this study, we found the rs721917_G allele to associate with a decreased risk of PRM, which is consistent with the previous findings, where the C allele was associated with an increased risk. However, the difference in allele significance among studies may partly be due to differences in study populations and disease processes.

TABLE 6 | Association of SP gene SNP interactions with persistent respiratory morbidity (PRM) at 6 and 12 months in a three-SNP model after adjusting for covariates (age, sex, race, and weight).

SNP#1	SNP#2	SNP#3	PRM at 6 months				PRM at 12 months			
			Interaction type	*p*-Value	FDR	OR (95%CI)	Interaction type	*p*-Value	FDR	OR (95%CI)
rs1059046	rs1130866	rs721917	a1a2d3	0.00E + 00	0.0001	0.07 (0.02–0.19)	a1d3	6.21E–03	0.01	5.5 (1.5–20.5)
			a2d3	2.30E–04	0.002	0.21 (0.07–0.62)	a2d3	2.33E–03	0.005	4.5 (1.4–14.6)
			d2a3	1.08E–03	0.005	0.25 (0.09–0.72)	d2a3	7.70E–03	0.01	3.8 (1.1–13.7)
			d2	5.20E–04	0.003	0.14 (0.04–0.44)				
rs1136450	rs1130866	rs721917	d2	9.00E–05	0.001	0.10 (0.03–0.35)				
			d2a3	2.20E–04	0.002	0.24 (0.08–0.72)				

Note. Interaction type: a and d denote additive and dominant effects of the particular SNP. Numbers 1, 2, and 3 denote the effect of the particular SNP at that position. For example, the rs1136450 × rs1130866 × rs721917 interaction exhibits two effect types, d2 and d2a3. This indicates that the main dominant effect of rs1130866 in d2 type and the dominant and additive effects of rs1130866 and rs721917, respectively, in the d2a3 type, are each significant. In the d2a3 interaction, the rs1136450 remained silent.

Increased risk of PRM at both timepoints was associated with rs1124 of the *SFTPC* (**Table 3**), whereas increased risk of PRM at 6 months was associated with the rs4715 (A allele) of the *SFTPC* only. Previously, we showed in the same dataset (Gandhi et al., 2020a) that the rs4715 (A allele) was associated with an increased risk of ARF (compared with nonARF newborns) but not with the short-term outcome, pulmonary dysfunction at discharge. Other studies have shown that haplotypes of these SNPs, but not of individual SNPs, are associated with severity of RSV infection but are protective against the long-term outcome, asthma (Puthothu et al., 2006). Although in the current study, haplotype analysis was not performed; this is a goal in future studies. Conversely, although 55% of the children in our study had RSV bronchiolitis, an increased risk of PRM was associated with each of the *SFTPC* SNPs. These contrasting findings could be due to difference in patient population, environmental conditions, case-control definitions, and/or statistical approaches used for the studies. To date, no studies have been done to examine the functional impact of these polymorphisms. Therefore, we can only speculate at this time. A preclinical study in mice has shown that SP-C encoded by the *SFTPC* gene is important for stabilization and recruitment of phospholipids in surfactant (Glasser et al., 2001). It is plausible that these polymorphisms may decrease surfactant stability and, in turn, increase susceptibility to PRM.

Increased risk of PRM at 6 months was associated with the rs1130866 (C allele) of *SFTPB*. The same SNP is shown to associate with an increased risk of various other pulmonary diseases, such as chronic obstructive pulmonary disease (Seifart et al., 2002), acute respiratory distress syndrome (Lin et al., 2000b), interstitial pulmonary fibrosis (Selman et al., 2003), and ARF in adults (Quasney et al., 2004), but with a decreased risk of HP (Gandhi et al., 2021) and neonatal RDS (Floros et al., 2001). This SNP is shown to increase apoptosis, lung injury, and mortality in humanized transgenic mice (Xu et al., 2016). Moreover, this SNP (rs1130866) is part of an N-linked glycosylation site [Asn(129)-Gln-Thr131] enabling posttranscriptional N-linked glycosylation of proSP-B (Wang et al., 2003). An *in vitro* study showed an allele-specific (Ile131Thr) delay in the secretion of SP-B as well as a lower rate of secretion under experimental conditions (Taponen et al., 2013). Furthermore, a transgenic mouse model of pneumonia and sepsis carrying the C allele of this SNP showed a decreased number of lamellar bodies, SP-B concentration, and increased surface tension compared with wild-type mice after infection (Yang et al., 2019). These biologic mechanisms may shed light on the association of the rs1130866 with increased risk of PRM in our patient population where the most common etiology of ARF was pneumonia. In summary, given the importance of SP-B and SP-C in normal lung function, we postulate that SNPs of the hydrophobic proteins play a central role in ARF and its disease progression even after 1 year of the initial insult in previously healthy children. These SNPs, although not part of the mature protein, may modulate various aspects of the encoded precursor proteins, function, or other, as discussed above for SP-B, although the mechanistic details are currently unknown.

Increased risk of PRM only at 12 months was associated with rs17886395 (G allele) of the *SFTPA2* gene in the single-SNP model. In contrast, the same SNP (G allele) was associated with a decreased risk of community-acquired pneumonia in Spanish adults (García-Laorden et al., 2011). Of note, the same SNP (G allele) was associated with an increased risk of TB and allergic bronchopulmonary aspergillosis in Indian study groups (Madan et al., 2002; Saxena et al., 2003). This SNP changes the amino acid from proline (C allele) to alanine (G allele). Proline is an important component of the repetitive subunit Gly-X-Pro in the collagen region of SP-A and is known to provide stability to triple helical collagenous structures (Improta et al., 2001). We speculate that this SNP (G allele) leads to unstable and/or partially functional SP-A, and this, in turn, may increase susceptibility to respiratory infections.

Interestingly, the majority of the significant SNPs in the single-SNP model are associated with increased risk of PRM; however,

when found in interactions with other SNPs, they are associated with a decreased risk of PRM at both timepoints. Recent studies have shown that a genetic variant in the presence of another variant can alter the susceptibility of an individual to certain diseases (Cordell, 2009). The additive and/or epistatic interactions among surfactant protein genetic variants may alter concentrations and/or functional capabilities of certain SPs, and/or host defense at the cellular, molecular, or tissue level (Cordell, 2009). In addition, we have previously shown association of SP SNP interactions (but not with a single SNP) with ARF and its short-term outcome (Gandhi et al., 2020a). Collectively, our results support that epistasis plays an important role in the development and progression of complex diseases, such as PRM (Marchini et al., 2005), and studying SNP–SNP interactions is crucial to our understanding of the regulation of physiological function and their impact in health and disease state.

Association of SP SNPs in the two- and three-SNP model: Decreased risk of PRM at both timepoints was associated with the majority of significant interactions and involved SNPs of both hydrophobic and hydrophilic SP genes. The rs721917 (C allele) of *SFTPD* is significant by itself and is associated with a decreased risk of PRM. This SNP interacted with other SNPs of the SP genes and was present in the majority of SNP–SNP interactions associated with a decreased risk of PRM. These indicate a protective role of the rs721917 (C allele) of the *SFTPD* gene in the long-term outcomes of ARF survivors; however, the underlying mechanism is unknown. At 12 months, some of the significant interactions are associated with increased or decreased risks of PRM depending on dominant or additive effects of each SNP in that particular interaction in the two-SNP model (**Table 5**). For example, the interaction between the rs1965708 of the *SFTPA2* and the rs1136450 of the *SFTPA1* is associated with an increased risk of PRM if the interaction type is a1d2, meaning that the rs1965708 and the rs1136450 exhibit additive and dominant effects, respectively. However, the susceptibility to PRM could reverse with reversal of the effects of the involved SNPs, as shown for these two SNPs, if the interaction type, for example, is d1a2. We observed eight such interactions with the same SNPs to associate with either increased or decreased risk of PRM at 12 months based on the effect of SNPs in the given interaction (**Table 5**).

In the current study, we applied principles of quantitative genetics that help to deconstruct the effects of each SNP on disease susceptibility. The gene dosage is an important factor for normal gene function in health and disease conditions (Veitia and Potier, 2015). Too much or too little of a gene product and their interactions could possibly lead to over-, under-, and/or nonfunction of genes in a disease state (Veitia and Potier, 2015). Furthermore, various studies have shown that the serum concentration and biochemical properties of surfactant proteins are altered in pediatric ARF as assessed by genetic and environmental factors (Sørensen et al., 2006; Dahmer et al., 2020; Saleh et al., 2021). The disease phenotype may change based on a quantitative or qualitative imbalance of a given gene product in a given microenvironment. Together, these observations may explain the change in susceptibility to PRM based on the effect of a particular SNP in a particular interaction. Of note, one intragenic interaction between SNPs (rs1059046 and rs1965707) of the *SFTPA2* is associated with an increased risk of PRM at 12 months. These SNPs, by themselves or in combination, have been shown to associate with an increased risk of severe RSV infection and asthma in children (Lüfgren et al., 2002; Pettigrew et al., 2007; El Saleeby et al., 2010). In the current study, about ~55% of the patients with PRM had RSV bronchiolitis as an etiology of ARF; hence, our findings are in line with previous observations.

In the three-SNP model, significant intergenic interactions between SNPs of both hydrophobic (*SFTPB*) and hydrophilic SPs (*SFTPA1*, *SFTPA2*, and *SFTPD*) exhibited disease-specific outcomes, meaning the same interaction with similar effects of the involved SNPs decreased the risk of PRM at 6 months but increased the risk of PRM at 12 months (**Table 6**). Currently, these observations are puzzling and difficult to understand. However, future *in vitro* and/or *in vivo* experiments studying the impact of these gene–gene interactions on the level and properties of SPs in health and disease may help to understand these observations.

The majority of SNPs and their interactions associated with PRM risk at 6 months remained significant at 12 months as well, yet the specific interactions are very distinct from ARF and its short-term outcome in the same cohort (Gandhi et al., 2020a). In fact, the pattern of SNPs and their interactions was unique to each disease population. For example, SNPs of the *SFTPB* and the *SFTPC* by themselves and/or through their interactions were significantly associated with cystic fibrosis (Lin et al., 2018), whereas, SNPs of the *SFTPA1* and *SFTPA2* and their interactions were associated with an increased HP risk in a Mexican population (Gandhi et al., 2021) and RDS in prematurely born neonates (Amatya et al., 2021). The majority of the significant interactions associated with an increased ARF risk involved *SFTPA2* SNPs, whereas the majority of the significant interactions associated with an increased risk of pulmonary dysfunction at discharge involved *SFTPA1 SNP*s in the same dataset (Gandhi et al., 2020a). This is an interesting observation because SP-A2 encoded by SFTPA2 and SP-A1 encoded by *SFTPA1* for the most part exhibit higher activity in innate host defense/inflammatory processes and in surfactant-related functions, respectively (Floros et al., 2021). In the current study, SNPs of the hydrophobic SPs by themselves were associated with an increased risk of PRM, whereas their interactions with the hydrophilic SPs were associated with a decreased risk of PRM at 6 and 12 months. These findings may point at significant roles of a particular set of SNPs and their interactions in ARF and disease progression (short term at 28 days, and long term at 6 and 12 months) in previously healthy children. Based on the odds ratio, of the two- and three-SNP interactions, there is only one for each with an OR of more than 5 that is associated with an increased risk for PRM at 12 months. In the two-SNP model, this interaction is of the a1 effect type, between rs1965707 of the *SFTPA2* × rs1136450 of *SFTPA1*, and in the three-SNP model, the

interaction is of the a1d3-effect type, among rs1059046 of the *SFTPA2* × rs1130866 of the *SFTPB* × rs721917 of the *SFTPD*. None of the interactions had ORs of more than 5 in PRM at 6 months. Of interest, infection was the major etiology of ARF in the studied cohort. Considering the vital role of the hydrophilic SPs, particularly SP-A, in innate immunity and host responses of the lung to infection, these findings are not surprising. In the future, if these results are duplicated in a validation cohort, identification of such high-risk interactions could possibly influence clinical decision making for prognostication and counselling of parents of pediatric ARF survivors.

Strengths of this study include 1) the multicenter prospective longitudinal study design enrolling previously healthy children and the well-characterized demographic, illness, and environmental exposure information for the study cohort, and 2) the use of two different statistical approaches adjusting for clinically important variables. Some limitations should be noted for the current study. First, based on the inherent drawbacks of case-control design, the cause–effect explanation is limited. Second, we did not measure the level of SPs in serum or bronchoalvolar lavage fluid; therefore, the impact of these SNPs on SP level is unknown. Third, we only have a moderate sample size and somewhat heterogeneous patient population, despite restricting the study to those with previously healthy lungs. According to the simulation studies of Wang et al. (2010), although this sample size may produce a power of approximately 50%, it can adequately reduce false-positive rates. Thus, while a portion of significant loci remains to be detected using a larger sample size, all significant genetic effects detected in this study deserve a further investigation. More importantly, our study has identified high-order epistatic interactions for persistent respiratory morbidity susceptibility, a genetic phenomenon that has been thought to be important but highly unexplored. The majority of enrolled patients were non-Hispanic Caucasian children; hence, generalization of our findings is limited. In addition, population stratification based on race and ethnicity, and the principal component analysis, were not done, and this omission may have introduced false-positive associations. However, we have adjusted for several variables, including race, to account for difference in allele frequencies among different races. Nonetheless, these associations should be validated and replicated in heterogeneous groups of patients in a sufficiently larger sample size.

In conclusion, we showed, for the first time, the association of SP SNPs with long-term sequelae of ARF survivors in previously healthy children. Our results indicate that both groups of SPs, those involved in normal lung function, and those involved in innate immunity, associate with PRM at 6 and 12 months via complex interactions. The SNP–SNP interaction statistical method helps to identify novel high-order interaction-mediated genotype–phenotype associations not found with the standard univariate/multivariate analyses in the same dataset. The knowledge gained from the current study could be used to develop specific markers to predict long-term sequelae of ARF survivors in previously healthy children, and thus, in the long term, an intervention may be initiated to attenuate the long-term pulmonary sequelae of ARF.

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 Human Subject Protection Office of the Pennsylvania State University College of Medicine and participating sites. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

Sample acquisition, NJT. Genotyping of the samples, JF. Project administrator for coordinating samples and clinical data acquisition from the various centers, DS. Data curation, CKG and DS. Formal analysis and clinical data: CKG and GK. SNP–SNP interaction analysis: MY and RW. Logistic regression analysis: CF and SZ. Oversight for statistical analysis, RW. Overall supervision, RW, NJT, and JF. Manuscript writing—original draft, CKG and JF. Writing—review and editing, CKG, NJT, RW, and JF.

FUNDING

This work was supported by NIH HL34788 to JF.

ACKNOWLEDGMENTS

The authors thank the parents and study participants. We also thank Susan DiAngelo for her help with the genotyping.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.815727/full#supplementary-material

REFERENCES

Amatya, S., Ye, M., Yang, L., Gandhi, C. K., Wu, R., Nagourney, B., et al. (2021). Single Nucleotide Polymorphisms Interactions of the Surfactant Protein Genes Associated With Respiratory Distress Syndrome Susceptibility in Preterm Infants. *Front. Pediatr.* 9 (1065), 682160. doi:10.3389/fped.2021.682160

Amigoni, A., Pettenazzo, A., Stritoni, V., and Circelli, M. (2017). Surfactants in Acute Respiratory Distress Syndrome in Infants and Children: Past, Present and Future. *Clin. Drug Investig.* 37 (8), 729–736. doi:10.1007/s40261-017-0532-1

Chan, K. S., Mangione-Smith, R., Burwinkle, T. M., Rosen, M., and Varni, J. W. (2005). The PedsQL? *Med. Care.* 43 (3), 256–265. doi:10.1097/00005650-200503000-00008

Cordell, H. J. (2009). Detecting Gene-Gene Interactions that Underlie Human Diseases. *Nat. Rev. Genet.* 10 (6), 392–404. doi:10.1038/nrg2579

Dahmer, M. K., Flori, H., Sapru, A., Kohne, J., Weeks, H. M., Curley, M. A. Q., et al. (2020). Surfactant Protein D Is Associated With Severe Pediatric ARDS, Prolonged Ventilation, and Death in Children With Acute Respiratory Failure. *Chest.* 158 (3), 1027–1035. doi:10.1016/j.chest.2020.03.041

Day, N. E., and Byar, D. P. (1979). Testing Hypotheses in Case-Control Studies-Equivalence of Mantel-Haenszel Statistics and Logit Score Tests. *Biometrics.* 35, 623–630. doi:10.2307/2530253

Depicolzuane, L., Phelps, D. S., and Floros, J. (2021). Surfactant Protein-A Function: Knowledge Gained From SP-A Knockout Mice. *Front. Pediatr.* 9, 799693. doi:10.3389/fped.2021.799693

DiAngelo, S., Lin, Z., Wang, G., Phillips, S., Ramet, M., Luo, J., et al. (1999). Novel, Non-radioactive, Simple and Multiplex PCR-cRFLP Methods for Genotyping Human SP-A and SP-D Marker Alleles. *Dis. Markers.* 15 (4), 269–281. doi:10.1155/1999/961430

El Saleeby, C. M., Li, R., Somes, G. W., Dahmer, M. K., Quasney, M. W., and DeVincenzo, J. P. (2010). Surfactant Protein A2 Polymorphisms and Disease Severity in a Respiratory Syncytial Virus-Infected Population. *J. Pediatr.* 156 (3), 409–414. doi:10.1016/j.jpeds.2009.09.043

Floros, J., Fan, R., Diangelo, S., Guo, X., Wert, J., and Luo, J. (2001). Surfactant Protein (SP) B Associations and Interactions with SP-A in white and Black Subjects with Respiratory Distress Syndrome. *Pediatr. Int.* 43 (6), 567–576. doi:10.1046/j.1442-200x.2001.01474.x

Floros, J., Lin, H. M., García, A., Salazar, M. A., Guo, X., DiAngelo, S., et al. (2000). Surfactant Protein Genetic Marker Alleles Identify a Subgroup of Tuberculosis in a Mexican Population. *J. Infect. Dis.* 182 (5), 1473–1478. doi:10.1086/315866

Floros, J., Thorenoor, N., Tsotakos, N., and Phelps, D. S. (2021). Human Surfactant Protein SP-A1 and SP-A2 Variants Differentially Affect the Alveolar Microenvironment, Surfactant Structure, Regulation and Function of the Alveolar Macrophage, and Animal and Human Survival under Various Conditions. *Front. Immunol.* 12 (2889), 681639. doi:10.3389/fimmu.2021.681639

Floros, J., and Tsotakos, N. (2021). Differential Regulation of Human Surfactant Protein A Genes, SFTPA1 and SFTPA2, and Their Corresponding Variants. *Front. Immunol.* 12, 766719. doi:10.3389/fimmu.2021.766719

Gandhi, C. K., Chen, C., Amatya, S., Yang, L., Fu, C., Zhou, S., et al. (2021). SNP and Haplotype Interaction Models Reveal Association of Surfactant Protein Gene Polymorphisms With Hypersensitivity Pneumonitis of Mexican Population. *Front. Med.* 7, 588404. doi:10.3389/fmed.2020.588404

Gandhi, C. K., Chen, C., Wu, R., Yang, L., Thorenoor, N., Thomas, N. J., et al. (2020a). Association of SNP-SNP Interactions of Surfactant Protein Genes with Pediatric Acute Respiratory Failure. *J Clin Med.* 9 (4), 1183. doi:10.3390/jcm9041183

Gandhi, C. K., Mikerov, A. N., Durrani, F., Umstead, T. M., Hu, S., Wang, G., et al. (2020b). Impact of Ozone, Sex, and Gonadal Hormones on Bronchoalveolar Lavage Characteristics and Survival in SP-A KO Mice Infected with *Klebsiella pneumoniae*. *Microorganisms.* 8 (9), 1354. doi:10.3390/microorganisms8091354

García-Laorden, M. I., Rodríguez de Castro, F., Solé-Violán, J., Rajas, O., Blanquer, J., Borderías, L., et al. (2011). Influence of Genetic Variability at the Surfactant Proteins A and D in Community-Acquired Pneumonia: a Prospective, Observational, Genetic Study. *Crit. Care.* 15 (1), R57. doi:10.1186/cc10030

Glasser, S. W., Burhans, M. S., Korfhagen, T. R., Na, C.-L., Sly, P. D., Ross, G. F., et al. (2001). Altered Stability of Pulmonary Surfactant in SP-C-Deficient Mice. *Proc. Natl. Acad. Sci.* 98 (11), 6366–6371. doi:10.1073/pnas.101500298

Greenley, R. N., Josie, K. L., and Drotar, D. (2008). Self-reported Quality of Life Among Inner-City Youth with Asthma: an Empirical Examination of the PedsQL 3.0 Asthma Module. *Ann. Allergy Asthma Immunol.* 100 (2), 106–111. doi:10.1016/s1081-1206(10)60418-8

Heidinger, K., König, I. R., Bohnert, A., Kleinsteiber, A., Hilgendorff, A., Gortner, L., et al. (2005). Polymorphisms in the Human Surfactant Protein-D (SFTPD) Gene: strong Evidence that Serum Levels of Surfactant Protein-D (SP-D) Are Genetically Influenced. *Immunogenetics.* 57 (1), 1–7. doi:10.1007/s00251-005-0775-5

Hochberg, Y. B. a. Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodological).* 57 (1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Hope, A. C. A. (1968). A Simplified Monte Carlo Significance Test Procedure. *J. R. Stat. Soc. Ser. B (Methodological).* 30, 582–598. doi:10.1111/j.2517-6161.1968.tb00759.x

Ibiebele, I., Algert, C. S., Bowen, J. R., and Roberts, C. L. (2018). Pediatric Admissions that Include Intensive Care: a Population-Based Study. *BMC Health Serv. Res.* 18 (1), 264. doi:10.1186/s12913-018-3041-x

Improta, R., Benzi, C., and Barone, V. (2001). Understanding the Role of Stereoelectronic Effects in Determining Collagen Stability. 1. A Quantum Mechanical Study of Proline, Hydroxyproline, and Fluoroproline Dipeptide Analogues in Aqueous Solution. *J. Am. Chem. Soc.* 123 (50), 12568–12577. doi:10.1021/ja010599i

Kala, P., Have, T. T., Nielsen, H., Dunn, M., and Floros, J. (1998). Association of Pulmonary Surfactant Protein A (SP-A) Gene and Respiratory Distress Syndrome: Interaction with SP-B. *Pediatr. Res.* 43 (2), 169–177. doi:10.1203/00006450-199802000-00003

Keim, G., Watson, R. S., Thomas, N. J., and Yehya, N. (2018). New Morbidity and Discharge Disposition of Pediatric Acute Respiratory Distress Syndrome Survivors. *Crit. Care Med.* 46 (11), 1731–1738. doi:10.1097/ccm.0000000000003341

Keim, G., Yehya, N., Spear, D., Hall, M. W., Loftis, L. L., Alten, J. A., et al. (2020). Development of Persistent Respiratory Morbidity in Previously Healthy Children After Acute Respiratory Failure. *Crit. Care Med.* 48 (8), 1120–1128. doi:10.1097/ccm.0000000000004380

Khemani, R. G., Smith, L., Lopez-Fernandez, Y. M., Kwok, J., Morzov, R., Klein, M. J., et al. (2019). Paediatric Acute Respiratory Distress Syndrome Incidence and Epidemiology (PARDIE): an International, Observational Study. *Lancet Respir. Med.* 7 (2), 115–128. doi:10.1016/s2213-2600(18)30344-8

Kishore, U., Greenhough, T. J., Waters, P., Shrive, A. K., Ghai, R., Kamran, M. F., et al. (2006). Surfactant Proteins SP-A and SP-D: Structure, Function and Receptors. *Mol. Immunol.* 43 (9), 1293–1315. doi:10.1016/j.molimm.2005.08.004

Lahti, M., Löfgren, J., Marttila, R., Renko, M., Klaavuniemi, T., Haataja, R., et al. (2002). Surfactant Protein D Gene Polymorphism Associated with Severe Respiratory Syncytial Virus Infection. *Pediatr. Res.* 51 (6), 696–699. doi:10.1203/00006450-200206000-00006

Lin, Z., deMello, D., Batanian, J., Khammash, H., DiAngelo, S., Luo, J., et al. (2000a). Aberrant SP-B mRNA in Lung Tissue of Patients with Congenital Alveolar Proteinosis (CAP). *Clin. Genet.* 57 (5), 359–369. doi:10.1034/j.1399-0004.2000.570506.x

Lin, Z., Pearson, C., Chinchilli, V., Pietschmann, S., Luo, J., Pison, U., et al. (2000b). Polymorphisms of humanSP-A,SP-B, andSP-Dgenes: Association ofSP-BThr131Ile with ARDS. *Clin. Genet.* 58 (3), 181–191. doi:10.1034/j.1399-0004.2000.580305.x

Lin, Z., Thorenoor, N., Wu, R., DiAngelo, S. L., Ye, M., Thomas, N. J., et al. (2018). Genetic Association of Pulmonary Surfactant Protein Genes, SFTPA1, SFTPA2, SFTPB, SFTPC, and SFTPD With Cystic Fibrosis. *Front. Immunol.* 9, 2256. doi:10.3389/fimmu.2018.02256

Lüfgren, J., Rämet, M., Renko, M., Marttila, R., and Hallman, M. (2002). Association between Surfactant Protein A Gene Locus and Severe Respiratory Syncytial Virus Infection in Infants. *J. Infect. Dis.* 185 (3), 283–289. doi:10.1086/338473

Madan, T., Saxena, S., Murthy, K. J., Muralidhar, K., and Sarma, P. U. (2002). Association of Polymorphisms in the Collagen Region of Human SP-A1 and

SP-A2 Genes with Pulmonary Tuberculosis in Indian Population. *Clin. Chem. Lab Med.* 40 (10), 1002–1008. doi:10.1515/CCLM.2002.174

Marchini, J., Donnelly, P., and Cardon, L. R. (2005). Genome-wide Strategies for Detecting Multiple Loci that Influence Complex Diseases. *Nat. Genet.* 37 (4), 413–417. doi:10.1038/ng1537

Matthay, M. A., McAuley, D. F., and Ware, L. B. (2017). Clinical Trials in Acute Respiratory Distress Syndrome: Challenges and Opportunities. *Lancet Respir. Med.* 5 (6), 524–534. doi:10.1016/s2213-2600(17)30188-1

Nogee, L. M., Wert, S. E., Proffit, S. A., Hull, W. M., and Whitsett, J. A. (2000). Allelic Heterogeneity in Hereditary Surfactant Protein B (SP-B) Deficiency. *Am. J. Respir. Crit. Care Med.* 161 (3 Pt 1), 973–981. doi:10.1164/ajrccm.161.3.9903153

Pettigrew, M. M., Gent, J. F., Zhu, Y., Triche, E. W., Belanger, K. D., Holford, T. R., et al. (2007). Respiratory Symptoms Among Infants at Risk for Asthma: Association with Surfactant Protein A Haplotypes. *BMC Med. Genet.* 8, 15. doi:10.1186/1471-2350-8-15

Pollack, M. M., Holubkov, R., Glass, P., Dean, J. M., Meert, K. L., Zimmerman, J., et al. (2009). Functional Status Scale: New Pediatric Outcome Measure. *Pediatrics.* 124 (1), e18–e28. doi:10.1542/peds.2008-1987

Puthothu, B., Krueger, M., Heinze, J., Forster, J., and Heinzmann, A. (2006). Haplotypes of Surfactant Protein C Are Associated with Common Paediatric Lung Diseases. *Pediatr. Allergy Immunol.* 17 (8), 572–577. doi:10.1111/j.1399-3038.2006.00467.x

Quasney, M. W., Waterer, G. W., Dahmer, M. K., Kron, G. K., Zhang, Q., Kessler, L. A., et al. (2004). Association between Surfactant Protein B + 1580 Polymorphism and the Risk of Respiratory Failure in Adults with Community-Acquired Pneumonia. *Crit. Care Med.* 32 (5), 1115–1119. doi:10.1097/01.ccm.0000124872.55243.5a

Rämet, M., Haataja, R., Marttila, R., Floros, J., and Hallman, M. (2000). Association between the Surfactant Protein A (SP-A) Gene Locus and Respiratory-Distress Syndrome in the Finnish Population. *Am. J. Hum. Genet.* 66 (5), 1569–1579. doi:10.1086/302906

Saleh, N. Y., Ibrahem, R. A. L., Saleh, A. A. h., Soliman, S. E. s., and Mahmoud, A. A. S. (2021). Surfactant Protein D: a Predictor for Severity of Community-Acquired Pneumonia in Children. *Pediatr. Res.* 91, 665–671. doi:10.1038/s41390-021-01492-9

Saxena, S., Madan, T., Shah, A., Muralidhar, K., and Sarma, P. U. (2003). Association of Polymorphisms in the Collagen Region of SP-A2 with Increased Levels of Total IgE Antibodies and Eosinophilia in Patients with Allergic Bronchopulmonary Aspergillosis. *J. Allergy Clin. Immunol.* 111 (5), 1001–1007. doi:10.1067/mai.2003.1395

Seid, M., Limbers, C. A., Driscoll, K. A., Opipari-Arrigan, L. A., Gelhard, L. R., and Varni, J. W. (2010). Reliability, Validity, and Responsiveness of the Pediatric Quality of Life Inventory (PedsQL) Generic Core Scales and Asthma Symptoms Scale in Vulnerable Children with Asthma. *J. Asthma.* 47 (2), 170–177. doi:10.3109/02770900903533966

Seifart, C., Plagens, A., Brödje, D., Müller, B., von Wichert, P., and Floros, J. (2002). Surfactant Protein B Intron 4 Variation in German Patients with COPD and Acute Respiratory Failure. *Dis. Markers.* 18 (3), 129–136. doi:10.1155/2002/194075

Selman, M., Lin, H.-M., Montao, M., Jenkins, A. L., Estrada, A., Lin, Z., et al. (2003). Surfactant proteinA and B Genetic Variants Predispose to Idiopathic Pulmonary Fibrosis. *Hum. Genet.* 113 (6), 542–550. doi:10.1007/s00439-003-1015-4

Serrano, A. G., and Pérez-Gil, J. (2006). Protein-lipid Interactions and Surface Activity in the Pulmonary Surfactant System. *Chem. Phys. Lipids.* 141 (1-2), 105–118. doi:10.1016/j.chemphyslip.2006.02.017

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI Database of Genetic Variation. *Nucleic Acids Res.* 29 (1), 308–311. doi:10.1093/nar/29.1.308

Silveyra, P., and Floros, J. (2012). Genetic Variant Associations of Human SP-A and SP-D with Acute and Chronic Lung Injury. *Front. Biosci.* 17, 407–429. doi:10.2741/3935

Sørensen, G. L., Hjelmborg, J. v. B., Kyvik, K. O., Fenger, M., Høj, A., Bendixen, C., et al. (2006). Genetic and Environmental Influences of Surfactant Protein D Serum Levels. *Am. J. Physiology-Lung Cell Mol. Physiol.* 290 (5), L1010–L1017. doi:10.1152/ajplung.00487.2005

Taponen, S., Huusko, J. M., Petäjä-Repo, U. E., Paananen, R., Guttentag, S. H., Hallman, M., et al. (2013). Allele-specific N-Glycosylation Delays Human Surfactant Protein B Secretion *In Vitro* and Associates with Decreased Protein Levels *In Vivo*. *Pediatr. Res.* 74 (6), 646–651. doi:10.1038/pr.2013.151

Thomas, N. J., DiAngelo, S., Hess, J. C., Fan, R., Ball, M. W., Geskey, J. M., et al. (2009). Transmission of Surfactant Protein Variants and Haplotypes in Children Hospitalized with Respiratory Syncytial Virus. *Pediatr. Res.* 66 (1), 70–73. doi:10.1203/PDR.0b013e3181a1d768

Thorenoor, N., Kawasawa, Y. I., Gandhi, C. K., and Floros, J. (2020). Sex-Specific Regulation of Gene Expression Networks by Surfactant Protein A (SP-A) Variants in Alveolar Macrophages in Response to *Klebsiella pneumoniae*. *Front. Immunol.* 11, 1290. doi:10.3389/fimmu.2020.01290

Thorenoor, N., Kawasawa, Y. I., Gandhi, C. K., Zhang, X., and Floros, J. (2019). Differential Impact of Co-expressed SP-A1/sp-A2 Protein on AM miRNome; Sex Differences. *Front. Immunol.* 10, 1960. doi:10.3389/fimmu.2019.01960

Veitia, R. A., and Potier, M. C. (2015). Gene Dosage Imbalances: Action, Reaction, and Models. *Trends Biochem. Sci.* 40 (6), 309–317. doi:10.1016/j.tibs.2015.03.011

Wang, G., Christensen, N. D., Wigdahl, B., Guttentag, S. H., and Floros, J. (2003). Differences in N-Linked Glycosylation between Human Surfactant Protein-B Variants of the C or T Allele at the Single-Nucleotide Polymorphism at Position 1580: Implications for Disease. *Biochem. J.* 369 (Pt 1), 179–184. doi:10.1042/bj20021376

Wang, Z., Liu, T., Lin, Z., Hegarty, J., Koltun, W. A., and Wu, R. (2010). A General Model for Multilocus Epistatic Interactions in Case-Control Studies. *PLoS One.* 5 (8), e11384. doi:10.1371/journal.pone.0011384

Wert, S. E., Whitsett, J. A., and Nogee, L. M. (2009). Genetic Disorders of Surfactant Dysfunction. *Pediatr. Dev. Pathol.* 12 (4), 253–274. doi:10.2350/09-01-0586.1

Wright, J. R. (2005). Immunoregulatory Functions of Surfactant Proteins. *Nat. Rev. Immunol.* 5 (1), 58–68. doi:10.1038/nri1528

Xu, H. N., Lin, Z., Gandhi, C. K., Amatya, S., Wang, Y., Li, L. Z., et al. (2020). Sex and SP-A2 Dependent NAD(H) Redox Alterations in Mouse Alveolar Macrophages in Response to Ozone Exposure: Potential Implications for COVID-19. *Antioxidants.* 9 (10), 915. doi:10.3390/antiox9100915

Xu, Y., Ge, L., Abdel-Razek, O., Jain, S., Liu, Z., Hong, Y., et al. (2016). Differential Susceptibility of Human Sp-B Genetic Variants on Lung Injury Caused by Bacterial Pneumonia and the Effect of a Chemically Modified Curcumin. *Shock (Augusta, Ga.).* 45 (4), 375–384. doi:10.1097/shk.0000000000000535

Yang, F., Zhang, J., Yang, Y., Ruan, F., Chen, X., Guo, J., et al. (2019). Regulatory Roles of Human Surfactant Protein B Variants on Genetic Susceptibility to Pseudomonas Aeruginosa Pneumonia-Induced Sepsis. *Shock.* 54, 507–519. doi:10.1097/shk.0000000000001494

Yehya, N., and Thomas, N. J. (2016). Relevant Outcomes in Pediatric Acute Respiratory Distress Syndrome Studies. *Front. Pediatr.* 4, 51. doi:10.3389/fped.2016.00051

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

ORIGINAL RESEARCH
published: 19 April 2022
doi: 10.3389/fgene.2022.818241

Genetic and Clinical Features of Heterotaxy in a Prenatal Cohort

*Tong Yi[1,2,3], Hairui Sun[1,4], Yuwei Fu[5], Xiaoyan Hao[1], Lin Sun[1], Ye Zhang[1], Jiancheng Han[1], Xiaoyan Gu[1], Xiaowei Liu[1], Yong Guo[1], Xin Wang[1], Xiaoxue Zhou[1], Siyao Zhang[1], Qi Yang[1], Jiaqi Fan[1] and Yihua He[1]**

[1]Beijing Anzhen Hospital, Capital Medical University, Beijing, China, [2]Beijing Lab for Cardiovascular PrecisionMedicine, Beijing, China, [3]Beijing Advanced Innovation Center for Big Data-based Precision Medicine, Capital Medical University, Beijing, China, [4]Key Laboratory of Medical Engineering for Cardiovascular Disease, Ministry of Education, Beijing, China, [5]Department of Ultrasound, Peking University International Hospital, Beijing, China

Objectives: Some genetic causes of heterotaxy have been identified in a small number of heterotaxy familial cases or animal models. However, knowledge on the genetic causes of heterotaxy in the fetal population remains scarce. Here, we aimed to investigate the clinical characteristics and genetic spectrum of a fetal cohort with heterotaxy.

Methods: We retrospectively investigated all fetuses with a prenatal diagnosis of heterotaxy at a single center between October 2015 and November 2020. These cases were studied using the genetic testing data acquired from a combination of copy number variation sequencing (CNV-seq) and whole-exome sequencing (WES), and their clinical phenotypes were also reviewed.

Result: A total of 72 fetuses diagnosed with heterotaxy and complete clinical and genetic results were enrolled in our research. Of the 72 fetuses, 18 (25%) and 54 (75%) had left and right isomerism, respectively. Consistent with the results of a previous study, intracardiac anomalies were more severe in patients with right atrial isomerism than in those with left atrial isomerism (LAI) and mainly manifested as atrial situs inversus, bilateral right atrial appendages, abnormal pulmonary venous connection, single ventricles or single atria, and pulmonary stenosis or atresia. In 18 fetuses diagnosed with LAI, the main intracardiac anomalies were bilateral left atrial appendages. Of the 72 fetuses that underwent CNV-seq and WES, 11 (15.3%) had positive genetic results, eight had definitive pathogenic variants, and three had likely pathogenic variants. The diagnostic genetic variant rate identified using WES was 11.1% (8/72), in which primary ciliary dyskinesia (PCD)-associated gene mutations (CCDC40, CCDC114, DNAH5, DNAH11, and ARMC4) accounted for the vast majority (n = 5). Other diagnostic genetic variants, such as KMT2D and FOXC1, have been rarely reported in heterotaxy cases, although they have been verified to play roles in congenital heart disease.

Conclusion: Thus, diagnostic genetic variants contributed to a substantial fraction in the etiology of fetal heterotaxy. PCD mutations accounted for approximately 6.9% of heterotaxy cases in our fetal cohort. WES was identified as an effective tool to detect genetic causes prenatally in heterotaxy patients.

Keywords: congenital heart, heterotaxy syndrome, whole exome, CNV (copy number variant), echocardiography

OPEN ACCESS

Edited by:
Wenhao Zhou,
Fudan University, China

Reviewed by:
Zirui Dong,
The Chinese University of Hong Kong, Hong Kong SAR, China
Maria Sabater-Molina,
Hospital Universitario Virgen de la Arrixaca, Spain

****Correspondence:***
Yihua He
heyihuaecho@hotmail.com

Specialty section:
This article was submitted to Genetics of Common and Rare Diseases, a section of the journal Frontiers in Genetics

Received: *19 November 2021*
Accepted: *09 March 2022*
Published: *19 April 2022*

Citation:
Yi T, Sun H, Fu Y, Hao X, Sun L, Zhang Y, Han J, Gu X, Liu X, Guo Y, Wang X, Zhou X, Zhang S, Yang Q, Fan J and He Y (2022) Genetic and Clinical Features of Heterotaxy in a Prenatal Cohort. Front. Genet. 13:818241. doi: 10.3389/fgene.2022.818241

INTRODUCTION

Heterotaxy syndrome is a congenital disorder resulting from incorrect establishment of left–right (LR) patterning during embryogenesis. It is a relatively uncommon syndrome, with an incidence rate of approximately 1:7,000–1:5,000 (Lin et al., 2000; Reller et al., 2008). Classic heterotaxy is characterized by abnormally arranged thoracic and visceral organs and a relatively high mortality mainly due to cardiac defects. Owing to the complexity of this disease, there is no consensus regarding the precise nomenclature of heterotaxy and nosological relationships between clinical and phenotypic features of patients with heterotaxy; however, it is generally classified as left or right isomerism. Clinical outcomes are remarkably poorer in patients with right isomerism than in those with left isomerism (Baban et al., 2018). Moreover, despite surgical management, the occurrence of congenital heart defects associated with heterotaxy has a relatively poor survival associated compared to that of patients with similar congenital heart defects but without heterotaxy (Bartz et al., 2006; Kim et al., 2008). Hence, heterotaxy is not only a disease with significant phenotypic heterogeneity but also a disorder with a considerable medical and economic burden associated with it.

The molecular mechanisms of LR patterning have been extensively investigated in animal models, and numerous implicated genes have been identified. However, only a few of them are candidate genes for causing heterotaxy in humans (Ma et al., 2012; Bellchambers and Ware, 2018). In 1997, the zinc finger protein of the cerebellum 3 gene (*ZIC3*) was first identified as the cause of heterotaxy based on familial X-linked pedigrees. However, mutations in *ZIC3* explain only 3–5% of sporadic heterotaxy cases (Bellchambers and Ware, 2018). Similarly, although the Nodal signaling pathway is a conserved and well-established pathway involved in embryonic LR development, point mutations in members of the Nodal pathway are found only in 5–10% of patients with heterotaxy (Dykes, 2014; Grimes and Burdine, 2017; Pelliccia et al., 2017). Currently, known genes account for only approximately 10–20% of sporadic cases (Sutherland and Ware, 2009; Fakhro et al., 2011). As its specific genetic etiology is currently identifiable in only a minority of patients, numerous novel genes and pathways implicated in the pathogeny of this disorder are expected to be discovered in the next years.

During the past decade, copy number variants (CNVs) and single-nucleotide polymorphisms (SNPs) that are associated with heterotaxy have been identified using G-banding and chromosome microarray analysis (Cowan et al., 2016). Moreover, whole-exome sequencing (WES) has been slowly implemented in the clinical setting in recent years as it can detect intragenic variants and lead to a more comprehensive genetic diagnosis (Li et al., 2019). WES enables the rapid analysis of a large number of gene sequences and has permitted the discovery of genes involved in numerous genetic diseases. To date, only a few studies have used WES to uncover the genetic etiology of congenital abnormalities, and most of them have focused on the study of postnatal cases (Lord et al., 2019; Diderich et al., 2021). To the best of our knowledge, no studies have used WES for the prenatal genetic diagnosis of heterotaxy.

In this study, the combination of CNV sequencing (CNV-seq) and WES was used to detect causative genes and mutations associated with heterotaxy in a prenatal cohort. We aimed to determine the characteristics of the genetic variants identified using WES and to evaluate the diagnostic value of WES in patients with heterotaxy. To this end, fetus, mother, and father trios were sequenced simultaneously. The detected variants were then interpreted following the guidelines recommended by the American College of Medical Genetics (ACMG).

MATERIALS AND METHODS

Patient Recruitment

We retrospectively investigated the previously diagnosed cases of fetuses with heterotaxy at the Beijing Anzhen Hospital, Capital Medical University, from October 2015 to November 2020. Clinical data, including gestational age at diagnosis, family history, indications for fetal echocardiography, and pregnancy outcomes, were collected. The available postnatal echocardiograms and autopsy studies were reviewed to confirm the diagnosis of heterotaxy. This study was approved by the research ethics board of Anzhen Hospital, and informed consent was obtained from the parents.

Echocardiographic Data and Phenotypic Classification

Fetal echocardiography was performed using a segmental approach with standardized anatomical planes and pulse-wave and color Doppler imaging. All cases were diagnosed by two independent investigators with extensive experience in the field of fetal cardiology (Carvalho et al., 2013). Heterotaxy was diagnosed when fetuses exhibited visceral situs ambiguity and complex congenital heart disease (CHD) (Primo et al., 2020).

Right atrial isomerism (RAI) was diagnosed in the presence of characteristic cardiac anomalies, including atrioventricular septal defect, conotruncal lesions, anomalous pulmonary venous return, and atrial appendages on both sides of the body that have the appearance of a morphologically right atrial appendage, and with at least one of the following findings: 1) juxtaposition of the abdominal aorta and inferior vena cava (IVC) or 2) visceral heterotaxy, namely, discordant laterality of the stomach, portal sinus, or gallbladder. Left atrial isomerism (LAI) was suspected in the presence of azygos continuation of the interrupted IVC and at least one of the common features of LAI: 1) cardiac defects, 2) heart block, and 3) visceral heterotaxy. The spleen was also assessed to classify the type of atrial isomerism, namely, asplenia in RAI and polysplenia in LAI.

CNV-Seq

Fetal samples, including dermal biopsies, umbilical cord blood, and parent blood, were collected. All fetus–mother–father trios underwent CNV-seq and WES. Both CNV-seq and WES were performed using a research-based protocol. CNV-seq was

performed as described previously (Kearney et al., 2011; Liu et al., 2015). Briefly, RNA-free high-molecular-weight genomic DNA (gDNA) was isolated from the umbilical cord using a QIAGEN DNA Blood Midi/Mini Kit (QIAGEN GmbH, Hilden, Germany) following the manufacturer's protocol. The quality and concentration of gDNA from the samples were assessed using a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, United States). Approximately 5 million sequencing reads from each sample were mapped to the NCBI human reference genome (hg19/GRCh37) using the Burrows–Wheeler Aligner and allocated to 20 kb sequencing bins with a 5 kb sliding window to achieve a high resolution in identifying CNVs. The expected resolution was 100 kb. The CNV profiles of each chromosome were represented as log2 of the mean sequencing reads of each sequencing bin along the chromosome.

Detected CNVs were evaluated based on a scientific literature review and the following public databases: DECIPHER (https://decipher.sanger.ac.uk/), DGV (http://dgv.tcag.ca/), 1000 Genomes Project (http://www.internationalgenome.org/), OMIM (http://omim.org/), ClinVar (http://www.ncbi.nlm.nih.gov/clinvar), ClinGen (https://www.clinicalgenome.org/), and ISCA CNV (https://www.iscaconsortium.org). Following the ACMG and Genomics standards and guidelines for the interpretation of CNVs, CNVs were classified into three categories: benign, uncertain clinical significance, and pathogenic (Riggs et al., 2021). In this study, we only report pathogenic CNVs.

WES and Data Analysis

gDNA was extracted from the umbilical cord and parental blood using a QIAGEN DNA Blood Midi/Mini Kit. DNA libraries were prepared using an Agilent liquid capture system (Agilent SureSelect Human All Exon V6) according to the manufacturer's protocol. The size distribution and concentration of the libraries were determined using an Agilent 2100 Bioanalyzer and quantified using real-time polymerase chain reaction. The DNA library was sequenced on an Illumina Hiseq 4000 or Illumina Novaseq for paired-end 150 bp-long reads according to the manufacturer's protocol. The mean sequencing coverage on target regions of whole-exome sequencing was 103-fold. Raw image files were processed using bcl2fastq (Illumina) for base calling and generating raw data. Low-quality sequencing reads were filtered using a quality score of 20 (Q20). The average read depths were 103X for each case. The reads were aligned with the NCBI human reference genome (hg19/GRCh37) using the Burrows–Wheeler Aligner. The BAM files were subjected to SNP analysis, duplication marking, indel realignment, and recalibration using GATK and Picard.

Variant Annotation, Filtering, and Classification

After variant detection, wANNOVAR (http://wannovar.wglab.org/) was used for annotation. Variant frequencies were determined using dbSNP150 (https://www.ncbi.nlm.nih.gov/SNP/), the 1000 Genomes Project, Exome Variant Server (http://evs.gs.washington.edu/EVS), ExAC (http://exac.broadinstitute.org/), and in-house databases. Common SNPs (minor allele frequency > 0.1%) were removed. Nonsynonymous, spliced, frameshift, nonframeshift variants and variants located in splice sites within 10 bp of an exon were prioritized for the study. SIFT (http://sift.jcvi.org), PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2), MutationTaster (http://www.mutationtaster.org), and CADD (http://cadd.gs.washington.edu) were used to predict the pathogenicity of missense variants, whereas Human Splicing Finder (http://www.umd.be/HSF) and MaxEntScan (http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html) were used to evaluate the splicing effects. Moreover, databases such as OMIM, ClinVar, and Human Gene Mutation Database (http://www.hgmd.org) were used to determine variant harmfulness and pathogenicity when appropriate. ACMG variant classification recommendations were utilized for all reported variants (Li et al., 2017). Variants were classified into one of the following five categories: pathogenic, likely pathogenic, of uncertain significance, likely benign, or benign.

Sanger Sequencing Confirmation

Sanger sequencing was performed to confirm all potentially diagnostic genetic variants identified.

Statistical Analyses

Categorical variables are presented as frequencies (percentages) and were compared using the Pearson χ^2 test or Fisher's exact test. Statistical analyses were performed using SPSS version 23 (SPSS Inc., Chicago, IL, United States). Statistical significance was set at $p < 0.05$.

RESULTS

Cohort Characteristics

From October 2015 to November 2021, 135 pregnant women found to have fetuses with heterotaxy were screened for inclusion in our study. Among them, eight chose to continue pregnancy without genetic testing and 127 chose to terminate their pregnancy (of which 49 women chose not to undergo genetic testing or to do it at other medical centers or companies). Hence, 78 parents chose to terminate the pregnancy and underwent CNV-seq and WES sequentially at our center. Six cases were excluded because of incomplete imaging or the lack of evidence of heterotaxy. Thus, 72 pregnancies were eligible for the analysis in our study (**Figure 1**). Autopsy was refused by the parents for 41 fetuses.

All parents were healthy according to their physical examination reports and were nonconsanguineous. The mean maternal age at diagnosis was 28.2 ± 4.11 years, and the median gestational age was 23.74 ± 2.84 weeks. Left isomerism was present in 18 (25%) fetuses, and right isomerism was present in 54 (75%) among all fetuses analyzed (72). The spectrum of cardiovascular abnormalities is shown in **Table 1**. The main intracardiac anomalies detected among the 54 fetuses with prenatally diagnosed RAI were atrial situs inversus, bilateral

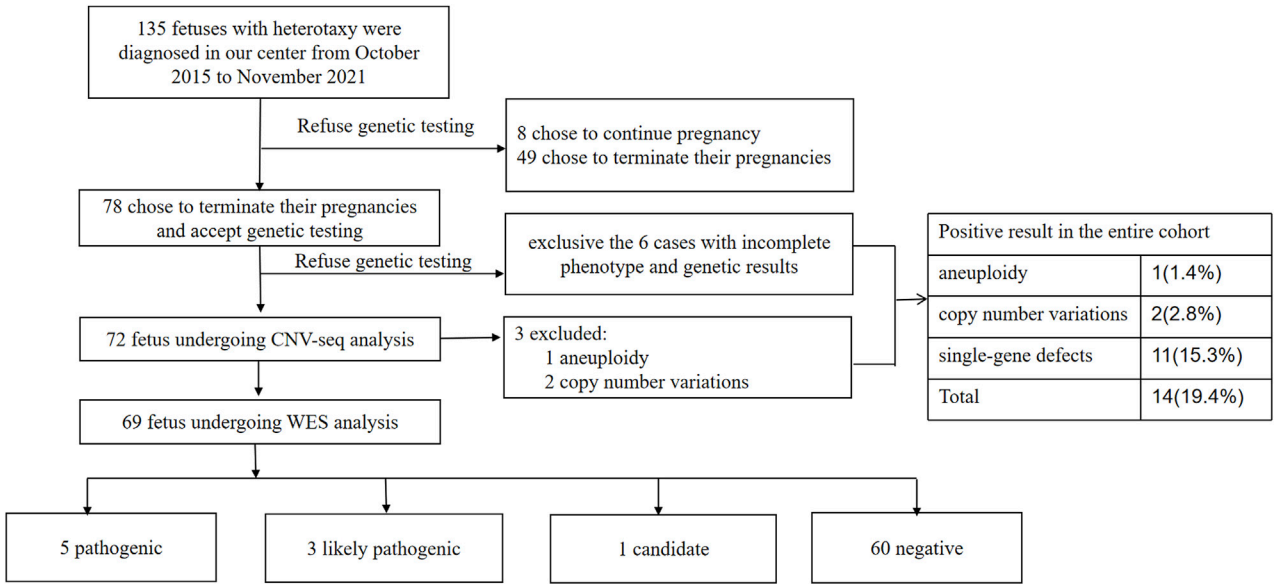


FIGURE 1 | Flowchart summarizing detection of genetic variants by CNV-Seq and WES performed in 72 prenatal samples with heterotaxy. CNV, copy number variation; WES, whole-exome sequencing.

right atrial appendages, abnormal pulmonary venous connection, single ventricles or atria, and pulmonary stenosis or atresia. Among the 18 fetuses diagnosed with LAI, the main intracardiac anomaly was the presence of bilateral left atrial appendages. **Table 2** summarizes the distribution of visceral abnormalities among patients with left and right isomerism; it also summarizes the visceral abnormalities in patients with RAI. Most RAI patients presented with juxtaposition of the descending aorta and IVC on the same side, aspenia, and bilateral right bronchi. The main extracardiac anomalies in LAI were interrupted IVC, bilateral left bronchi (long), polysplenia, and arrhythmias.

Frequency of the Genetic Variants in Different Presentations of Heterotaxy

Among the 72 cases analyzed, eight had definitive pathogenic variants and three had likely pathogenic variants. The positive detection rate was approximately 15.3% (11/72). The frequencies of the genetic variants in patients with different presentations of heterotaxy are shown in **Supplementary Table S1**. The positive genetic detection rate for isomerism of the right atrial appendages was significantly higher than that for isomerism of the left atrial appendage ($p = 0.04$).

Diagnostic Yield of CNV-Seq

Among the 72 cases for whom CNV-seq analysis was performed, three (16.9%) had a chromosomal abnormality, including one (7.2%) with aneuploidy (21 chromosome trisomy) and two (9.6%) with a *de novo* pathogenic CNV (22q11.2 deletion). The three cases with such chromosomal abnormalities were prospectively excluded from the WES analysis.

Contribution of Single-Gene Defects

Of the 69 patients for whom WES analysis was performed, 11.6% (8/69) had a positive result, indicating an increased yield of 11.1% (8/72) and an overall diagnostic yield of 15.3% (11/72) in the entire cohort (**Table 1**). The genotype–phenotype information of the eight cases is shown in **Table 3**. The ratio of recessive genotypes was 8.3% (6/72), whereas dominant mutations accounted for 4.2% (3/72) of heterotaxy patients. *De novo* mutations accounted for 2.8% (2/72) of patients.

In our cohort, the largest contribution of pathogenic diagnostic genetic variant-related phenotypes was primary ciliary dyskinesia (PCD). We found evidence of recessive mutations (compound heterozygous alleles) in *CCDC40*, *CCDC114*, *DNAH5*, *DNAH11*, and *ARMC4*. All these genes are known to cause PCD (MIM 242650). The remaining likely pathogenic and candidate variants were identified in genes that have not been previously reported to be associated with heterotaxy, including *FOXC1*, *KMT2D*, and *FGFR3*. These genes have been reported to be related to heart development, except for *FGFR3*. The patient with the mutation in *FGFR3* had extracardiac phenotypes of disorders of the cervical vertebrae (C4–7), lymph node cysts of the neck, abnormal appearance of the right ear, defects of the right radius, defects of the right thumb, single umbilical arteries, abnormal running of the portal vein, asplenia, and a double right main bronchus.

DISCUSSION

To the best of our knowledge, this is the first cohort-based study assessing the contribution of genetic variants to heterotaxy in a fetal population. Most importantly, we identified genetic

TABLE 1 | Spectrum of cardiovascular abnormalities in right and left atrial isomerism patients.

	RAI	LAI	Number of patients	P
	N% (95% CI)	N% (95% CI)		
Total number of patients in each column	54	18		
Cardiac abnormality				
Cardiac position				0.035[a]
Levocardia	23 (42.6)	14 (77.7)	37 (51.4)	
Dextrocardia	22 (40.7)	3 (16.7)	25 (34.7)	
Mesocardia	9 (16.7)	1 (5.6)	10 (13.9)	
Atrial arrangement				0.000[b]
Atrial situs inversus	8 (14.8)	1 (5.6)	9 (12.5)	
Bilateral right atrial appendages	29 (53.7)	0	29 (40.3)	
Bilateral left atrial appendage	0	14 (77.8)	14 (19.4)	
No record	17 (31.5)	3 (16.7)	20 (27.8)	
SV/SA	27 (50.0)	3 (16.7)	30 (41.7)	0.013[a]
Non-SV/SA	27 (50.0)	15 (83.3)	42 (58.3)	
AVSD	35 (64.8)	9 (50.0)	44 (61.1)	0.264[a]
Non-AVSD	19 (35.2)	9 (50.0)	28 (38.9)	
Outflow tracts and great vessels				
DORT	21 (38.9)	6 (33.3)	27 (37.5)	0.673[a]
Non-DORT	33 (61.1)	12 (66.7)	45 (62.5)	
TGA	7 (13.0)	2 (11.1)	9 (12.5)	1.000[b]
Non-TGA	47 (87.0)	16 (88.9)	63 ()	
PS or PA	43 (79.6)	8 (44.4)	51 (70.8)	0.004[a]
Non-(PS or PA)	11 (20.4)	10 (55.6)	21 (29.2)	
Truncus arteriosus				0.434[b]
Truncus arteriosus	9 (16.7)	1 (5.6)	10 (86.1)	
Non-truncus arteriosus	45 (83.3)	17 (94.4)	62 (13.9)	
Aortic arch				0.380[a]
Left aortic arch	21 (38.9)	8 (44.4)	29 (40.3)	
Right aortic arch	15 (27.8)	7 (38.9)	22 (30.6)	
Unknown	18 (33.3)	3 (16.7)	21 (29.2)	
Venous anomalies				
SVC				0.642[b]
Right SVC	19 (35.2)	8 (44.4)	27 (37.5)	
Left SVC	4 (7.4)	0 ()	4 (5.6)	
Bilateral SVC	20 (37.0)	5 (27.8)	25 (34.7)	
Unknown	11 (20.4)	5 (27.8)	16 (22.2)	
TAPVC/PAPVC				0.028[a]
TAPVC/PAPVC	28 (51.9)	4 (22.2)	32 (44.4)	
Non-TAPVC/PAPVC	26 (48.1)	14 (77.8)	40 (55.6)	

The Chi square or Fisher's exact test was performed as appropriate; a. Pearson's Chi-squared test; b. Fisher's exact test; RAI, right atrial isomerism; LAI, left atrial isomerism; AVSD, atrioventricular septal defect; SV/SA, single ventricle or single atrium; DORT, double-outlet right ventricle; TGA, transposition of great arteries; PS or PA, pulmonary stenosis or atresia; SVC, superior vena cava; TAPVC/PAPVC, anomalous pulmonary venous return.

abnormalities in 11 (15.3%) fetuses. These abnormalities represented aneuploidies, CNVs, and diagnostic genetic variants in 1.4, 2.8, and 11.1% of the cases, respectively. WES combined with CNV-seq, rather than just CNV-seq or targeted panels, which only provide partial information about the genome, led to a dramatic rise in the detection rate of genetic variants associated with heterotaxy in our prenatal cohort. This finding supports the notion that WES is an effective tool for detecting the genetic causes of heterotaxy as it can detect both chromosomal abnormalities and novel candidate genes.

Importantly, the WES results showed a spectrum of gene variants involved in the occurrence of heterotaxy, highlighting the need for implementing the use of this technique in a clinical setting to facilitate perinatal decision making and management when conventional tests such as karyotype testing or microarray analyses are inconclusive. Nearly half of the diagnostic genetic mutations identified in prenatal heterotaxy patients were associated with PCD, which is a recessive heterogeneous disorder of motile cilia that presents with chronic otosinopulmonary disease and organ lateral defects in 50% of the cases and with heterotaxy in 10% of the cases (Knowles et al., 2013). Moreover, a previous study reported that patients with heterotaxy and PCD had a higher prevalence of respiratory symptoms compared to patients with heterotaxy without PCD (Shapiro et al., 2014). In addition, another study has shown that patients with PCD and mutations in *CCDC40* have a more severe disease course with an earlier onset and a higher prevalence of neonatal respiratory distress than PCD patients without such

TABLE 2 | Spectrum of visceral abnormalities in right and left atrial isomerism.

	RAI	LAI	Total (number of patients, %)	P (b)
IVC				0.000
Interrupted IVC, azygos/hemiazygos vein continuation	1 (1.9)	16 (88.9)	17 (23.6)	
Noninterrupted IVC	53 (98.1)	2 (11.1)	55 (76.4)	
Relationship of IVC and descending aorta				0.053
IVC right of the spine and descending aorta left of the spine	8 (14.8)	0	8 (11.1)	
IVC left of the spine and the descending aorta right of the spine	1 (1.9)	0	1 (1.4)	
IVC and descending aorta same side	22 (40.7)	0	22 (30.6)	
IVC left of the spine and the descending aorta anterior of the spine	6 (11.1)	0	6 (8.3)	
IVC anterior of the spine and the descending aorta left of the spine	0	1 (5.6)	1 (1.4)	
Unknown	16 (29.6)	1 (5.6)	17 (23.6)	
Bronchi				0.000
Bilateral right bronchi (short)	18 (33.3)	0	18 (25)	
Bilateral left bronchi (long)	1 (1.9)	8 (44.4)	9 (12.5)	
Multiple pulmonary lobes	1 (1.9)	0	1 (1.4)	
Unknown	34 (63.0)	10 (55.6)	44 (61.1)	
Spleen				0.000
Polysplenia	1 (1.9)	12 (66.7)	13 (18.1)	
Aspenia	25 (46.3)	0	25 (34.7)	
Single right spleen	8 (14.8)	0	8 (11.1)	
Single left spleen	1 (1.9)	1 (5.6)	2 (2.8)	
Unknown	19 (35.2)	5 (27.8)	24 (33.3)	
Stomach				0.29
Right-sided stomach	28 (51.9)	6 (33.3)	34 (47.2)	
Left-sided stomach	18 (33.3)	9 (50)	27 (37.5)	
Stomach centrally situated	3 (5.6)	0	3 (4.2)	
Unknown	5 (9.3)	3 (16.7)	8 (11.1)	
Liver				0.511
Left-sided liver	24 (44.4)	6 (33.3)	30 (41.7)	
Liver centrally situated	18 (33.3)	9 (50.0)	27 (37.5)	
Right-sided liver	7 (13.0)	3 (16.7)	10 (13.9)	
Unknown	5 (9.3)	0	5 (6.9)	

The Chi square or Fisher's exact test as appropriate; a. Pearson's Chi-squared test; b. Fisher's exact test. Values are presented as numbers (%). RAI, right atrial isomerism; LAI, left atrial isomerism; IVC, inferior vena cava.

mutations (Davis et al., 2015). Hence, we recommend the use of WES in combination with standard diagnostic testing in patients prenatally diagnosed with heterotaxy to identify the pathogenic genetic variants and predict severity of the disease after birth.

Point mutations in Nodal signaling pathway members, which are known to be associated with laterality defects, were not routinely identified in our cohort. This may be due to the limitations of the screening conditions used in the previous studies. Numerous mutations identified in *Nodal* by previous studies were classified as benign in our research, according to the ACMG guidelines (Mohapatra et al., 2009). Mutations in *ZIC3* are also well-documented to be related to situs abnormalities. In our cohort, we did not find any pathogenic or likely pathogenic variants of *ZIC3*. This may be because *ZIC3* mutations underlie only a minority (3–5%) of sporadic heterotaxy cases (Bellchambers and Ware, 2018). Expanding the size of the cohort used may contribute to test this hypothesis. Finally, in our study, many genes related to PCD encoding components of cilia or the assembly of motile cilia were identified. This result is consistent with previous research showing that ultrastructural dynein arm defects strongly correlate with the development of situs abnormalities (Knowles et al., 2016). Collectively, these data highlight how the use of different research techniques or cohorts reveals differences in the genetic spectrum of heterotaxy. These differences should be considered in the clinical setting when performing molecular diagnostics.

Two genes that had not been previously associated with heterotaxy but have been found to be involved in early cardiac development (*KMT2D* and *FOXC1*) were identified in this study. The sequence alteration in coding exon 31 of *KMT2D* is identified in one family: a heterozygous deletion of one nucleotide, c.6595delT, predicting a frameshift with a premature stop codon at position 2,262 of the protein p.(Y2199fs*64). The variant was not detected in both parents, confirming that the variant is *de novo*. This variant has not yet been reported in public databases, including the Genome Aggregation Database (gnomAD), the Exome Aggregation Consortium (Exac), 1000 Genomes Project (1KGP), and Exome Sequencing Project

TABLE 3 | Genotype–phenotype information of patients diagnosed with pathogenic variants, likely pathogenic variants, or variants of uncertain significance.

Gene	Patient ID	Clinical phenotypes	NM	Variants		Parental origin	Zygosity	Variant (novel/ reported) (PMID)	Gene-related phenotypes (phenotype MIM number)	Variant class
				cDNA	protein					
ARMC4	21	RAL (AVSD, RVOT, PS, SRS, RSS, LCS)	NM_001290021	c.1454G > A c.722T > G	p.G485D p.L241R	Mat/Pat	Compound heterozygous	Novel Novel	Ciliary dyskinesia, primary,23,AR (615451)	P
CCDC114	1	RAL (HLHS, TAPVC, asplenia, RSS, BRB)	NM_144577.3	c.761_768delGCGTCTGG c.88G > A	p.G254Efs*17 p.R30W	Mat/Pat	Compound heterozygous	Novel Novel	Ciliary dyskinesia, primary,20,AR (615067)	P
DNAH11	2	RAL (AVSD, PA, TAPVC, SVC, asplenia, RSS, LSL)	NM_003777.3	c.3470T > G c.7628G > T	p.L1157R p.C2543F	Mat/Pat	Compound heterozygous	32502479; 31040315; 31507630 Novel	Ciliary dyskinesia, primary,7,AR (611884)	P
KMT2D	55	RAL (SV, polysplenia, SA, RSS, ICP)	NM_003482	c.6595delT	p.Y2199fs*64	*De novo*	Heterozygous	Novel	Kabuki syndrome 1,AD (147920)	P
STRA6	17	RAL (SV, PA,SVC, TAPVC, asplenia, RSS, LCS)	NM_001199042.1	c.523+5C > T		Mat/Pat	Homozygous	Novel	Microphthalmia syndromic 9,AR (601186)	P
CCDC40	57	RAL (SV, PS, RSS, LSL)	NM_017950.3	c.2552G > A c.2843_2874de	p.R851Q p.G948fs*69	Mat/Pat	Compound heterozygous	Novel Novel	Ciliary dyskinesia, primary, 15,AR (613808)	LP
DNAH5	25	RAL (SA, SV, PS, SVC, RSS, LSL)	NM_001369	c.1126G > T c.2047C > T	p.A376S p.R683W	Mat/Pat	Compound heterozygous	Novel Novel	Ciliary dyskinesia, primary,3,AR (608644)	LP
FOXC1	43	RAL (LCS, RSS, CAT, SV, AVSD, CAT)	NM_001453	c.1124_1125insCGA	p.G375delinsGD	*De novo*	Heterozygous	Novel	Anterior segment dysgenesis 3,AD (601631) Axenfeld–Rieger syndrome,AD (602482)	LP
FGFR3	70	RAL (PA-VSD, asplenia, BRB, and other extracardiac abnormalities*)	NM_001163213	c.1144G > A	p.G382R	*De novo*	Heterozygous	29080836	Achondroplasia,AD, (100800)	Candidate

*Maternal (Mat); Paternal (Pat); Single right spleen (SRS); Right-sided stomach (RSS); Left-sided liver (LSL); Liver centrally situated (LCS); Single right spleen (SRS); Bilateral right bronchi (BRB); Superior vena cava (SVC); Single ventricle (SV); Single atria (SA); Transportation of great arteries (TGA); Atrioventricular septal defect (AVSD); Isolate cleft palate (ICP); Pulmonary atresia (PA); Pulmonary stenosis (PS);.Common arterial trunk (CAT); Double outlet of the right ventricle (DORV); Hypoplastic left heart syndrome (HLHS); Total anomalous pulmonary venous connection (TAPVC); *Extracardiac abnormalities shown in the last paragraph in the conclusion part.*

(ESP6500). *KMT2D* is a major cause for Kabuki syndrome (OMIM:147920), in which CHD is predominantly presented. Nearly 31–58% of the patient had CHD, including a single ventricle with a common atrium, VSD, ASD, ToF, coarctation of aorta, and patent ductus arteriosus (Niikawa et al., 1988; Digilio et al., 2001). For this reason, the mutation was considered as a pathogenic mutation. Another sequence alteration in coding exon 6 of *FOXC1* is identified in one proband: a heterozygous insertion of three nucleotides, c.11241125insCGA, predicting that the amino acid residue aspartic acid was inserted after position 375, p.G375delinsGD (**Table 3**). This variant has not yet been reported in public databases, including gnomAD, Exac, 1KGP, and ESP6500. The variant was not detected in both parents, confirming that the variant is *de novo*. One *FOXC1*-related phenotype showed in OMIM was Axenfeld–Rieger syndrome, type 3, the features of which included hearing loss, CHD, dental anomalies, developmental delay, and a characteristic facial appearance, central nervous hypoplasia (Maclean et al., 2005). In conclusion, the mutation was considered as a likely pathogenic mutation.

Additionally, we also identified *FGFR3* as a candidate gene. *FGFR3* is known to cause achondroplasia with rhizomelic shortening of the limbs, characteristic facies, exaggerated lumbar lordosis, a limitation of elbow extension, genu varum, and trident hand (Bellus et al., 1995). In our study, the fetus with a *de novo FGFR3* mutation was diagnosed with RA (PA-VSD, right-sided aortic arch accompanied by a mirror branch, bilateral right bronchi, and the absence of the spleen) and achondroplasia (disordered arrangement of cervical vertebrae, missing right radius, abnormal appearance of the right ear, trident right hand, and a higher measurement volume of the lateral ventricle) (Nelson et al., 1988). However, *FGFR3* has not been reported to be the genetic cause of heterotaxy to date and has rarely been shown to be related to cardiac abnormalities, except for one case of *FGFR3*-related craniosynostosis combined with a cleft mitral valve (Agochukwu et al., 2012). Functional experiments have shown that FGFR3 is highly expressed in primary cardiomyocytes (Touchberry et al., 2013) and regulates the intracellular calcium levels and cardiac contractility in cardiac hypertrophy, suggesting that it plays a role in cardiomyocytes or heart development.

Although prenatal diagnosis of heterotaxy has been well studied, the spectrum of underlying gene variants has rarely been reported. Our retrospective cohort study demonstrated the spectrum of cardiac and other associated anomalies in heterotaxy in a prenatal cohort and confirmed previous findings of fetal atrial isomerism. Our results showed a greater incidence of RAI than LAI, which is consistent with other studies in Asian cohorts (Yan et al., 2008). It has been previously reported that patients with RAI have more severe cardiac abnormalities compared with patients with LAI (Buca et al., 2018). We observed that the positive genetic detection rate is 3 times higher in RAI than in LAI, even though the difference is not significant for right and left isomerism (**Supplementary Table S1**). In the future, it is necessary to evaluate this in larger cohorts. Based on these results, we recommended the use of genetic testing, particularly for fetuses diagnosed with RAI. Additionally, the results of this study indicate that the genetic spectrum of fetal heterotaxy must be evaluated in larger studies.

Study Limitations

This study had some limitations. First, the sample size of this study was relatively small. Second, subtle dysmorphic features cannot be determined using fetal ultrasound, and some phenotypes, particularly neurodevelopmental disorders, cannot be determined in the prenatal setting.

Another major limitation of our study is the deviation of our cohort. As a matter of fact, we work in a referral center which is famous for diagnosing fetal cardiac abnomalities, and the patients we see normally have complicate malformations which may be difficult for primary doctors to diagnose. Therefore, the abnormality of the fetus heart we see in our center is more complicated compared to other centers. For this reason, our cohort may not represent the clinical characteristic of whole prenatal heterotaxy patients neither the TOP rate under normal conditions.

If we calculate the termination of pregnancy rate in our cohort, we may be shocked by the extremely high rate. This may be partly explained by the deviation of our cohort we mentioned above, and more importantly, it may be caused by two reasons: first, in China, the decision to choose TOP is considered a private matter of the couple (Lou et al., 2018). Second, high-quality, targeted counseling following prenatal diagnosis of CHD was not available in China in previous years. In China, the TOP following prenatal diagnose was 97.4% in Hunan and 84.8% in Beijing (Yang et al., 2009; Xie et al., 2017; Li et al., 2018). For now, high-quality, targeted counseling following the prenatal diagnosis of CHDs was established and disseminated, and this condition may improve in the future.

CONCLUSION

Clinical phenotyping and next-generation sequencing of a cohort of fetuses with heterotaxy revealed that 1. diagnostic genetic variants were the main genetic cause of fetal heterotaxy, 2. mutations related to PCD were common in the prenatal heterotaxy cohort, and 3. mutations in genes related to Nodal signaling were rare. In contrast to other studies based on animal models or human case reports, our study was conducted on a cohort of prenatal heterotaxy cases, which provided novel insights into the genetic etiology of heterotaxy syndrome. Our results raise the question of whether the insights obtained from animal models and isolated cases with heterotaxy can be extended to fetal patients, especially when determining the candidate genes to be included in sequencing panels. Our results also demonstrate that WES is a promising method for identifying diagnostic genetic variants and novel candidate genes in heterotaxy patients.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/ **Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Medical Ethics Committee of Beijing Anzhen 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 individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

TY and YH designed the study. YF, XH, XZ, LS, XW, YZ, XG, JH and YG collected the study materials and samples. HS and TY performed sequencing experiments. YF, XH, XZ and XW collected and aggregated the data. TY and HS analyzed and interpreted the data. TY drafted the manuscript. All the authors have contributed to the manuscript and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (8217020250 and 82100322), the Beijing Lab for Cardiovascular Precision Medicine, Beijing, China (PXM 2020_014226_000054), and the Beijing Advanced Innovation Center for Big Data-based Precision Medicine (PXM 2021_014226_000026).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.818241/full#supplementary-material

REFERENCES

Agochukwu, N. B., Solomon, B. D., Doherty, E. S., and Muenke, M. (2012). Palatal and Oral Manifestations of Muenke Syndrome (FGFR3-Related Craniosynostosis). *J. Craniofac. Surg.* 23 (3), 664–668. doi:10.1097/scs.0b013e31824db8bb

Baban, A., Cantarutti, N., Adorisio, R., Lombardi, R., Calcagni, G., Piano Mortari, E., et al. (2018). Long-term Survival and Phenotypic Spectrum in Heterotaxy Syndrome: A 25-year Follow-Up Experience. *Int. J. Cardiol.* 268, 100–105. doi:10.1016/j.ijcard.2018.02.050

Bartz, P. J., Driscoll, D. J., Dearani, J. A., Puga, F. J., Danielson, G. K., O'Leary, P. W., et al. (2006). Early and Late Results of the Modified Fontan Operation for Heterotaxy Syndrome. *J. Am. Coll. Cardiol.* 48 (11), 2301–2305. doi:10.1016/j.jacc.2006.07.053

Bellchambers, H. M., and Ware, S. M. (2018). ZIC3 in Heterotaxy. *Adv. Exp. Med. Biol.* 1046, 301–327. doi:10.1007/978-981-10-7311-3_15

Bellus, G. A., Hefferon, T. W., Ortiz de Luna, R. I., Hecht, J. T., Horton, W. A., Machado, M., et al. (1995). Achondroplasia Is Defined by Recurrent G380R Mutations of FGFR3. *Am. J. Hum. Genet.* 56 (2), 368–373.

Buca, D. I. P., Khalil, A., Rizzo, G., Familiari, A., Di Giovanni, S., Liberati, M., et al. (2018). Outcome of Prenatally Diagnosed Fetal Heterotaxy: Systematic Review and Meta-Analysis. *Ultrasound Obstet. Gynecol.* 51 (3), 323–330. doi:10.1002/uog.17546

Carvalho, J. S., Allan, L. D., Chaoui, R., Copel, J. A., DeVore, G. R., Hecher, K., et al. (2013). ISUOG Practice Guidelines (Updated): Sonographic Screening Examination of the Fetal Heart. *Ultrasound Obstet. Gynecol.* 41 (3), 348–359. doi:10.1002/uog.12403

Cowan, J. R., Tariq, M., Shaw, C., Rao, M., Belmont, J. W., Lalani, S. R., et al. (2016). Copy Number Variation as a Genetic Basis for Heterotaxy and Heterotaxy-Spectrum Congenital Heart Defects. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371 (1710). doi:10.1098/rstb.2015.0406

Davis, S. D., Ferkol, T. W., Rosenfeld, M., Lee, H.-S., Dell, S. D., Sagel, S. D., et al. (2015). Clinical Features of Childhood Primary Ciliary Dyskinesia by Genotype and Ultrastructural Phenotype. *Am. J. Respir. Crit. Care Med.* 191 (3), 316–324. doi:10.1164/rccm.201409-1672oc

Diderich, K. E. M., Romijn, K., Joosten, M., Govaerts, L. C. P., Polak, M., Bruggenwirth, H. T., et al. (2021). The Potential Diagnostic Yield of Whole Exome Sequencing in Pregnancies Complicated by Fetal Ultrasound Anomalies. *Acta Obstet. Gynecol. Scand.* 100 (6), 1106–1115. doi:10.1111/aogs.14053

Digilio, M. C., Marino, B., Toscano, A., Giannotti, A., and Dallapiccola, B. (2001). Congenital Heart Defects in Kabuki Syndrome. *Am. J. Med. Genet.* 100 (4), 269–274. doi:10.1002/ajmg.1265

Dykes, I. (2014). Left Right Patterning, Evolution and Cardiac Development. *Jcdd* 1 (1), 52–72. doi:10.3390/jcdd1010052

Fakhro, K. A., Choi, M., Ware, S. M., Belmont, J. W., Towbin, J. A., Lifton, R. P., et al. (2011). Rare Copy Number Variations in Congenital Heart Disease Patients Identify Unique Genes in Left-Right Patterning. *Proc. Natl. Acad. Sci. U.S.A.* 108 (7), 2915–2920. doi:10.1073/pnas.1019645108

Grimes, D. T., and Burdine, R. D. (2017). Left-Right Patterning: Breaking Symmetry to Asymmetric Morphogenesis. *Trends Genet.* 33 (9), 616–628. doi:10.1016/j.tig.2017.06.004

Kearney, H. M., Thorland, E. C., Brown, K. K., Quintero-Rivera, F., and South, S. T. (2011). American College of Medical Genetics Standards and Guidelines for Interpretation and Reporting of Postnatal Constitutional Copy Number Variants. *Genet. Med.* 13 (7), 680–685. doi:10.1097/gim.0b013e3182217a3a

Kim, S.-J., Kim, W.-H., Lim, H.-G., and Lee, J.-Y. (2008). Outcome of 200 Patients after an Extracardiac Fontan Procedure. *J. Thorac. Cardiovasc. Surg.* 136 (1), 108–116. doi:10.1016/j.jtcvs.2007.12.032

Knowles, M. R., Daniels, L. A., Davis, S. D., Zariwala, M. A., and Leigh, M. W. (2013). Primary Ciliary Dyskinesia. Recent Advances in Diagnostics, Genetics, and Characterization of Clinical Disease. *Am. J. Respir. Crit. Care Med.* 188 (8), 913–922. doi:10.1164/rccm.201301-0059ci

Knowles, M. R., Zariwala, M., and Leigh, M. (2016). Primary Ciliary Dyskinesia. *Clin. Chest Med.* 37 (3), 449–461. doi:10.1016/j.ccm.2016.04.008

Li, A. H., Hanchard, N. A., Azamian, M., D'Alessandro, L. C. A., Coban-Akdemir, Z., Lopez, K. N., et al. (2019). Genetic Architecture of Laterality Defects Revealed by Whole Exome Sequencing. *Eur. J. Hum. Genet.* 27 (4), 563–573. doi:10.1038/s41431-018-0307-z

Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I., Roy, S., et al. (2017). Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. *J. Mol. Diagn.* 19 (1), 4–23. doi:10.1016/j.jmoldx.2016.10.002

Li, Y., Rongming, X., and Qiuju, L. (2018). The Analysis of 860 Congenital Heart Disease Cases in Beijing, China. *Women Child. Health China* 33 (03), 588–591.

Lin, A. E., Ticho, B. S., Houde, K., Westgate, M.-N., and Holmes, L. B. (2000). Heterotaxy: Associated Conditions and Hospital-Based Prevalence in Newborns. *Genet. Med.* 2 (3), 157–172. doi:10.1097/00125817-200005000-00002

Liu, S., Song, L., Cram, D. S., Xiong, L., Wang, K., Wu, R., et al. (2015). Traditional Karyotypingvscopy Number Variation Sequencing for Detection of Chromosomal Abnormalities Associated with Spontaneous Miscarriage. *Ultrasound Obstet. Gynecol.* 46 (4), 472–477. doi:10.1002/uog.14849

Lord, J., McMullan, D. J., Eberhardt, R. Y., Rinck, G., Hamilton, S. J., Quinlan-Jones, E., et al. (2019). Prenatal Exome Sequencing Analysis in Fetal Structural

Anomalies Detected by Ultrasonography (PAGE): a Cohort Study. *Lancet* 393 (10173), 747–757. doi:10.1016/S0140-6736(18)31940-8

Lou, S., Carstensen, K., Petersen, O. B., Nielsen, C. P., Hvidman, L., Lanther, M. R., et al. (2018). Termination of Pregnancy Following a Prenatal Diagnosis of Down Syndrome: A Qualitative Study of the Decision-Making Process of Pregnant Couples. *Acta Obstet. Gynecol. Scand.* 97 (10), 1228–1236. doi:10.1111/aogs.13386

Ma, L., Selamet Tierney, E. S., Lee, T., Lanzano, P., and Chung, W. K. (2012). Mutations in ZIC3 and ACVR2B Are a Common Cause of Heterotaxy and Associated Cardiovascular Anomalies. *Cardiol. Young* 22 (2), 194–201. doi:10.1017/s1047951111001181

Maclean, K., Smith, J., Heaps, L. S., Chia, N., Williams, R., Peters, G. B., et al. (2005). Axenfeld-Rieger Malformation and Distinctive Facial Features: Clues to a Recognizable 6p25 Microdeletion Syndrome. *Am. J. Med. Genet.* 132A (4), 381–385. doi:10.1002/ajmg.a.30274

Mohapatra, B., Casey, B., Li, H., Ho-Dawson, T., Smith, L., Fernbach, S. D., et al. (2009). Identification and Functional Characterization of NODAL Rare Variants in Heterotaxy and Isolated Cardiovascular Malformations. *Hum. Mol. Genet.* 18 (5), 861–871. doi:10.1093/hmg/ddn411

Nelson, F. W., Hecht, J. T., Horton, W. A., Butler, I. J., Goldie, W. D., and Miner, M. (1988). Neurological Basis of Respiratory Complications in Achondroplasia. *Ann. Neurol.* 24 (1), 89–93. doi:10.1002/ana.410240117

Niikawa, N., Kuroki, Y., Kajii, T., Matsuura, N., Ishikiriyama, S., Tonoki, H., et al. (1988). Kabuki Make-Up (Niikawa-Kuroki) Syndrome: a Study of 62 Patients. *Am. J. Med. Genet.* 31 (3), 565–589. doi:10.1002/ajmg.1320310312

Pelliccia, J. L., Jindal, G. A., and Burdine, R. D. (2017). Gdf3 Is Required for Robust Nodal Signaling during Germ Layer Formation and Left-Right Patterning. *Elife* 6. doi:10.7554/eLife.28635

Primo, A. E. L., Brock Zacharias, R. S., de Moraes, A. D., da Silva Filho, L. V. R. F., Trujillo, D. R., Yoshida, R. A. M., et al. (2020). Primary Ciliary Dyskinesia as a Cause of Repeating Atelectasis in the Neonatal Period. *Am. J. Case Rep.* 21, e921949. doi:10.12659/AJCR.921949

Reller, M. D., Strickland, M. J., Riehle-Colarusso, T., Mahle, W. T., and Correa, A. (2008). Prevalence of Congenital Heart Defects in Metropolitan Atlanta, 1998-2005. *J. Pediatr.* 153 (6), 807–813. doi:10.1016/j.jpeds.2008.05.059

Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., et al. (2021). Correction: 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.* 23 (11), 2230. doi:10.1038/s41436-021-01150-9

Shapiro, A. J., Davis, S. D., Ferkol, T., Dell, S. D., Rosenfeld, M., Olivier, K. N., et al. (2014). Laterality Defects Other Than Situs Inversus Totalis in Primary Ciliary Dyskinesia. *Chest* 146 (5), 1176–1186. doi:10.1378/chest.13-1704

Sutherland, M. J., and Ware, S. M. (2009). Disorders of Left-Right Asymmetry: Heterotaxy and Situs Inversus. *Am. J. Med. Genet.* 151C (4), 307–317. doi:10.1002/ajmg.c.30228

Touchberry, C. D., Green, T. M., Tchikrizov, V., Mannix, J. E., Mao, T. F., Carney, B. W., et al. (2013). FGF23 Is a Novel Regulator of Intracellular Calcium and Cardiac Contractility in Addition to Cardiac Hypertrophy. *Am. J. Physiology-Endocrinology Metab.* 304 (8), E863–E873. doi:10.1152/ajpendo.00596.2012

Xie, D., Wang, H., Liu, Z., Fang, J., Yang, T., Zhou, S., et al. (2017). Perinatal Outcomes and Congenital Heart Defect Prognosis in 53313 Non-selected Perinatal Infants. *PLoS One* 12 (6), e0177229. doi:10.1371/journal.pone.0177229

Yan, Y. L., Tan, K. B., and Yeo, G. S. (2008). Right Atrial Isomerism: Preponderance in Asian Fetuses. Using the Stomach-Distance Ratio as a Possible Diagnostic Tool for Prediction of Right Atrial Isomerism. *Ann. Acad. Med. Singap* 37 (11), 906–912.

Yang, X. Y., Li, X. F., Lü, X. D., and Liu, Y. L. (2009). Incidence of Congenital Heart Disease in Beijing, China. *Chin. Med. J. (Engl)* 122 (10), 1128–1132.

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership