Pharmacogenomics and Personalized Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

NOTCH2, ATIC, MRI1, SLC6A13, ATP13A2 Genetic Variations are Associated with Ventricular Septal Defect in the Chinese Tibetan Population Through Whole-Exome Sequencing

Xiaohui Zhang[1–3], Da Zhen[4], Xuemei Li[1,2,5], Faling Yi[1,2,5], Zhanhao Zhang[1,2,5], Wei Yang[1,2,6], Xuguang Li[1,2,5], Yemeng Sheng[1,2,5], Xiaoli Liu[1,2,5], Tianbo Jin[1,2,5], Yongjun He[1,2,5]

[1]Key Laboratory of High Altitude Hypoxia Environment and Life Health, School of Medicine, Xizang Minzu University, Xianyang, People's Republic of China; [2]Key Laboratory of Molecular Mechanism and Intervention Research for Plateau Diseases of Tibet Autonomous Region, School of Medicine, Xizang Minzu University, Xianyang, People's Republic of China; [3]Department of Ultrasound, the Affiliated Hospital of Xizang Minzu University, Xianyang, People's Republic of China; [4]Department of Medical, Tibet Autonomous Region Maternity and Children's Hospital, Lhasa, People's Republic of China; [5]School of Medicine, Xizang Minzu University, Xianyang, People's Republic of China; [6]Department of Emergency, the Affiliated Hospital of Xizang Minzu University, Xianyang, People's Republic of China

Correspondence: Tianbo Jin; Yongjun He, Xizang Minzu University, #6 East Wenhui Road, Xianyang, Shaanxi, 712082, People's Republic of China, Email tianbo__jin@163.com; 545139647@qq.com

Background: Ventricular septal defect (VSD) is the most common congenital cardiac abnormality in children and the second most common in adults. This study aimed to explore the potentially causative genes in VSD patients in the Chinese Tibetan population, and to provide a theoretical basis for the genetic mechanism of VSD.

Methods: Peripheral venous blood was collected from 20 VSD subjects, and whole-genome DNA was extracted. High-throughput sequencing was performed on qualified DNA samples using whole-exome sequencing (WES) technology. After filtering, detecting, and annotating qualified data, single nucleotide variations (SNVs) and insertion-deletion (InDel) markers were analyzed, and data processing software such as GATK, SIFT, Polyphen, and MutationTaster were used for comparative evaluation and prediction of pathogenic deleterious variants associated with VSD.

Results: A total of 4793 variant loci, including 4168 SNVs, 557 InDels and 68 unknown loci and 2566 variant genes were obtained from 20 VSD subjects through bioinformatics analysis. According to the screening of the prediction software and database, the occurrence of VSD was predicted to be associated with five inherited pathogenic gene mutations, all of which were missense mutations, including *NOTCH2* (c.1396C >A:p.Gln466Lys), *ATIC* (c.235C >T:p.Arg79Cys), *MRI1* (c.629G >A:p.Arg210Gln), *SLC6A13* (c.1138G >A:p.Gly380Arg), *ATP13A2* (c.1363C >T:p.Arg455Trp).

Conclusion: This study demonstrated that *NOTCH2, ATIC, MRI1, SLC6A13, ATP13A2* gene variants were potentially associated with VSD in Chinese Tibetan population.

Keywords: ventricular septal defect, genes, genetic variation, whole-exome sequencing

Introduction

Congenital heart disease (CHD) refers to cardiovascular malformations caused by abnormal development of cardiac vessels during the fetal period, which is the most common congenital dysplasia and also the main cause of non-infectious death in newborns and infants.[1] CHD includes atrial septal defect (ASD), ventricular septal defect (VSD), pulmonary atresia (PA), patent ductus arteriosus (PDA), tetralogy of Fallot (TOF) and other cardiac malformations, among which the incidence of VSD is the highest, especially isolated VSD, and 30% of cases are combined with other malformations.[2] The incidence of isolated muscular VSD has been reported to be 5.7% in preterm infants and 1.1–5.3% in term infants.[3] While many VSD can close spontaneously, if they do not close, large defects can lead to deleterious complications such

Received: 11 January 2023
Accepted: 31 March 2023
Published: 27 April 2023

as pulmonary arterial hypertension (PAH), ventricular dysfunction, and an increased risk of arrhythmias.[4] The low pressure and hypoxia in the Qinghai-Tibet Plateau can cause altitude sickness, and the incidence of CHD will increase with the increase of altitude.[5] Therefore, it is inferred that the hypoxic and low pressure environment at plateau is closely related to the pathogenesis of CHD. According to epidemiology, the incidence of neonatal CHD in high-altitude areas is about 20 times higher than that in low-altitude areas.[6] In the Columbia region, rate of CHD has also been found to be much lower in plains than that in high-altitude regions.[7] Ethnic and regional differences have significant roles in the occurrence of CHD, and the data have indicated that there may be ethnic differences in the prevalence of CHD in China.[8] Malg et al have found that in China, the incidence of CHD in plain areas is not as high as that in Huangnan Tibetan Autonomous Prefecture and Yushu Tibetan Autonomous Prefecture in Qinghai Province.[9] The above studies have shown that CHD in the Qinghai-Tibet Plateau area in China is characterized by high incidence and high mortality. The pathogenesis of CHD is not yet clear, and researchers think that there are two main factors for its occurrence, environmental factors and genetic factors. Among the genetic factors, gene mutation plays an important role in the occurrence of VSD. The study by Basson et al has demonstrated for the first time that *TBX5* mutations is associated with the development of hereditary CHD.[10] Subsequently, some gene mutations, including *NOTCH1, CITED2, NDRG4*, etc. were found to be related to VSD.[11–13] Ji et al have reported that *NOTCH1* rs6563G > A variant may affect the regulation of miR-3691-3p on the *NOTCH1* gene, thereby influencing the susceptibility to VSD.[11] The study by Zheng et al has found that functional changes caused by variants in the promoter region of the *CITED2* gene may affect a set of downstream genes and pathways and eventually lead to VSD.[12] Peng et al have revealed that the p.T256M variant in *NDRG4* is a pathogenic variant associated with the impaired proliferation of human cardiomyocytes and cell cycle arrest, which may be involved in the pathogenesis of VSD.[13] Taken together, genetic factors are crucial in the occurrence of VSD. However, some genetic variants associated with VSD in the Chinese Tibetan population remain to be further identified. At present, there are few studies on molecular mechanisms of VSD in the high-altitude environment.

Whole-exome sequencing (WES) can detect low-frequency variants in the coding regions of a patient's genes, rare variants with different effects on disease, and disease-related genetic variants. This study intended to perform WES analysis on 20 Tibetan subjects with VSD, and use bioinformatics technology to screen rare or low-frequency variants in VSD genes and annotate them with pathogenicity, so as to further provide accurate molecular diagnosis for VSD.

Materials and Methods

Study Participants

All participants under the age of 18 years in this study were diagnosed with CHD in the Second People's Hospital of Tibet Autonomous Region, and a total of 20 subjects with congenital heart type VSD were screened. This study fully followed the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Xizang Minzu University (201808). Informed consent form was obtained from parental/legal guardian of all participants. The clinical data of all subjects are shown in Table 1.

Inclusion criteria for the subjects in this study: (1) Three family generations of all enrolled subjects were free from VSD, CHD and other genetic disease. (2) Subjects with clinically diagnosed VSD. (3) Subjects themselves and their parents who signed informed consent and underwent WES examination. Exclusion criteria: (1) Subjects with other CHD. (2) Subjects with FCHD, FPAH, respiratory system-related PAH or other familial hereditary diseases. (3) Any other cardiovascular disease, chronic anemia, thyroid disease, electrolyte imbalance, systemic immune disease, malignant tumor, or other diseases that may cause VSD.

Collection of Blood Samples and WES

With the consent of subjects and their guardians, about 5-mL of fasting peripheral venous blood was collected in anticoagulation tube and stored at ultra-low temperature refrigerator (−80 degrees C). DNA extraction kit was used to extract genomic DNA from blood samples, and Nanodrop 2000/Qubit was carried out to detect the quality and concentration of genomic DNA (the total amount of samples ≥ 1.5 μg, the concentration ≥ 50 ng/μL and OD260/280 = 1.8~2.0). The genomic DNA was fragmented using Covaris, so that most of the genomic DNA was between 100 and

Table 1 Sample Information of Enrolled Patients

Sample Number	Gender	Age	Diagnostic Result	Echocardiogram Report
VSD-1	Female	2.9 years	CHD	CHD: After VSD operation, left atrial enlargement, CDFI: no shunt color beams at the level of atrium, ventricle, and great vessels, tricuspid regurgitation (small amount)
VSD-2	Male	7 months	CHD	CHD: Tricuspid valve atresia, VSD (perimembranous type 7 mm) ventricular level left-to-right shunt, ASD (central type 11 mm) atrial level left-to-right shunt, consider: complete endocardial cushion defect may be pulmonary valve stenosis
VSD-3	Female	3.7 years	CHD	CHD: After right ventricular double outlet + VSD + PAH surgery: VSD (30 mm) ventricular level bidirectional shunt, ASD (12 mm) atrial level left to right, PAH (80 mmHg)
VSD-4	Male	10 years	CHD	CHD: After VSD, CDFI: no shunting color beams at the ventricular level, PAH (43 mmHg), tricuspid regurgitation (moderate volume)
VSD-5	Male	7 months	CHD	CHD: VSD (membranous 6 mm), CDFI: left-to-right shunt at ventricular level, PAH (46 mmHg), tricuspid regurgitation (small amount)
VSD-6	Female	6 years	CHD	VSD (membranous 5.6 mm), CDFI: left-to-right shunt at ventricular level, patent foramen ovale (3 mm), CDFI: left-to-right shunt at atrial level, PAH (90 mmHg), tricuspid regurgitation, three cuspid regurgitation (moderate-large)
VSD-7	Male	6.11 years	CHD	CHD: VSD (membranous 3 mm), CDFI: left-to-right shunt at ventricular level
VSD-8	Female	4 years	CHD	Postoperative VSD: no shunting color beam at ventricular level, PAH (38 mmHg), tricuspid regurgitation (small amount)
VSD-9	Female	11 years	CHD	CHD: After VSD + tricuspid valvuloplasty, right atrium and right ventricle enlargement, tricuspid valve depression deformity? (Tricuspid septal valve is 20 mm away from atrioventricular valve), PAH (50 mmHg), tricuspid regurgitation (moderate volume)
VSD-10	Female	4.7 years	CHD	CHD: Postoperative VSD: no shunt color beam at the ventricular level, tricuspid regurgitation (small amount)
VSD-11	Female	2.2 years	CHD	CHD: VSD (membranous 4 mm), CDFI: left-to-right shunt at ventricular level
VSD-12	Male	3.9 years	CHD	Postoperative VSD: no shunt color beam at the ventricular level, tricuspid regurgitation (small amount)
VSD-13	Male	9 months	CHD	CHD: VSD (membranous 5 mm), CDFI: left-to-right shunt at ventricular level, PAH (40mmHg), tricuspid regurgitation (small amount)
VSD-14	Female	4.5 years	CHD	CHD: Postoperative VSD: no shunting color beam at ventricular level, tricuspid regurgitation (small amount)
VSD-15	Female	2.9 years	CHD	CHD: VSD (membrane 7 mm), CDFI: left-to-right shunt at ventricular level, mitral insufficiency, mitral regurgitation (moderate volume), PAH (80 mmHg), tricuspid regurgitation (moderate volume), Pulmonary regurgitation (small amount)
VSD-16	Female	2.1 years	CHD	CHD: VSD (membranous 4 mm), CDFI: left-to-right shunt at ventricular level
VSD-17	Male	4.9 years	CHD	CHD: After VSD: ventricular level residual small bundle shunt (2 mm), tricuspid regurgitation (small amount)
VSD-18	Female	2.3 years	CHD	CHD: After VSD surgery, no shunting color beams were seen at the ventricular level
VSD-19	Female	1.5 months	CHD	CHD: VSD (membranous 4 mm), CDFI: left-to-right shunt at ventricular level, patent foramen ovale (2 mm), CDFI: left-to-right shunt at atrial level, mitral regurgitation (small amount), PAH (36 mmHg), tricuspid regurgitation (small amount)
VSD-20	Female	8 years	CHD	CHD: After VSD surgery, no shunt color beams were seen at the ventricular level, PAH (44 mmHg), tricuspid regurgitation (small amount)

Abbreviations: CHD, Congenital heart disease; VSD, ventricular septal defect; CDFI, color Doppler flow imaging; ASD, atrial septal defect; PAH, Pulmonary Arterial Hypertension.

https://doi.org/10.2147/PGPM.S404438 DovePress

500 bp in length. DNA libraries were prepared by adding "A" bases to the 3' ends of DNA fragments. Whole-exome capture was performed using the SureSelectXT Reagent kit and probe hybridization was performed with the library using the SureSelectXT Human All Exon Kit V6. Concentration (> 5 ng/μL). The fragment length (300~400 bp) of libraries were detected using Qubit and Agilent 2100 Bioanalyzer. Additionally, high-throughput sequencing was conducted on the Illumina Hiseq platform using 2×150 bp paired-end sequencing mode to obtain FastQ data.

WES Data Analysis

The raw data of each sample was compared with the human reference genome (UCSC, https://hgdownload.soe.ucsc.edu/downloads.html) using the mem algorithm of BWA (http://bio-bwa.sourceforge.net/), and preliminary alignment results in BAM format (http://samtools.sourceforge.net/) were obtained. Picard software (https://broadinstitute.github.io/picard/) was used to count the alignment results of each sample, including the number and ratio of sequences on the alignment, Q20 and Q30 sequence ratio, average coverage depth, etc. In order to accurately identify SNVs and InDels, we used the GATK standard procedure (https://software.broadinstitute.org/gatk/best-practices/) to correct base quality, repeat sequences caused by PCR amplification, and misalignments generated by InDels. The SNV/InDel of each sample was detected using the GATK HaplotypeCaller method (https://software.broadinstitute.org/gatk/best-practices/) and filtered according to the screening protocol recommended by the software. In order to quickly find the most biologically meaningful SNV/InDel loci from a large amount of variant information under the premise of ensuring the accuracy of research data, we used ANNOVAR (http://annovar.openbioinformatics.org/en/latest/) to compare all SNV/InDel sites with these in the latest published population databases (dbSNP, 1000 Genomes, esp6500, ExAC03, ExAC03_EAS, gnomAD_exome_EAS), disease databases (HGMD, InterVar, MGI, HPO) and other known information for alignment analysis, so as to evaluate the mutation frequencies, functional characteristics, conservation, pathogenicity of these SNV/InDel sites, etc. Conservation and protein hazard were predicted by SIFT, PolyPhen-2, Mutation Taster, Cadd, Dann to analyze whether the mutation would affect the structure and function of the protein. In detail, SIFT prediction results can be divided into two types: D (damaging) and T (tolerated); PolyPhen prediction results can be divided into three types: B (benign), P (possibly damaging), and D (probably damaging); Mutation Taster prediction results can be divided into four types: A (disease causing automatic), D (disease causing), N (polymorphism), and P (polymorphism automatic). Cadd and Dann predictions were used to calculate SNV/InDel risk scores. And the higher the score, the higher the probability that the mutation site is deleterious.

Bioinformatics Analysis

Firstly, the filtering criteria for low-frequency functional mutations included population database frequency filtering (1000 Genome ≤ 0.01, ExAC03_EAS ≤ 0.01, gnomAD_exome_EAS ≤ 0.01) and functional mutation filtering (retaining mutations located in exons or splice regions, nonsynonymous SNV, stopgain, and other non-synonymous SNV types of mutations). The second step was to prioritize these SNV/InDel sites (divided into First1, First2, Second, Third, according to the reliability from high to low). Finally, the screening strategy for candidate mutations was generated as follows. (1) Only keep mutations with priority class First1. (2) Screening of known pathogenic genes. According to the standard English name of the disease, the gene and disease phenotype databases (HGMD, HPO, MGI, ClinVar) were used to check whether there is a known causative gene. (3) The loci with lower frequency and stronger pathogenicity were preferentially selected. Combined with information such as the incidence of disease, only the variant sites with the frequency screening threshold < 0.01 were retained according to the database frequency. Reference frequency databases included Freq_Alt (1000g), ExAC03_EAS, esp6500, gnomAD_exome_EAS, etc. Meanwhile, combining with the predicted of pathogenicity software (score of D and Dann ≥ 0.93), the loci were screened.

Results

Whole-Exome Sequencing Data and the Reference Sequence Alignment

The raw data of 20 VSD samples were obtained through the Illumina HiSeq sequencing platform. The statistical results of WES data are shown in Table S1, and a total of 1,237,197,436 reads were obtained. The clean reads of each sample had high Q20 and Q30, and the average proportion of Q30 was 95.83%, indicating good sequencing quality of sequencing data.

The quality of the paired-end sequenced bases of all samples met the standard (Figure S1). As shown in Figure S2, qualified DNA will form a library mainly distributed between 300 and 400 bp in size by randomizing fragments. The comparison results of each sample were counted by Picard software (Table S2), and it was found that after removing low-quality reads, the proportion of bases with a sequencing depth of 10X was > 90%. When the base coverage depth of a locus reached more than 10X, the SNV detected at that locus was considered to be relatively credible.

Variant Detection

By using the GATK HaplotypeCaller method, we found a total of 225,735 mutation sites, including 196,458 SNV mutation sites and 29,277 InDel sites. According to the classification statistics of SNV/InDel loci relative to the genome, the distribution ratio of SNV/InDel loci in different regions was obtained, as shown in Figure 1, mainly including intronic mutations (53.74%), exonic mutations (22.92%), splicing mutations (4.91%), etc. Classification statistics were performed according to the functional types of SNV/InDel sites, as shown in Figure 2. Function types mainly includes missense mutation or nonsynonymous mutation (50.43%), synonymous mutation (44.3%), unknown (0.99%) etc. We further counted the genotypes (homozygous/heterozygous) of SNV/InDel loci according to genotype classification, and the number of SNV/InDel loci of different genotypes in 20 patients is shown in Table 2.

Bioinformatics Analysis of Identified Gene Variants

We obtained a total of 4793 variant loci (including 4168 SNVs, 557 InDels and 68 unknown loci) and 2566 genes, of which 3232 belonged to heterozygous mutations, 62 belonged to homozygous mutations, 24 belonged to heterozygous and homozygous mutations, and all variant loci were filtered by frequency in population database and mutations' function. Confidence site classifications were obtained according to the criteria of site priority classification: First1 (1396 sites), First2 (1218 sites), Second (1891 sites) and Third (288 sites). According to the candidate mutation screening strategy, a total of five predicted deleterious genes and five mutation sites were screened, and they were all located in the first1 with priority. The frequencies of the five genes in the population database were all far below 0.01, indicating that the selected genes were rare in normal people, and the predicted results of software were all harmful genes (Table 3). The predicted deleterious genes, including five genes related to VSD, were located in the exon region with nonsynonymous mutation SNVs, namely *NOTCH2* (c.1396C >A:p.Gln466Lys), *ATIC* (c.235C >T:p.Arg79Cys), *MRI1* (c.629G >A:p.Arg210Gln), *SLC6A13* (c.1138G >A:p.Gly380Arg), and *ATP13A2* (c.1363C >T:p.Arg455Trp). In the gene inheritance

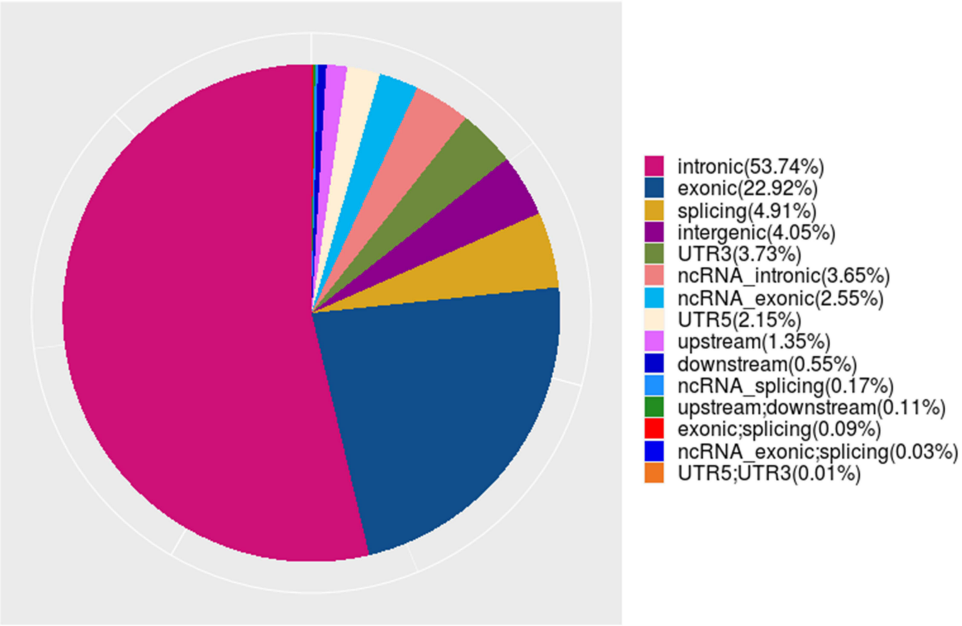

Figure 1 Distribution ratio of SNV/InDel sites in different regions.

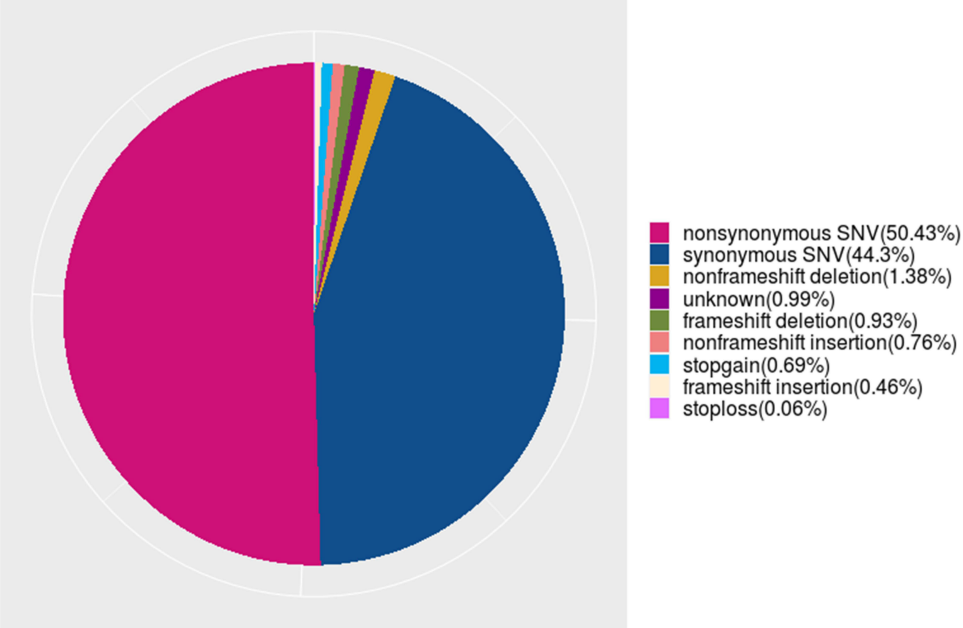

Figure 2 Distribution ratio of SNV/InDel sites for different functional types.

pattern, *NOTCH2* was autosomal dominant; *ATIC* and *ATP13A2* were autosomal recessive (Table 4). Subsequently, we further studied the functions of SNP through the RegulomeDB database. The results showed that rs141935585, rs367615313, rs141094096 and rs377179026 could affect transcription factor (TF) binding and DNase. And rs371445094 was associated with TF binding, motif and DNase (Table 4). According to the results of function annotation of predicted disease-associated genes, it was found that *NOTCH2* was associated with VSD, and *ATIC* was associated with ASD and ventricular disease. *MRI1* was annotated as pathogenic (Table 5).

Table 2 Number of SNV/InDel Loci for Different Genotypes

Sample	Het_SNV	Hom_SNV	Novel_SNV	Het_InDel	Hom_InDel	Novel_InDel	Het_SNV/Hom_SNV
VSD-1	54830	37066	1842	7967	3742	978	1.479253224
VSD-2	53015	38341	1887	7802	3849	950	1.382723455
VSD-3	54107	36634	1912	7779	3648	1067	1.476961293
VSD-4	56339	39923	2109	9221	4365	1199	1.411191544
VSD-5	55212	39111	2057	8074	4087	985	1.411674465
VSD-6	57515	38627	1945	9196	4154	1142	1.488984389
VSD-7	55699	38855	1878	8547	4018	1026	1.433509201
VSD-8	56486	38165	1990	8131	4036	944	1.480047164
VSD-9	53427	37515	1737	6976	3663	768	1.42415034
VSD-10	54386	37075	1919	7492	3688	803	1.466918409
VSD-11	53930	35809	1798	6756	3459	708	1.506045966
VSD-12	52216	37604	1718	6647	3674	787	1.388575683
VSD-13	49489	35433	1756	5805	3307	637	1.396692349
VSD-14	51391	36103	1804	6370	3417	759	1.423455115
VSD-15	53257	36629	1707	7047	3645	799	1.453957247
VSD-16	51707	36832	1694	6563	3498	749	1.403860773
VSD-17	51174	36150	1762	5935	3537	643	1.41560166
VSD-18	55913	38715	2073	9230	4058	1082	1.444220586
VSD-19	55254	37057	1876	7933	3711	893	1.491054322
VSD-20	54505	37566	1831	7660	3858	966	1.45091306

Abbreviations: SNV, single nucleotide variations; InDel, insertion-deletion; VSD, ventricular septal defect; Het_SNV, SNV of heterozygous genotype; Hom_SNV, SNV of homozygous genotype; Novel_SNV, new (not annotated by dbSNP) SNV.

Table 3 Predicted Gene Frequencies and Harmful Prediction Tables

Gene	First Priority	Freq_Alt (1000g)	ExAC03_EAS	esp6500	gnomAD_exome_EAS	SIFT Pred	POLYPhen V2 Pred	MutationTaster Pred	Cadd Raw	Dann
NOTCH2	First1	0.00159744	0	0.0002	0	D	D	D	2.971	0.995
ATIC	First1	0.000199681	0	0.000077	0.00005437	D	D	D	4.162	0.999
MRI1	First1	0.000399361	0.0001	0.000077	0.00005452	D	D	D	4.39	1
SLC6A13	First1	0.000199681	0.0003	0.000077	0.0002	D	D	D	5.478	0.999
ATP13A2	First1	0.000199681	0.0005	0.000077	0.0004	D	D	D	4.313	0.999

Abbreviations: First Priority, the best considered gene; Freq_Alt (1000g), the frequency of non-reference alleles in the 1000 Genome Project; ExAC03_EAS, allele frequency data of 65,000 exons in East Asia; esp6500, from ESP (NHLBI GO Exome Sequencing Project) Frequencies in 6500 samples of whole exome sequencing; gnomAD_exome_EAS, gnomAD data in East Asia; SIFT Pred, POLYPhen V2 Pred, Mutation Taster Pred, predicting the effect of gene variation on function, protein hazard prediction; D, pathogenic; Cadd Raw and Dann, SNV deleterious score (the higher the value, the more deleterious the mutation is).

https://doi.org/10.2147/PGPM.S404438

Table 4 Genetic Information for Predicted Genes

SNV NO.	Gene	Chrs	SNP ID	Gene Region	Function	Predicted Protein Variants	Genetic Model	SNP Function
SNV10061	*NOTCH2*	chr1	rs141935585	Exonic	Nonsynonymous SNV	c.1396C>A:p.Gln466Lys:exon8	AD	TF binding, DNase peak
SNV01363	*ATIC*	chr2	rs367615313	Exonic	Nonsynonymous SNV	c.235C>T:p.Arg79Cys:exon4	AR	TF binding, DNase peak
SNV08935	*MRI1*	chr19	rs141094096	Exonic	Nonsynonymous SNV	c.629G>A:p.Arg210Gln:exon4	–	TF binding, DNase peak
SNV13971	*SLC6A13*	chr12	rs377179026	Exonic	Nonsynonymous SNV	c.1138G>A:p.Gly380Arg:exon10	–	TF binding, DNase peak
SNV01397	*ATP13A2*	chr1	rs371445094	Exonic	Nonsynonymous SNV	c.1363C>T:p.Arg455Trp:exon15	AR	TF binding, any motif, DNase peak

Abbreviations: SNV NO., single nucleotide variation number; Chrs, chromosome number; SNP ID, the number of the locus in the dbSNP of NCBI; Gene Region, the gene region where the variation is located; Function, gene function; Predicted Protein Variants, Using snpEFF for mutation annotation, the amino acid changes of the protein caused by the gene predicted from the reference sequence; Genetic model, Gene inheritance model; nonsynonymous SNV, missense mutation; AD, autosomal dominant; AR, autosomal recessive; TF binding, Transcription factor binding.

Table 5 Predicting Gene Disease Functional Annotation

Gene	InterVar	HGMD	HGMD_gene	HPO	MGI
NOTCH2	Uncertain significance	–	Atrioventricular septal defect; Bicuspid aortic valve; PAH	VSD; ASD; TOF; Peripheral pulmonary artery stenosis; Patent ductus arteriosus	Cardiovascular system phenotype; Respiratory system phenotype; Hematopoietic system phenotype
ATIC	Uncertain significance	–	Left ventricular obstruction with extracardiac anomalies and Neurodevelopmental disorder; Transposition of the great arteries with extra-cardiac anomalies and neurodevelopmental disorder	ASD	-
MRI1	Uncertain significance	DM	-	-	-
SLC6A13	Uncertain significance	-	-	-	-
ATP13A2	Uncertain significance	-	-	-	-

Abbreviations: InterVar, Mutation clinical interpretation information according to the American College of Medical Genetics and Genomics (ACMG) scoring rules; HGMD, Human Gene Mutation Database; HPO, Human Phenotype Ontology; MGI, Mouse Genome Informatics; DM, Pathogenic mutation; PAH, Pulmonary Arterial Hypertension; VSD, ventricular septal defect; ASD, atrial septal defect; TOF, tetralogy of fallot.

Discussion

In this study, the genetic etiology of VSD subjects in the Chinese Tibetan population was studied by WES technology, and rare or low-frequency gene mutations related to VSD, namely five disease-causing genes with heterozygous mutations were found. This study is the first to identify pathogenic genes associated with VSD in the Chinese Tibetan population. According to the results of function annotation of predicted disease-associated genes, it was found that *NOTCH2* was associated with VSD, and *ATIC* was associated with ASD. Meanwhile, *MRI1* was annotated as pathogenic.

The *NOTCH* receptor family is highly conserved and belongs to the membrane protein receptor family, including 4 homologous receptors in mammals, namely *Notch1~4*. Many studies have shown that *NOTCH2* plays an extremely important role in cardiac development, and its deletion may lead to diseases such as reduced cardiac compaction, cardiac hypertrophy, and VSD, due to its involvement in cardiomyocyte apoptosis, proliferation, and differentiation and other important cellular biological processes.[14–16] For example, miR-29b-3p inhibits cardiomyocyte proliferation through *NOTCH2*, which demonstrates the important role of *NOTCH2* in cardiac development.[14] Some studies have indicated that *NOTCH2* is essential for the correct formation of the cardiac outflow tract in the proliferation of cardiac neural crest-derived smooth muscle cells.[17] Besides, another studied have showed that the use of Cre-lox technology to specifically inhibit *NOTCH* signaling in neural crest cells (NCCs) results in outflow tract defects, such as VSD, pulmonary stenosis, ASD, and OA.[18] Moreover, mutations in *NOTCH2* cause alagille syndrome, pulmonary artery stenosis and CHD.[19] Taken together, *NOTCH2* mutations may play an important role in the occurrence of VSD. Through a comprehensive review of the literature, we found that the *NOTCH* signaling pathway related to the *NOTCH2* gene has been reported in VSD. *NOTCH* signaling is an evolutionarily conserved pathway whose aberrations contribute to the development of cardiac malformations,[20] such as bicuspid aortic valve disease, heart valve calcification, alagille syndrome, and VSD.[21] Mutations in the *NOTCH* receptors and their ligands have been identified as the cause of CHD in humans, suggesting the importance of *NOTCH* signaling during cardiac development.[17] Taken the above, we speculated that *NOTCH2* could affect VSD through *NOTCH* signaling pathway.

5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase (*ATIC*) is a novel rate-limiting enzyme in the purine biosynthesis pathway, catalyzing the last 2 reactions of the purine synthesis pathway.[22] The abnormal expression of *ATIC* is closely related to the prognosis of hepatocellular carcinoma patients and can cause significant changes in the

proliferation and migration of hepatocellular carcinoma cells.[23] Other studies have also shown that the abnormal expression of *ATIC* is closely related to the occurrence of various diseases, including multiple myeloma,[24] lung cancer,[25] lymphoma,[26] etc. We found that *ATIC* gene was associated with VSD in the Chinese Tibetan population, and it was associated with ASD, ventricular disease. It has been suggested that synthetase deficiency may lead to pregnancy complications through reducing purine synthesis and cell proliferation.[27] There is an undeniable relationship between the incidence of VSD and the developmental status of *ATIC* in the mother, and it is inferred that this enzyme may affect the pathogenesis of VSD through purine synthesis.

The *MRI1* gene, encoding the translation initiation factor eIF-2B subunit alpha/beta/delta-like protein 2, belongs to a little-known protein family in eukaryotes, the eIF2B-related protein.[28] The *MRI1* gene serves as a candidate locus for infantile epilepsy with severe cerebral degeneration,[29] and it is also significant in the differential gene expression of peripheral blood DNA methylation in infants with maternal asthma.[30] The role of this gene in VSD has not been found, but is pathogenic in the functional annotation of the disease-related genes. Therefore, we speculated that *MRI1* gene may play an important role in VSD risk, which needs further validation.

SLC6A13, also known as *GAT2*, belongs to the gene encoding a high-affinity GABA transporter.[31] It is mainly expressed in the liver, kidney and other peripheral tissues such as testis, retina and lung. Currently, there are few studies on the *SLC6A13* gene, not mention to the studies on its expression in the heart.[32] We detected *SLC6A13* gene in VSD for the first time, but its specific underlying mechanism needs further exploration.

ATPase cation transporting 13A2 (*ATP13A2*) is a late endolysosomal transporter that encodes a lysosomal transmembrane P5B-type ATPase,[33] and it is genetically linked to a range of neurodegenerative diseases, for instance, Kufor-Rakeb syndrome, early-onset PD, neuronal ceroid lipofuscinosis, juvenile parkinsonism, and complex hereditary spastic paraplegia.[34,35] We detected *ATP13A2* gene in VSD patients for the first time. *ATP13A2* deficiency causes mitochondrial and lysosomal damage in various models,[36] including mitochondrial fragmentation, increased oxygen consumption, and mitochondrial DNA damage.[37] Hypoxia stimulates *ATP13A2* transcription through HIF1α, and hypoxia signaling plays an important role in regulating *ATP13A2* gene expression.[38] The subjects in this study live in a plateau hypoxic environment for a long time, so we may speculate that hypoxia regulates the expression of *ATP13A2* gene to control the occurrence of VSD.

This study only carried out a preliminary screening test, and did not verify key genes and pathways and elucidate the mechanism of VSD through cell culture or animal models, which is still to be improved in the future. In future research work, we will continue to expand the sample size of the included VSDs to conduct more comprehensive researches.

Conclusion

In this study, WES biological analysis was performed on Tibetan subjects with VSD, and five pathogenic genes with low-frequency mutations were found: *NOTCH2, ATIC, MRI1, SLC6A13, ATP13A2*. It provides a bioinformatics basis for exploring the relationship between gene mutations and the pathogenesis of VSD, and provides clues for the molecular mechanism of VSD and new treatment strategies.

Data Sharing Statement

The datasets used or analyzed during the current study are available from the corresponding author of Tianbo Jin on reasonable request.

Ethical Approval and Consent to Participate

This study fully followed the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Xizang Minzu University (201808). All participants under the age of 18 years of age, parental/legal guardian consent was obtained.

Acknowledgments

We are very grateful to all the volunteers, clinicians and hospital staff who participated in this study.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This work were supported by National Natural Science Fund (81860600), Open Fund for Key Laboratory of Plateau Hypoxia Environment and Life and Health (XZMU-2022M-H06), Scientific research project of Xizang Minzu University (2022MDY011), Natural Science Foundation of Tibet Autonomous Region (XZ2018ZRG-79(Z)), Postgraduate research innovation and practice project of Xizang Minzu University (Y2022099).

Disclosure

The authors declare that they have no competing interests.

References

1. Clark KL, Yutzey KE, Benson DW. Transcription factors and congenital heart defects. *Annu Rev Physiol*. 2006;68:97–121. doi:10.1146/annurev.physiol.68.040104.113828
2. Yu H, Smallwood PM, Wang Y, Vidaltamayo R, Reed R, Nathans J. Frizzled 1 and frizzled 2 genes function in palate, ventricular septum and neural tube closure: general implications for tissue fusion processes. *Development*. 2010;137(21):3707–3717. doi:10.1242/dev.052001
3. Miyake T. A review of isolated muscular ventricular septal defect. *World J Pediat*. 2020;16(2):120–128. doi:10.1007/s12519-019-00289-5
4. Dakkak W, Oliver TI. Ventricular septal defect. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing Copyright © 2022, StatPearls Publishing LLC; 2022.
5. González-Andrade F. High altitude as a cause of congenital heart defects: a medical hypothesis rediscovered in Ecuador. *High Alt Med Biol*. 2020;21(2):126–134. doi:10.1089/ham.2019.0110
6. Li JJ, Liu Y, Xie SY, et al. Newborn screening for congenital heart disease using echocardiography and follow-up at high altitude in China. *Int J Cardiol*. 2019;274:106–112. doi:10.1016/j.ijcard.2018.08.102
7. García A, Moreno K, Ronderos M, Sandoval N, Caicedo M, Dennis RJ. Differences by altitude in the frequency of congenital heart defects in Colombia. *Pediatr Cardiol*. 2016;37(8):1507–1515. doi:10.1007/s00246-016-1464-x
8. Han S, Wei CY, Hou ZL, et al. Prevalence of congenital heart disease amongst schoolchildren in Southwest China. *Indian Pediatr*. 2020;57(2):138–141. doi:10.1007/s13312-020-1731-z
9. Ma LG, Chen QH, Wang YY, et al. Spatial pattern and variations in the prevalence of congenital heart disease in children aged 4–18 years in the Qinghai-Tibetan Plateau. *Sci Total Environ*. 2018;627:158–165. doi:10.1016/j.scitotenv.2018.01.194
10. Basson CT, Bachinsky DR, Lin RC, et al. Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat Genet*. 1997;15(1):30–35. doi:10.1038/ng0197-30
11. Ji L, Hou H, Zhu K, et al. NOTCH1 gene MicroRNA target variation and ventricular septal defect risk. *Omics*. 2019;23(1):28–35. doi:10.1089/omi.2018.0171
12. Zheng SQ, Chen HX, Liu XC, Yang Q, He GW. Genetic analysis of the CITED2 gene promoter in isolated and sporadic congenital ventricular septal defects. *J Cell Mol Med*. 2021;25(4):2254–2261. doi:10.1111/jcmm.16218
13. Peng J, Wang Q, Meng Z, et al. A loss-of-function mutation p.T256M in NDRG4 is implicated in the pathogenesis of pulmonary atresia with ventricular septal defect (PA/VSD) and tetralogy of Fallot (TOF). *FEBS Open Bio*. 2021;11(2):375–385. doi:10.1002/2211-5463.13044
14. Yang Q, Wu F, Mi Y, et al. Aberrant expression of miR-29b-3p influences heart development and cardiomyocyte proliferation by targeting NOTCH2. *Cell Prolif*. 2020;53(3):e12764. doi:10.1111/cpr.12764
15. Wang L, Song G, Liu M, et al. MicroRNA-375 overexpression influences P19 cell proliferation, apoptosis and differentiation through the Notch signaling pathway. *Int J Mol Med*. 2016;37(1):47–55. doi:10.3892/ijmm.2015.2399
16. Yang J, Bücker S, Jungblut B, et al. Inhibition of Notch2 by Numb/Numblike controls myocardial compaction in the heart. *Cardiovasc Res*. 2012;96(2):276–285. doi:10.1093/cvr/cvs250
17. Varadkar P, Kraman M, Despres D, Ma G, Lozier J, McCright B. Notch2 is required for the proliferation of cardiac neural crest-derived smooth muscle cells. *Dev Dynam*. 2008;237(4):1144–1152. doi:10.1002/dvdy.21502
18. High FA, Zhang M, Proweller A, et al. An essential role for Notch in neural crest during cardiovascular development and smooth muscle differentiation. *J Clin Invest*. 2007;117(2):353–363. doi:10.1172/JCI30070
19. McDaniell R, Warthen DM, Sanchez-Lara PA, et al. NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *Am J Hum Genet*. 2006;79(1):169–173. doi:10.1086/505332
20. van den Akker NM, Molin DG, Peters PP, et al. Tetralogy of fallot and alterations in vascular endothelial growth factor-A signaling and notch signaling in mouse embryos solely expressing the VEGF120 isoform. *Circ Res*. 2007;100(6):842–849. doi:10.1161/01.RES.0000261656.04773.39
21. Niessen K, Karsan A. Notch signaling in cardiac development. *Circ Res*. 2008;102(10):1169–1181. doi:10.1161/CIRCRESAHA.108.174318
22. Boutchueng-Djidjou M, Collard-Simard G, Fortier S, et al. The last enzyme of the de novo purine synthesis pathway 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase (ATIC) plays a central role in insulin signaling and the Golgi/endosomes protein network. *Mol Cell Proteom*. 2015;14(4):1079–1092. doi:10.1074/mcp.M114.047159

23. Li M, Jin C, Xu M, Zhou L, Li D, Yin Y. Bifunctional enzyme ATIC promotes propagation of hepatocellular carcinoma by regulating AMPK-mTOR-S6 K1 signaling. *Cell Commun Signal*. 2017;15(1):52. doi:10.1186/s12964-017-0208-8
24. Li R, Chen G, Dang Y, et al. Upregulation of ATIC in multiple myeloma tissues based on tissue microarray and gene microarrays. *Int J Lab Hematol*. 2021;43(3):409–417. doi:10.1111/ijlh.13397
25. Zhu J, Wang M, Hu D. Development of an autophagy-related gene prognostic signature in lung adenocarcinoma and lung squamous cell carcinoma. *PeerJ*. 2020;8:e8288. doi:10.7717/peerj.8288
26. van der Krogt JA, Bempt MV, Ferreiro JF, et al. Anaplastic lymphoma kinase-positive anaplastic large cell lymphoma with the variant RNF213-, ATIC- and TPM3-ALK fusions is characterized by copy number gain of the rearranged ALK gene. *Haematologica*. 2017;102(9):1605–1616. doi:10.3324/haematol.2016.146571
27. Christensen KE, Deng L, Leung KY, et al. A novel mouse model for genetic variation in 10-formyltetrahydrofolate synthetase exhibits disturbed purine synthesis with impacts on pregnancy and embryonic development. *Hum Mol Genet*. 2013;22(18):3705–3719. doi:10.1093/hmg/ddt223
28. Bumann M, Djafarzadeh S, Oberholzer AE, et al. Crystal structure of yeast Ypr118w, a methylthioribose-1-phosphate isomerase related to regulatory eIF2B subunits. *J Biol Chem*. 2004;279(35):37087–37094. doi:10.1074/jbc.M404458200
29. Sunker A, Alkuraya FS. Identification of MRI1, encoding translation initiation factor eIF-2B subunit alpha/beta/delta-like protein, as a candidate locus for infantile epilepsy with severe cystic degeneration of the brain. *Gene*. 2013;512(2):450–452. doi:10.1016/j.gene.2012.10.063
30. Gunawardhana LP, Baines KJ, Mattes J, Murphy VE, Simpson JL, Gibson PG. Differential DNA methylation profiles of infants exposed to maternal asthma during pregnancy. *Pediatr Pulmonol*. 2014;49(9):852–862. doi:10.1002/ppul.22930
31. Liu QR, López-Corcuera B, Mandiyan S, Nelson H, Nelson N. Molecular characterization of four pharmacologically distinct gamma-aminobutyric acid transporters in mouse brain [corrected]. *J Biol Chem*. 1993;268(3):2106–2112. doi:10.1016/S0021-9258(18)53968-5
32. Christiansen B, Meinild AK, Jensen AA, Braüner-Osborne H. Cloning and characterization of a functional human gamma-aminobutyric acid (GABA) transporter, human GAT-2. *J Biol Chem*. 2007;282(27):19331–19341. doi:10.1074/jbc.M702111200
33. Ramirez A, Heimbach A, Gründemann J, et al. Hereditary parkinsonism with dementia is caused by mutations in ATP13A2, encoding a lysosomal type 5 P-type ATPase. *Nat Genet*. 2006;38(10):1184–1191.
34. Bras J, Verloes A, Schneider SA, Mole SE, Guerreiro RJ. Mutation of the parkinsonism gene ATP13A2 causes neuronal ceroid-lipofuscinosis. *Hum Mol Genet*. 2012;21(12):2646–2650. doi:10.1093/hmg/dds089
35. Estrada-Cuzcano A, Martin S, Chamova T, et al. Loss-of-function mutations in the ATP13A2/PARK9 gene cause complicated hereditary spastic paraplegia (SPG78). *Brain*. 2017;140(2):287–305. doi:10.1093/brain/aww307
36. van Veen S, Martin S, Van den Haute C, et al. ATP13A2 deficiency disrupts lysosomal polyamine export. *Nature*. 2020;578(7795):419–424. doi:10.1038/s41586-020-1968-7
37. Grünewald A, Arns B, Seibler P, et al. ATP13A2 mutations impair mitochondrial function in fibroblasts from patients with Kufor-Rakeb syndrome. *Neurobiol Aging*. 2012;33(8):1843.e1841–1847. doi:10.1016/j.neurobiolaging.2011.12.035
38. Xu Q, Guo H, Zhang X, et al. Hypoxia regulation of ATP13A2 (PARK9) gene transcription. *J Neurochem*. 2012;122(2):251–259. doi:10.1111/j.1471-4159.2012.07676.x