CLINICAL RESEARCH

e-ISSN 1643-3750
© Med Sci Monit, 2018; 24: 1340-1358
DOI: 10.12659/MSM.908923

Received: 2018.01.11
Accepted: 2018.02.08
Published: 2018.03.05

Novel Genetic Variants of Sporadic Atrial Septal Defect (ASD) in a Chinese Population Identified by Whole-Exome Sequencing (WES)

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1,2,3 **Yong Liu***
A 1,2 **Yu Cao***
G 1,3 **Yaxiong Li***
E 4 **Dongyun Lei***
E 1,3 **Lin Li**
G 1,3 **Zong Liu Hou**
F 1,3 **Shen Han**
G 1 **Mingyao Meng**
C 1 **Jianlin Shi**
G 1,3 **Yayong Zhang**
B 1,3 **Yi Wang**
B 1 **Zhaoyi Niu**
G 1 **Yanhua Xie**
B 1 **Benshan Xiao**
B 1 **Yuanfei Wang**
C 1 **Xiao Li**
F 1 **Lirong Yang**
ACDE 1,3 **Wenju Wang**
AG 2 **Lihong Jiang**

1 Department of Cardiovascular Surgery, Yan'an Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, P.R. China
2 Department of Cardiovascular Surgery, The First Peoples' Hospital of Yunnan Province, Kunming, Yunnan, P.R. China
3 Key Laboratory of Cardiovascular Disease of Yunnan Province, Kunming, Yunnan, P.R. China.
4 The First Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, P.R. China

* Joint first authors; Yong Liu, Yu Cao, Yaxiong Li, Dongyun Lei

Corresponding Authors: Lihong Jiang, e-mail: jianglihong_yayy@163.com, Wenju Wang, e-mail: wangwenju_vip@hotmail.com

Source of support: This study was supported by grants from National Natural Science Foundation of China (No. 81560060, 31160230, 81760059), and the Yun Ling Scholarship from the Yunnan/Kunming Key Laboratory of Cardiovascular Surgery

Background: Recently, mutations in several genes have been described to be associated with sporadic ASD, but some genetic variants remain to be identified. The aim of this study was to use whole-exome sequencing (WES) combined with bioinformatics analysis to identify novel genetic variants in cases of sporadic congenital ASD, followed by validation by Sanger sequencing.

Material/Methods: Five Han patients with secundum ASD were recruited, and their tissue samples were analyzed by WES, followed by verification by Sanger sequencing of tissue and blood samples. Further evaluation using blood samples included 452 additional patients with sporadic secundum ASD (212 male and 240 female patients) and 519 healthy subjects (252 male and 267 female subjects) for further verification by a multiplexed MassARRAY system. Bioinformatic analyses were performed to identify novel genetic variants associated with sporadic ASD.

Results: From five patients with sporadic ASD, a total of 181,762 genomic variants in 33 exon *loci*, validated by Sanger sequencing, were selected and underwent MassARRAY analysis in 452 patients with ASD and 519 healthy subjects. Three *loci* with high mutation frequencies, the 138665410 *FOXL2* gene variant, the 23862952 *MYH6* gene variant, and the 71098693 *HYDIN* gene variant were found to be significantly associated with sporadic ASD ($P<0.05$); variants in *FOXL2* and *MYH6* were found in patients with isolated, sporadic ASD ($P<5\times10^{-4}$).

Conclusions: This was the first study that demonstrated variants in *FOXL2* and *HYDIN* associated with sporadic ASD, and supported the use of WES and bioinformatics analysis to identify disease-associated mutations.

MeSH Keywords: **Genetic Variation • Heart Defects, Congenital • Heart Septal Defects, Atrial**

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/908923

3625 8 9 29

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Background

Atrial septal defect (ASD) is the most common subtype of congenital heart disease (CHD), and its prevalence can reach between 8.93–10.6 per 1,000 live births in China [1,2]. ASD can lead to several clinical complications, including infective endocarditis, chronic heart failure, and repeated lung infection, which can severely affect the physical and psychological health of affected patients [3]. The process of cardiac development is regulated by several genes and is precisely controlled; any disruption in cardiac development can cause congenital cardiac defects. Gene mutations have been shown to play important roles in the etiology and pathogenesis of ASD, including mutations in *NKX2-5, GATA4, MYH6,* and *TBX5* [4,5]. However, some genetic variants associated with sporadic ASD remain to be identified.

Whole-exome sequencing (WES) is a powerful and efficient tool to obtain sequencing information on the whole exome with high resolution and low cost [6]. Due to the limitations in current knowledge of the genomic noncoding regions, many recent studies have only focussed on the identification of pathogenic mutations in protein coding regions by using WES [7]. The rapid development in the techniques used in WES has shed light on the complex mechanisms involved in several forms of CHD, including patent ductus arteriosus (PDA), familial ASD, and ventricular septal defect (VSD) [8]. However, few studies have studied the mutations associated with the etiology and pathogenesis of sporadic ASD. Therefore, there is a need to extend studies on the spectrum of ASD-related mutations using WES and to provide a foundation for further functional analysis, which may further clarify the gene associations, the pathogenesis, and possibly lead to new diagnostic markers for sporadic ASD.

The aim of this study was to use WES combined with bioinformatics analysis to identify novel gene variants in cases of sporadic congenital ASD, followed by validation by Sanger sequencing and a multiplexed MassARRAY system.

Material and Methods

Study participants and collection of tissue samples

All participants in this study were from the Chinese population and were enrolled by the Department of Cardiovascular Surgery of Yan'an Affiliated Hospital of Kunming Medical University. Five Han patients with secundum atrial septal defect (ASD) were recruited, and their tissue samples were collected for whole-exome sequencing (WES). To confirm the WES findings, verification was conducted by Sanger sequencing with tissue samples and blood samples from the same five patients with sporadic ASD. For further evaluation, 452 additional patients with sporadic secundum ASD (212 male and 240 female patients) and 519 healthy subjects (252 male and 267 female subjects) were recruited, and the blood samples were obtained for further verification using a multiplexed MassARRAY system.

The study was approved by the Medical Ethics Committee of Kunming Medical University. Informed consent was signed by each study participant or their legal guardians to participate in the study. The diagnosis of the ASD was based on echocardiography and during routine surgery for ASD surgical repair. Details of the clinical characteristics of the study participants are shown in Supplementary Table 1.

The clinical data of the five patients with isolated ASD

Case 1: A female patient aged 6 months was found to have a grade 3/6 systolic murmur in the second left parasternal space and fixed splitting of the second heart sound. Left-to-right shunting was detected by echocardiogram and the diameter of the defect in the atrial septum was 22 mm.

Case 2: A male patient aged 13 months was found to have a grade 2/6 systolic murmur in the second left parasternal space. Left-to-right shunting was detected by echocardiogram and the diameter of the defect in the atrial septum was 10 mm.

Case 3: A female patient aged 19 months was found to have a grade 2/6 systolic murmur in the second left parasternal space and fixed splitting of the second heart sound. Left-to-right shunting was detected by echocardiogram and the diameter of the defect in the atrial septum was 16 mm.

Case 4: A female patient aged 23 months was found to have a grade 2/6 systolic murmur in the second left parasternal space and fixed splitting of the second heart sound. Left-to-right shunting was detected by echocardiogram and the diameter of the defect in the atrial septum was 12 mm.

Case 5: A male patient aged 17 months was found to have a grade 2/6 systolic murmur in the second left parasternal space and fixed splitting of the second heart sound. Left-to-right shunting was detected by echocardiogram and the diameter of the defect in the atrial septum was 11 mm.

The secundum ASD of the five patients with isolated ASD was further confirmed during the ASD repair surgery, performed through a median sternotomy, under cardiopulmonary bypass. Atrial septal tissue samples from the five patients with ASD were collected from the rims of the defect in the atrial septum. The collected tissue samples were stored at –80°C. Serum was extracted from blood and stored at –80°C.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

None of the five patients had a family history of congenital heart disease (CHD), Down's syndrome, or Marfan's syndrome. Patients with other common developmental defects or chromosomal abnormalities were excluded from this study. The 519 healthy subjects were recruited from the Department of Health Examination Centers of Yan'an Affiliated Hospital of Kunming Medical University. Echocardiograms, dynamic electrocardiograms, treadmill exercise tests, measurements of blood pressure and blood lipids were performed to exclude cardiovascular disease, including coronary artery disease, hypertension, and arrhythmias.

Whole-exome sequencing

Genomic DNA of tissue samples from five patients with ASD was extracted using Thermo DNA and the Lab-Serv Cell and Tissue DNA Extraction Kit (Thermo Scientific). The purity and quality of DNA from the samples were evaluated using an ultramicro-spectrophotometer (SpectraMax QuickDrop), and the optical density (OD) value of DNA was identified as between 1.8–2.0. Then the DNA was aliquoted and preserved in 0.5mL Eppendorf tubes at –80°C for WES. The enrichment of exon was performed by Agilent Sure Select Human All Exon V5 Kit (Agilent, USA) from 1.0 ug genomic DNA according to the manufacturer's protocol. First, the genomic DNA was broken randomly into 150–200bp fragments. Then, the DNA libraries were prepared by the addition of "A" bases to the 3' end of the DNA fragment. Finally, the DNA libraries were assessed for quality control and sequenced by Illumina Hiseq 2500 Sequencer (Supplementary Figure 1).

Mapping to reference sequences

All single nucleotide variants (SNVs) of each sample were obtained by comparing the valid sequencing data with the human reference genome (UCSC Genome Browser hg19) using Burrows-Wheeler Aligner (BWA) software [9] to gain the primary mapping results. Then, the aligned data were sorted by SAMtools [10] to select the best mapping positions, and the duplicated reads were marked by Picard (*http://sourceforge.net/projects/picard/*) so that they could be used in the next analysis.

Annotation, data filtering, and gene ontology analysis

Functional annotation was conducted to find the genetic variation associated with ASD. First, all variants were annotated using the ANNOVAR software tool [11]. Then, the normal population variant databases, including 1,000 Genomes Project (version 2012), the Single Nucleotide Polymorphism database (DbSNP) (version 138), and the National Heart, Lung, and Blood Institute (NHLBI) database, were performed to exclude the common variations occurring with no more than 1% minor allele frequency (MAF) and variants not related to congenital heart disease (CHD). Then, the rare variants obtained in the previous step were further analyzed using the Venn analysis, gene ontology (GO) analysis, and literature review and protein database (*http://www.uniprot.org/* and *http://www.genecards.org/*) to classify the variants associated with cardiac development or ASD. GO analysis was applied to analyze the main function of SNVs according to the GO which was the key functional classification of the National Center for Biotechnology Information (NCBI). The genes of the Gene Ontology (GO) term enrichment analysis included the following: GO.0048739_cardiac muscle fiber development, GO.0003143_embryonic heart tube morphogenesis, GO.0003300_cardiac muscle hypertrophy, GO.0001539_cilium or flagellum-dependent cell motility, GO.0055008_cardiac muscle tissue morphogenesis, GO.0060038_cardiac muscle cell proliferation, GO.0060956_endocardial cell differentiation, GO.0007512_adult heart development, GO.0003007_heart morphogenesis and GO.0030509_BMP signaling pathway.

Candidate susceptible variants and gene selection

The SIFT [12], PolyPhen-2 [13] and MutationTaster [14] were performed to predict whether the substitution of amino acid affected the function of the protein. Based on the degree of CHD, the mutation was selected as candidate susceptible variant if one of the three software showed it was pathological. Subsequently, the expression level of the corresponding genes of the candidate variants was selected via genecards (*http://www.genecards.org*).

Variant validation and statistical analysis

Mutation validation was initially conducted by Sanger sequencing. Sanger sequencing was performed using genomic DNA from both tissue samples and peripheral blood of the same five patients with ASD recruited for WES. The mutations in both tissue samples and blood samples were selected as positive variants. Additionally, the positive variants were consistent with corresponding data from WES and were further tested using the multiplexed MassARRAY analysis in large-scale samples. The primers involved in MassARRAY were designed using AssayDesigner 3.1. The sequences of the primers are listed in Supplementary Table 2.

The genomic DNA was amplified under the following conditions: initial denaturation, 94°C for 15 min, then 45 cycles of 94°C for 20 s, 56°C for 30 s and 72°C for 60 s, followed by 72°C for 3 min. The final products were preserved at 4°C for Shrimp alkaline phosphatase (SAP) purification reaction to remove unincorporated dinucleotide triphosphates (dNTPs). The products were purified under the following conditions: 37°C for 40 min, 85°C for 5 min, the products were preserved at 4°C for extension polymerase chain reaction (PCR).

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

The conditions of extension PCR were listed as follows: 94°C for 30 s, 94°C for 5 s, then 40 cycles of 52°C for 5 s, 5 cycles of 80°C for 5 s, followed by 72°C for 3 min. The final products were preserved at 4°C for further analysis. The products and water were robotically dispensed into 384 sample plate and mass spectra were collected by the multiplexed MassARRAY compact matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) analyzer. The MassARRAY analysis was performed by EpiTYPER (Version: 4.0). The frequencies of the mutations were calculated by the χ^2 test or Fisher exact test using SPSS version 22.0 (IBM, USA). P<0.05 represented statistical significance.

Results

Clinical characteristics of five patients with sporadic atrial septal defect (ASD)

Five patients with secundum atrial septal defect (ASD) diagnosed by echocardiogram were recruited into the study. The diameter of the defects in the atrial septum ranged from 10 mm to 22 mm (Table 1). Except for ASD, other cardiac structural abnormalities and congenital anomalies were not found in these five patients. ASD was not diagnosed in other family members, which ensured that all patients included in this study had sporadic ASD.

Overview of whole-exome sequencing (WES) data

The total raw reads of ASD samples ranged from 17.57–28.46 million. More than 95% of the sequenced bases showed a quality score of ≥Q20. The raw sequencing data yielded an average of 5000 Mb of effective data and the average depth of the target areas was over 60 × coverage. The quality and depth of target areas are listed in Supplementary Figure 2 and Supplementary Table 3.

In total, 181,762 genomic variants in five samples from patients with isolated ASD were identified. There were an average of over 20,000 heterozygous variants and 15,000 homozygous variants in the samples (Supplementary Table 4). The circos map is shown to demonstrate the distributions of the variants on the chromosomes (Supplementary Figure 3).

Identification of 33 ASD-associated gene variants by bioinformatics analysis

The advanced bioinformatic analysis was performed to identify the single nucleotide variants (SNVs) associated with ASD (Figure 1). The 181,762 SNVs from the five patients were filtered by 1,000 Genome, the Single Nucleotide Polymorphism database (DbSNP), and the National Heart, Lung, and Blood Institute (NHLBI) database. After filtering, there were 713 rare variants, and the distribution of the rare variants on chromosomes was presented by circus map (Supplementary Figure 3, Supplementary Table 4). Furthermore, through performing Venn analysis, GO analysis and literature review, 13, 21 and 20 pathogenic variants perhaps involved in heart development were obtained (Supplementary Figure 4, Supplementary Tables 5–7). Also, the genes that overlapped and or were only minimally expressed in the heart were excluded from the present study. Finally, 33 variants in 25 genes were chosen for further analysis (Table 1).

Validation and exploration of novel ASD variants

The 33 ASD-associated variants selected in the above step were further validated by Sanger sequencing in samples from the atrial septum and the peripheral blood (Supplementary Figure 5). The findings showed that the mutations of 33 *loci* were consistent with corresponding data from WES, verifying the accuracy and reliability of WES findings. The consistency of mutations in tissue samples and blood samples not only confirmed our results but also excluded somatic mutations. Then, the multiplexed MassARRAY analysis was performed to verify the 33 variants in additional 452 samples from patients with ASD, and 519 samples from healthy subjects.

The results showed that nine variants were positive mutations, which were respectively located in nine genes. The variants were: 138665410 (NM_023067_c.C155G) in *FOXL2*, 23862952 (NM_002471: c.G2851T) in *MYH6*, 71098693 (NM_001198542: c.A2207C) in *HYDIN*, 88600890 (NM_153813: c.G2524A) in *ZFPM*, 39913231 (NM_001123384: c.T4728G) in *BCOR*, 156186376 (NM_001128209: c.A845G) in *SGCD*, 61708404 (NM_003400: c.G2985C) in *XPO1*, 65652097 (NM_004214: c.G850A) in *FIBP*, 179480499 (NM_003319: c.C21134G) in *TTN* (Supplementary Table 2).

Also, three mutations with a high mutation frequency were found to be significantly associated with ASD (P<0.05), including 138665410 in the *FOXL2* gene (26.11%), 23862952 in the *MYH6* gene (2.43%), 71098693 in the *HYDIN* gene (3.10%) (Figure 2, Supplementary Table 2). Among these loci, variants in the *FOXL2* gene and the *MYH6* gene were only found in patients with ASD ($P<5\times10^{-4}$), and variants in the *FOXL2* gene and the *HYDIN* gene were first identified in isolated ASD. No significant difference in the other six variants were found between patients with ASD and healthy subjects (P>0.05) (Figure 3, Supplementary Table 2). The location of the three variants was found to be in located in a region that is highly conserved among species (Supplementary Figure 6).

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1. 33 variants in 25 ASD related genes from WES data.

Gene	The information of variants	The position of mutation site on a chromosome	The number of sample	Harmful prediction by SIFT/ Ployphen-2/ Mutation Taster	Bioinformatic methods		
					Venn analysis	GO analysis	Literature review
TTN	NM_003319: exon178: c.C70564T: p.R23522C	179406045	ASD-5	D/D/D	+		
TTN	NM_003319: exon154: c.G47144A: p.R15715Q	179436520	ASD-4	D/D/D	+		
TTN	NM_003319: exon86: c.C21134G: p.T7045S	179480499	ASD-2	D/B/N	+		
TTN	NM_003319: exon186: c.C76340T: p.P25447L	179397807	ASD-3	D/D/D	+		
TTN	NM_003319: exon154: c.G48662T: p.G16221V	179435002	ASD-2	D/D/D	+	+	
HYDIN	NM_001270974: exon47: c.G7930A: p.E2644K	70952188	ASD-1	T/B/D	+	+	
HYDIN	NM_001198542: exon16: c.A2207C: p.H736P	71098693	ASD-2, ASD-3	T/P/D	+	+	
IGSF3	NM_001542: exon3: c.A619G: p.S207G	117156600	ASD-2, ASD-5	T/B/D	+		
IGSF3	NM_001007237: exon10: c.G3300C: p.E1100D	117122048	ASD-4	T/B/D	+		
ZFPM	NM_153813: exon10: c.C2359T: p.P787S	88600725	ASD-2	T/B/D	+	+	
ZFPM	NM_153813: exon10: c.G2524A: p.A842T	88600890	ASD-5	D/P/N	+	+	
MYH6	NM_002471: exon11: c.G985T: p.E329X	23871923	ASD-1	D/./A	+		
MYH6	NM_002471: exon22: c.G2851T: p.E951X	23862952	ASD-3	D/./A	+	+	
FMO5	NM_001144830: exon4: c.T572A: p.I191N	146684019	ASD-1	D/D/D			+
NSD1	NM_022455: exon5: c.A2608G: p.R870G	176638008	ASD-1	./D/N			+
OBSCN	NM_001098623: exon27: c.A7301G: p.H2434R	228467050	ASD-1	T/D/D			+
FOXL2	NM_023067: exon1: c.C155G: p.A52G	138665410	ASD-1	T/P/D			+
NUP188	NM_015354: exon28: c.C3047T: p.P1016L	131756681	ASD-1	D/D/D			+
SOX17	NM_022454: c.A595T	55371905	ASD-1	T/B/N			+
XPO1	NM_003400: exon24: c.G2985C: p.K995N	61708404	ASD-1	T/P/D		+	+
KIAA0196	NM_014846: exon18: c.A2186C: p.N729T	126062819	ASD-3	T/D/D			+
EP300	NM_001429: exon13: c.G2261A: p.R754H	41545061	ASD-3	T/D/D			+
ZNF638	NM_001014972: exon2: c.A254G: p.E85G	71576338	ASD-3	D/D/D			+
BBS1	NM_024649: exon11: c.G1067A: p.R356H	66291310	ASD-5	T/B/D			+
ARL6	NM_001278293: exon5: c.G283T: p.D95Y	97503827	ASD-5	D/P/D		+	+
ACVR1	NM_001105: exon4: c.A275G: p.E92G	158636905	ASD-5	T/P/D			+
VEGFA	NM_001025366: exon6: c.C997T: p.R333X	43748503	ASD-5	T/./D			+
USP44	NM_001042403: exon6: c.A2134C: p.S712R	95911935	ASD-3	D/B/D			+
FGB	NM_001184741: exon3: c.A263C: p.Y88S	155487774	ASD-3	T/B/N			+
BCOR	NM_001123384: exon13: c.T4728G: p.D1576E	39913231	ASD-4	T/B/N			+
SGCD	NM_001128209: exon8: c.A845G: p.Q282R	156186376	ASD-4	T/D/D		+	+
BBS9	NM_014451: exon19: c.A2389G: p.M797V	33573776	ASD-1	T/B/D			+
FIBP	NM_004214: exon8: c.G850A: p.A284T	65652097	ASD-2	T/B/D		+	+

According to SIFT, related gene was noted on tolerated (T, score >0.5) or deleterious (D, score <0.5). According to Ployphen-2, related gene was noted on probably damaging (D, Polyphen-2 ≥0.909), possibly damaging (P, 0.447 ≤Polyphen-2 <0.909), and benign (B, Polyphen-2 <0.447). According to Mutation Taster, related genes were noted on disease-causing automatic (A), disease-causing (D), polymorphism (N), and polymorphism automatic (P) (*http://www.mutationtaster.org/*).

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

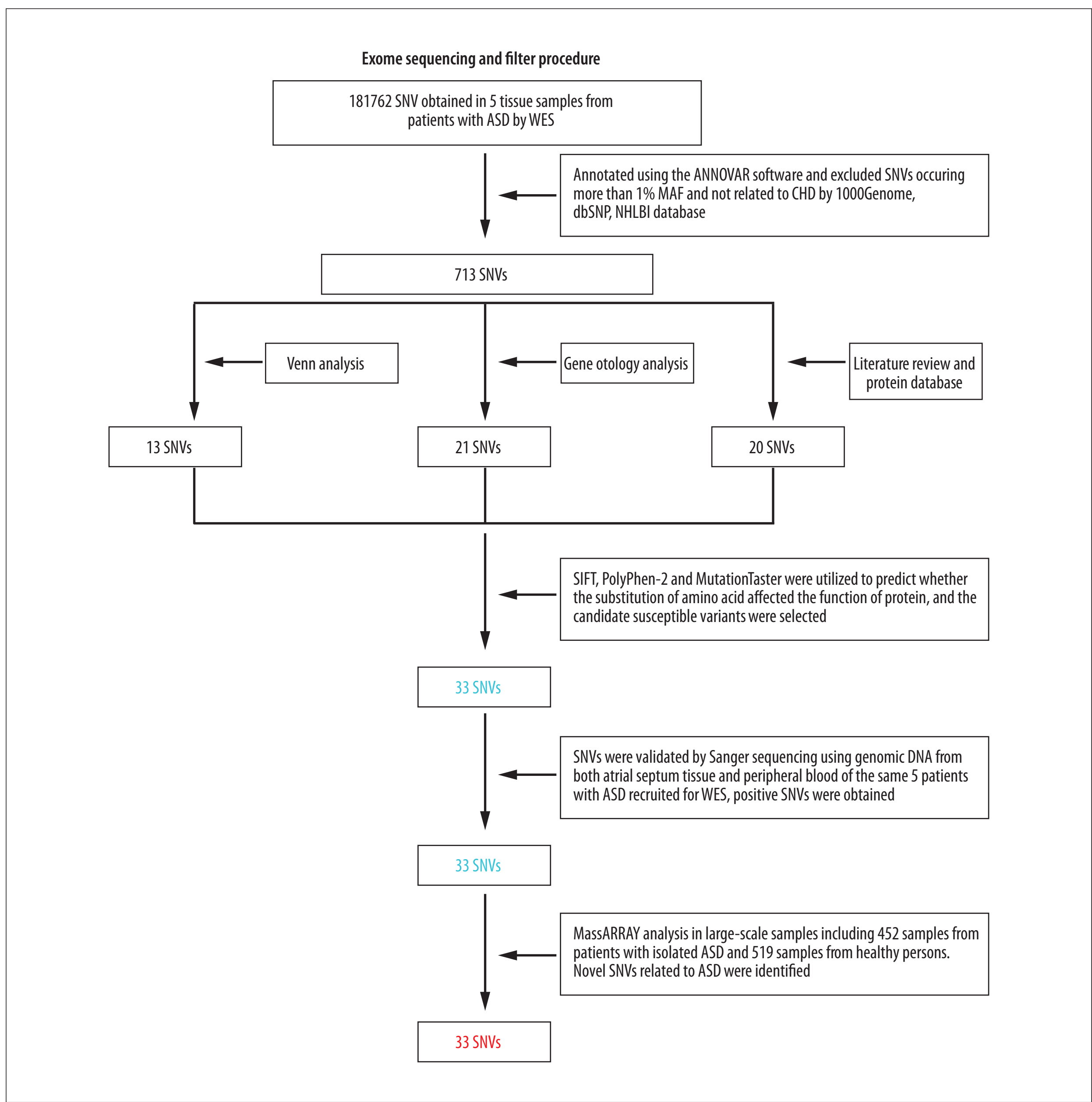


Figure 1. Schematic representation of the study design and protocol. First, all variants were annotated using the ANNOVAR bioinformatics software. The normal population variant databases, including 1,000 Genomes Project, the Single Nucleotide Polymorphism database (DbSNP), and the National Heart, Lung, and Blood Institute (NHLBI) databases were used to exclude the common variants occurring with more than 1% minor allele frequency (MAF). Second, the rare variants obtained in the previous step were further analyzed using Venn analysis, gene ontology (GO) analysis, and literature review. Third, three function predictor scores, SIFT [12], Polymorphism Phenotyping v2 (PolyPhen-2) [13], and MutationTaster [14] were used to predict whether the substitution of amino acids affected the function of the protein, and 33 single nucleotide variants (SNVs) were selected and validated as positive mutations by Sanger sequencing. Finally, a multiplexed MassARRAY system was performed to verify the 33 variants and three mutations with high mutation frequency, which were found to be strongly associated with atrial septal defect (ASD) ($P<0.05$).

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

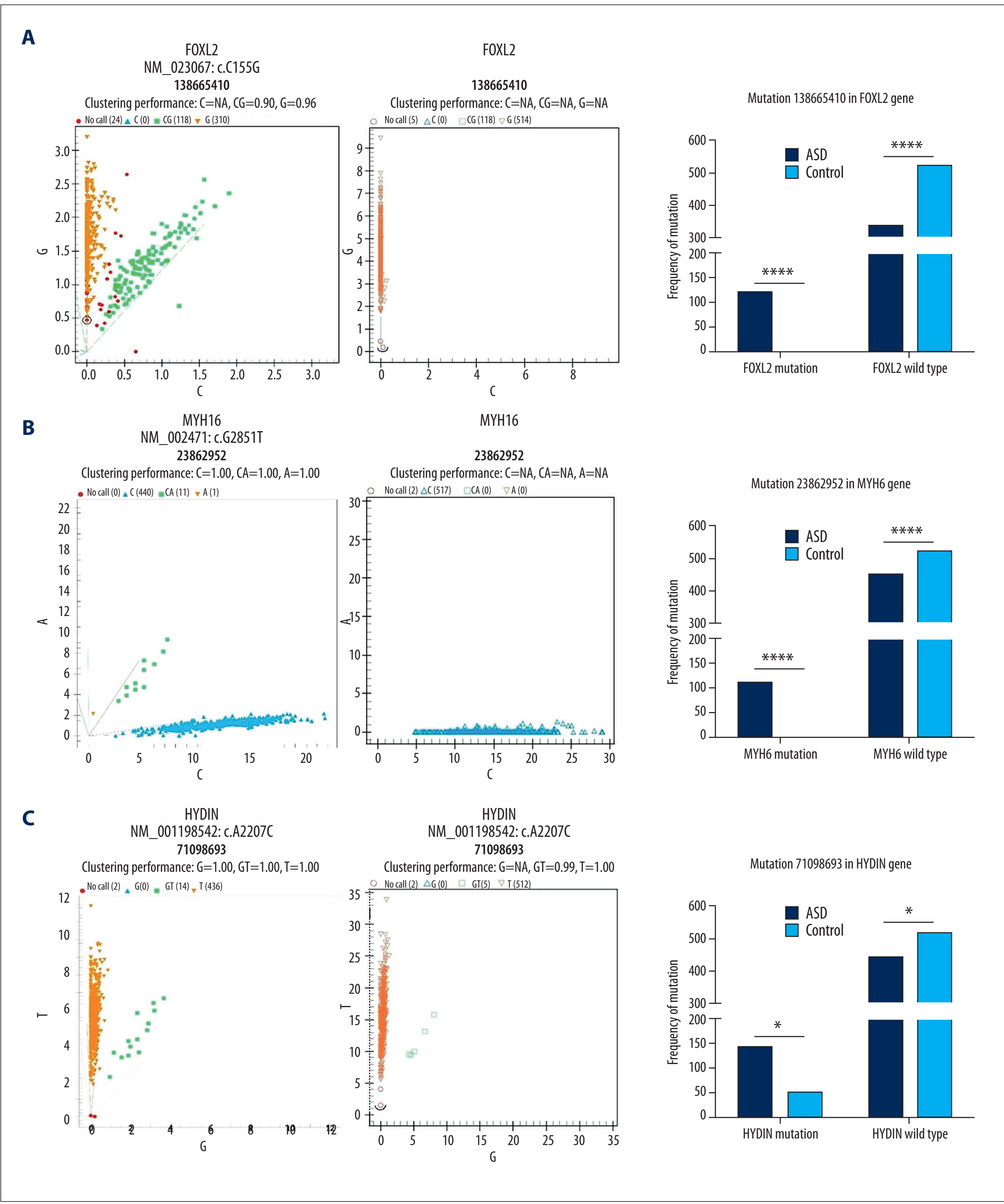
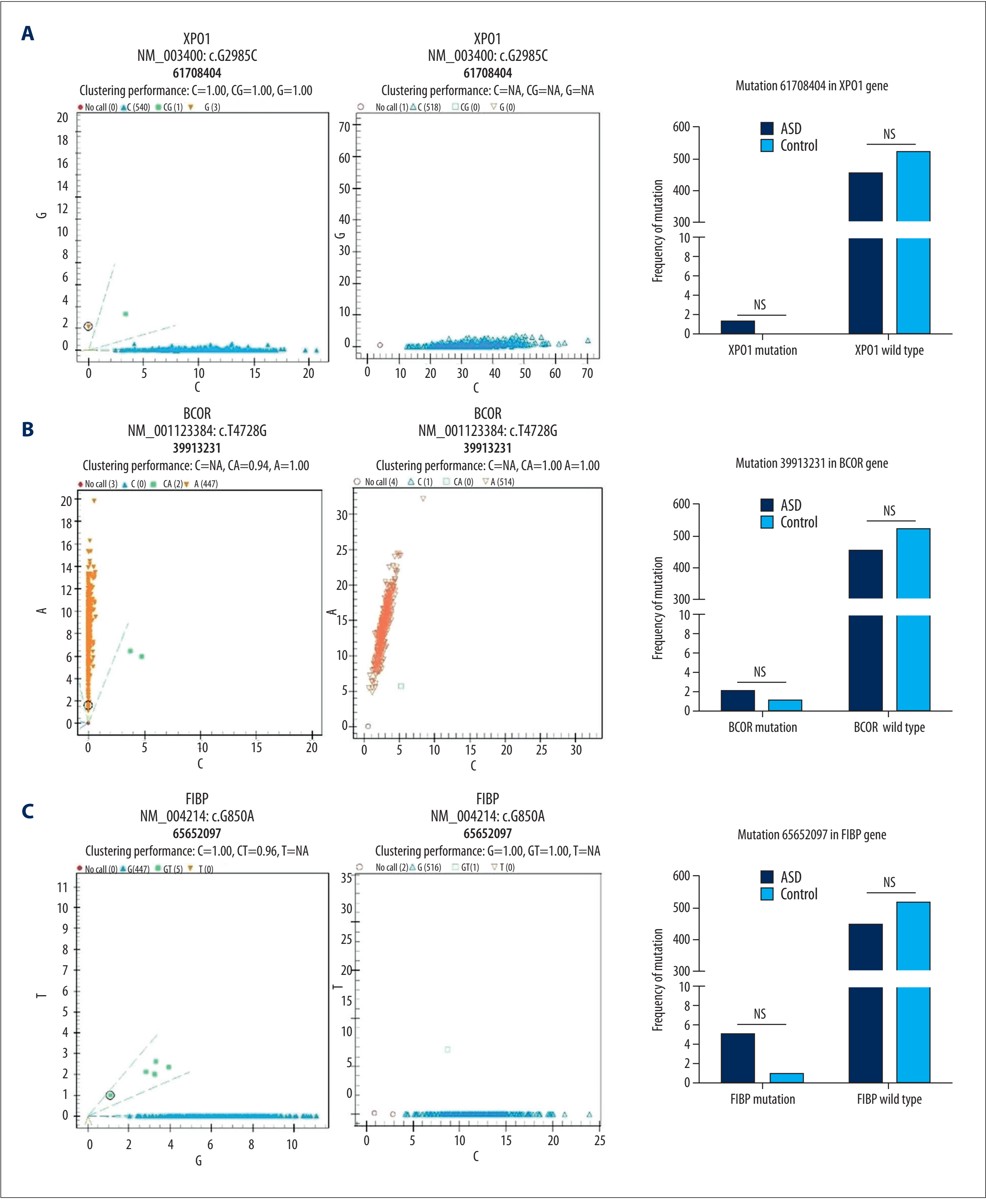


Figure 2. Selected gene variants in patients with atrial septal defect (ASD) were validated by MassARRAY analysis. (**A**) The mutation frequency of the 138665410 variant in the *FOXL2* gene. (**B**) The mutation frequency of the 23862952 variant in the *MYH6* gene. (**C**) The mutation frequency of the 71098693 variant in the *HYDIN* gene. Shown in patients with ASD (**left**) and healthy subjects (**right**) in a scatter chart. These *loci* with high mutation frequency were found to be strongly associated with ASD ($P<0.05$). Among these *loci*, variants in *FOXL2* and *MYH6* were only found in patients with ASD ($P<5\times10^{-4}$). * $P<0.05$, **** $P<0.0005$.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

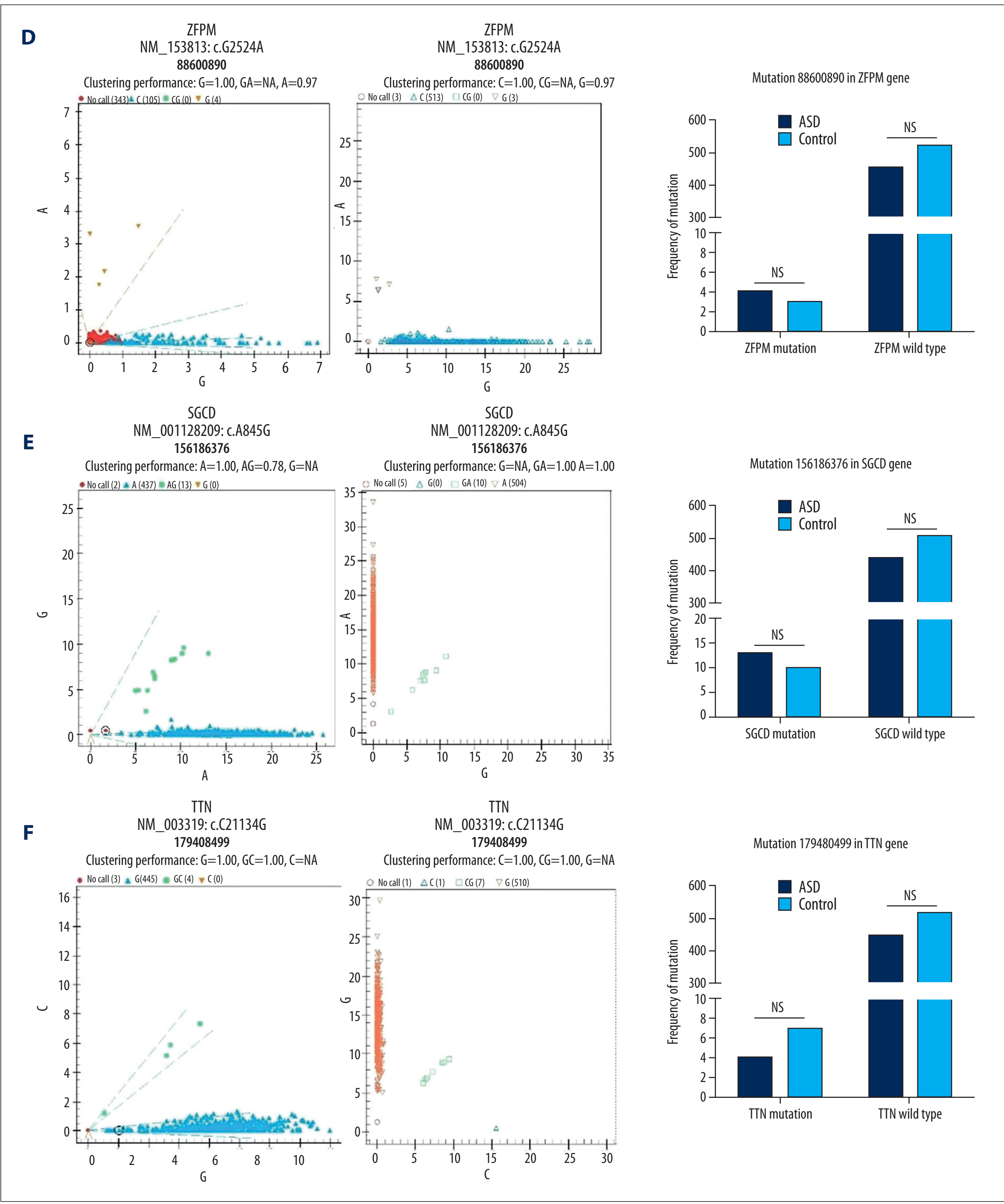


Figure 3. Further selected gene variants in patients with atrial septal defect (ASD) were validated by MassARRAY analysis. (**A**) The mutation frequency of the 61708404 variant in the *XPO1* gene. (**B**) The mutation frequency of the 39913231 variant in the *BCOR* gene. (**C**) The mutation frequency of the 65652097 variant in the *FIBP* gene. (**D**) The mutation frequency of the 88600890 variant in the *ZFPM* gene. (**E**) The mutation frequency of the 156186376 variant in the *SGCD* gene. (**F**) The mutation frequency of the 179480499 variant in the *TTN* gene. Shown in patients with atrial septal defect (ASD) (**left**) and healthy subjects (**right**) in a scatter chart. There was no significant difference in these six *loci* between patients with ASD and healthy subjects (P>0.05). (NS, P>0.05).

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Discussion

The etiology of congenital heart disease (CHD), including atrial septal defect (ASD), is complex and associated with both environmental and genetic factors. Epidemiological data has shown that environmental factors, including viral infection during pregnancy, can increase the risk of CHD, but that the genetic causes are mainly associated with CHD [15]. The role of genomic variants in CHD, with rare mutations in cardiac genes, are more likely to result in CHD, including in sporadic ASD [4,5]. Therefore, the aim of this study was to use whole-exome sequencing (WES) combined with bioinformatics analysis to identify novel genetic variants in cases of sporadic congenital ASD, followed by validation by Sanger sequencing.

In the present study, 181,762 genomic variants were identified by the WES approach. To the best of our knowledge, this is the first study to identify exon mutations associated with ASD by WES in sporadic ASD. Through verification by Sanger sequencing and MassARRAY, three *loci* were identified, 138665410 in the *FOXL2* gene, 71098693 in the *HYDIN* gene, and 23862952 in the *MYH6* gene, with high mutation frequency, likely to be associated with ASD. Also, two variants, 138665410 (NM_023067_c.C155G) in the *FOXL2* gene and 71098693 (NM_001198542: c.A2207C) in the *HYDIN* gene were identified for the first time in ASD. The following variants, 138665410 (NM_023067_c.C155G) in the *FOXL2* gene and 23862952 (NM_002471: c.G2851T) in the *MYH6* gene were only found in ASD in this study. These findings probably represent disease-associated mutations for sporadic ASD.

The *FOXL2* gene acts as an important transcription factor belonging to the winged helix/forkhead transcription factor family, including Foxa2, Foxc1/c2 and Foxos [16]. These family members have been identified as crucial components of the signaling pathways for controlling cardiogenesis and embryonic cardiac development [16]. Previous studies have shown that FOXL2 protein was expressed in several tissues, including the heart [16–19]. However, the role of the *FOXL2* gene has not been well studied in ASD. In the present study, we found that the mutation frequency of 138665410 variant (NM_023067_c.C155G) in *FOXL2* gene was 26.11%, the G>C mutation was found in 118 cases with ASD but not in healthy subjects. The mutation frequency was significantly different between patients with ASD and healthy subjects ($P<5\times10^{-4}$). The variant in *FOXL2* that resulted in a replacement of alanine by glycine might affect atrial septum morphogenesis. The mechanism of the involvement of *FOXL2* in ASD remains to be elucidated.

Previously published studies have shown that the activation of aberrant Notch signaling was induced the down-regulation of *FOXL2* expression during the development of the eyelid levator muscle [20]. Also, the Notch signaling pathway has been shown to participate in the development of the heart, while abnormal Notch signaling might cause aberrant atrioventricular canal, proximal outflow tract and coronary system development [21,22]. Therefore, it may be speculated that the function of *FOXL2* during atrioventricular canal and proximal outflow tract formation is Notch signaling-dependent, but the specific mechanisms of *FOXL2* mutations participating in the ASD development require further investigation.

The *MYH6* gene encodes the alpha heavy chain subunit of cardiac myosin, which is a fast ATP-ase that is primarily expressed in atrial tissue, and plays a crucial role in muscle contraction and sarcomeric structure, with studies showing that aberrant expression of *MYH6* ablated the atrial septal formation, and mutation of *MYH6* contributed to ASD development [23–25]. In the present study, the 23862952 variant of the *MYH6* gene (NM_002471: c.G2851T) was found in patients with ASD but not in healthy subjects ($P<5\times10^{-4}$). The mutation was localized in the 22nd exon of *MYH6* gene, which resulted in a shift of the open reading frame (ORF), termination of protein synthesis, and loss of 988 amino acids by the premature stop codon. Due to this variant, a highly conserved amino acid was altered in the myosin tail domain, which led to the loss of about 95% of tail domains of *MYH6*. This mutation in the *MYH6* gene might affect the development of the atrial septum, explaining its involvement in the pathogenesis of ASD.

The *HYDIN* gene is localized on the chromosome 16 (16q22.2) and encodes an axonal and cilial protein [26]. This gene has been predominantly found in the fetal heart and bronchial ciliated epithelia [27], and mutations in the *HYDIN* gene have been shown to impair ciliary motility in a mouse model [28]. A large-scale mouse mutagenesis screening study showed that primary cardiac cilia were required in formation and development of endocardial cushions during embryogenesis [29]. A high rate of ciliary dysfunction and mutations in cilia-related pathways have also been identified in mice with cardiac defects [29]. However, the role of the *HYDIN* gene has not previously been well studied in human ASD. However, in this study, a novel mutation, the 71098693 variant (NM_001198542: c.A2207C) in the *HYDIN* gene was identified in patients with ASD, and its mutation frequency was significantly different between patients with ASD compared with healthy subjects ($P<0.05$). This variant caused a replacement of alanine by cysteine with charge modification, which may result in the mutation in the *HYDIN* gene disabling ciliary motility that affects the formation of endocardial cushions in cardiac embryogenesis. However, the exact function of *HYDIN* during atrial septum morphogenesis remains unknown.

Genomic studies rely on precision, reliability, and reproducibility with reliable methods of identification and analysis. In the current study, a combination of WES and generation sequencing

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

was used to reduce the risk of false-positive results. In the selection of susceptible and reliable genetic variants related to ASD, the study design included five main steps. To exclude the interference of familial and syndromic ASD, sporadic cases of ASD were included. The use of WES yielded an average of 5000 Mb of data and the average depth of the target area was >60× coverage, with more than 95% of the sequenced bases shown to have a quality score of ≥Q20, reflecting the accuracy of sequencing and the quality of samples in the study. The process of gene mutation filtering was performed as rigorously as possible, using the 1,000 Genomes Project, the Single Nucleotide Polymorphism database (DbSNP), and the National Heart, Lung, and Blood Institute (NHLBI) databases to exclude the common variants occurring with more than 1% minor allele frequency (MAF). Through a series of analyses that included Venn analysis, gene ontology (GO) analysis, and literature review, and the use of three predictive bioinformatics software programs, including SIFT [12], PolyPhen-2 [13], and MutationTaster [14], it was possible to classify the variants associated with cardiac development or CHD and their potential impact on the function of proteins. To exclude false-positive variants, Sanger sequencing was performed, which showed that mutations from 33 *loci* were consistent with corresponding data from WES, which verified the accuracy and supported the reliability of the WES findings. Finally, the fifth important consideration in the study design was confirmation of the findings based on validation using large-scale clinical populations with ASD by using multiplexed MassARRAY analysis.

This study had several limitations. Because WES allowed the amplification of between 80–90% of all coding exons, some gene exons may have been missed in the process of sequencing. Also, WES cannot be performed to identify deep intronic mutations and is not an effective method to test for large genomic events, such as gene deletions and insertions.

Conclusions

The present study was the first study that demonstrated variants in the *FOXL2* and the *HYDIN* genes were associated with sporadic atrial septal defect (ASD), and supported the use of whole-exome sequencing (WES), Sanger sequencing, and bioinformatics analysis to identify disease-associated mutations. The results showed that mutations in the *FOXL2, MYH6* and *HYDIN* genes were associated with cases of ASD and that the presence of mutations in the *FOXL2*, *MYH6*, and *HYDIN* genes might contribute to the etiology of sporadic cases of ASD.

Conflict of Interest

None.

Supplementary data

Supplementary Table 1. Characteristics of 5 ASD patients.

Number of sample	Gender	Age	Diagnosis and diameter
ASD-1	Female	6 months	Secundum ASD, 22 mm
ASD-2	Male	13 months	Secundum ASD, 10 mm
ASD-3	Female	19 months	Secundum ASD, 16 mm
ASD-4	Female	23 months	Secundum ASD, 12 mm
ASD-5	Male	17 months	Secundum ASD, 11 mm

Supplementary Table 2. Sequences of the primers involved in Mass-Array.

Variants under test	Forward (5'-3')	Reverse (3'-5')	Product (bp)
23871923	AGCTTGTAGACGCCAGCTTT	CTGGGTTCACTCCTTGGTCC	433
33573776	ACAGCCAGCTGAACATACCC	GGGATGTCAGAGGGGAAACG	281
55371905	CGCAAGCAGGTGAA	CCAGCGTAGTCCGAGA	359
61708404	TATCCGAATAACATCTC	GAATGTCTTGAACCC	450
70952188	TGGGGACTGCAAAGGTTGTT	ATCGTCTCCTACCCGGTGAA	420

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Variants under test	Forward (5'-3')	Reverse (3'-5')	Product (bp)
131756681	ATGAAGTGACCACCAAA	GAGAAGACCACAGAGCC	390
138665410	GGTCCAGCGTCCAGTAGTTGC	GCACAGTCAAGGAGCCAGAAGG	330
146684019	TCTATTAGCACCAATCACC	GAAGTTCAAAGGGCAGTA	315
176638008	AAAACCAGGGATTCAAGT	CAGTAAGCCAGGTAGGGA	465
228467050	ATGGTTTGCCCTACACTCCT	TGTCCTCTTGTGGCTCTGC	89
65652097	GACCAGGTTGGGTGTCAGAG	GGGACTTGGTTGATGGAGGG	248
88600725	CCACGAGACCTACACCG	CCAGGCGGAAATGCT	545
117156600	GACAAGAGGCTCAGAGGG	CTGTGGCTTTGTGACTGC	479
179435002	AATGCCTTCTTTATCCCG	AACCAGCCGCTTAGTTTG	310
179480499	CCCATAACTTCAACTCTA	ACACCTCATTTGCCATCT	461
23862952	GCTGCTCCAGCTTGACCTTA	TCTCCTCCCCTCCCCTAGAT	686
41545061	CCTCTTCAGCCCTCTTCACC	GAGACACTGGAGCTTGACCG	281
71098693	GAACTCACCTGAGGCTGGAC	GTATCTCGCTTGGGGCTCTG	224
71576338	ACTTCAAAGGCCACGAGCAC	ATTAGGTAATCGGCGCCCCA	526
95911935	ACAAGTTCATCATCCGAGCC	CAGGGTAACACTGCCCAATCT	631
126062819	TGAAATGGGATCAATAGG	CAAAGCAGTTGCTGGAAG	369
155487774	CCCAAATCCTTCATCTAA	CAGTCTAACGGTTCCAAT	286
179397807	TACTTGCGTGGCTCTGGT	AAGGCTTGGATTATTATGCTCT	403
39913231	TCACCCCTGAGCCACAGATA	GAGGGAAAGTGGATGGTGGG	303
117122048	TTGCATGGGTAATCAAGGGTCA	GCTGGACTGTAGCATCGTGT	610
156186376	ACGCCTACAGGAACGAGG	GGAATGTTGGGACGGATG	450
179436520	GAAGTGTCCCGTTTCTCA	CAGCCTTCAAACTCCTGT	409
43748503	TGGCTTTGCTTTGGTCGTTC	AAACCAGTTGGGTGAGCAGG	520
66291310	AGCACCCCAAGTACTGCATC	CGGGGTGTGGATGACATTGA	419
88600890	CGACGGCCCCATCGACCTGA	CGTGCCCTTGCTGGGAGTCTGG	520
97503827	TCGGTGTAATAGGGTTATGGTATT	TTCTTCTTTGGCCACAACCATT	399
158636905	ACACGGACCCAGGACAAC	TTCCCTCAATGAAGGTGAAACT	403
179406045	TTTGATTGTGGTGGTGAT	CTCCCAAGTGACTGGATAT	341

Supplementary Table 3. Quality of data obtained by WES.

Sample name	Raw reads	Raw data (G)	Effective (%)	Error (%)	Q20 (%)	Q30 (%)
ASD-1	28464122	7.12	95.73	0.03; 0.04	95.76; 93.25	91.57; 87.61
ASD-2	18423348	4.61	95.92	0.03; 0.04	95.79; 93.10	91.62; 87.36
ASD-3	22964836	5.74	95.87	0.03; 0.04	95.81; 93.28	91.67; 87.70
ASD-4	17576512	4.39	95.78	0.03; 0.04	95.72; 92.50	91.48; 86.34
ASD-5	18637077	4.66	95.88	0.04; 0.04	95.12; 93.79	90.10; 87.89

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Supplementary Table 4. Quantity of variants before and after filtering by 1000Genome, dbSNP, and NHLBI.

Number of sample	Quantity of variants before filtering	Heterozygous variants	Homozygous variants	Quantity of variants after filtering
ASD-1	36774	21711	15063	159
ASD-2	36031	20896	15135	142
ASD-3	36443	21416	15027	143
ASD-4	36240	21528	14712	133
ASD-5	36274	21477	14797	136
Total	181762	107028	74734	713

Supplementary Table 5. 13 variants from WES data of 5 ASD patients by Venn analysis.

Gene	The information of variants	The position of mutation site on a chromosome	The number of sample	Harmful prediction by SIFT/ Ployphen2/Mutation Taster
TTN	NM_003319: c.C70564T	179406045	ASD-5	D/D/D
TTN	NM_003319: c.G47144A	179436520	ASD-4	D/D/D
TTN	NM_003319: c.C21134G	179480499	ASD-2	D/B/N
TTN	NM_003319: c.G48662T	179435002	ASD-2	D/D/D
TTN	NM_003319: c.C76304T	179397807	ASD-3	D/D/D
HYDIN	NM_001270974: c.G7930A	70952188	ASD-1	T/B/D
HYDIN	NM_001198542: c.A2207C	71098693	ASD-2, ASD-3	T/P/D
IGSF3	NM_001542: c.A619G	117156600	ASD-2, ASD-5	T/B/D
IGSF3	NM_001007237: c: G3300C	117122048	ASD-4	T/B/D
ZFPM	NM_153813: c.C2359T	88600725	ASD-2	T/B/D
ZFPM	NM_153813: c.G2524A	88600890	ASD-5	D/P/N
MYH6	NM_002471: c.G985T	23871923	ASD-1	D/./A
MYH6	NM_002471: c.G2851T	23862952	ASD-3	D/./A

According to SIFT, related gene was noted on tolerated (T, score >0.5) or deleterious (D, score <0.5). According to PolyPhen-2, related gene was noted on probably damaging (D, Polyphen-2 ≥0.909), possibly damaging (P, 0.447 ≤Polyphen-2 <0.909), and benign (B, Polyphen-2 <0.447). According to MutationTaster, related genes were noted on disease-causing automatic (A), disease-causing (D), polymorphism (N), and polymorphism automatic (P) (*http://www.mutationtaster.org/*).

Supplementary Table 6. 21 variants associated with cardiac development or CHD obtained by Gene Ontology analysis.

Gene	The information of variants	The position of mutation site on a chromosome	The number of sample	Harmful prediction by SIFT/ Ployphen2/Mutation Taster
BMP4	NM_001202: c.C125T	54418816	ASD-5	T/P/D
TTN	NM_003319: c.G48662T	179435002	ASD-2	D/D/D
DSG4	NM_001134453: c.G790A	28971146	ASD-5	D/D/D
XPO1	NM_003400: c.G2985C	61708404	ASD-1	T/P/D

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Gene	The information of variants	The position of mutation site on a chromosome	The number of sample	Harmful prediction by SIFT/ Ployphen2/Mutation Taster
TGFBR3	NM_001195683: c.G1763A	92181893	ASD-5	T/B/N
HYDIN	NM_001270974: c.G7930A	70952188	ASD-1	TTN
HYDIN	NM_001270974: c.G7930A	70952188	ASD-1	T/B/D
TENM4	NM_001098816: c.G8189C	78369224	ASD-1	./B/N
SGCD	NM_001128209: c.A845G	156186376	ASD-4	T/D/D
MYH6	NM_002471: c.G2851T	23862952	ASD-3	D/./A
SMARCA4	NM_001128845: c.A602T	11097111	ASD-5	T/P/D
ZFPM	NM_153813: c.C2359T	88600725	ASD-2	T/B/D
ZFPM	NM_153813: c.G2524A	88600890	ASD-5	D/P/N
FOXL2	NM_023067: c.C155G	138665410	ASD-1	T/P/D
DNAH7	NM_018897: c.T6053C	196741332	ASD-1	T/D/D
DNAH17	NM_173628: c.C12828A	76422625	ASD-1	D/P/D
FIBP	NM_004214: c.G850A	65652097	ASD-2	T/B/D
ATP2B2	NM_001001331: c.A182G	10491046	ASD-3	T/D/D
ZNF638	NM_001014972: c.A254G	71576338	ASD-3	D/D/D
ALMS1	NM_015120: c.A6808G	73680465	ASD-3	D/D/N
MCHR1	NM_005297: c.G671A	41077334	ASD-3	D/D/D

According to SIFT, related gene was noted on tolerated (T, score >0.5) or deleterious (D, score <0.5). According to PolyPhen-2, related gene was noted on probably damaging (D, Polyphen-2 ≥0.909), possibly damaging (P, 0.447 ≤Polyphen-2 <0.909), and benign (B, Polyphen-2 <0.447). According to MutationTaster, related genes were noted on disease-causing automatic (A), disease-causing (D), polymorphism (N), and polymorphism automatic (P) (*http://www.mutationtaster.org/*).

Supplementary Table 7. 20 variants associated with cardiac development or CHD obtained by literature review and protein database.

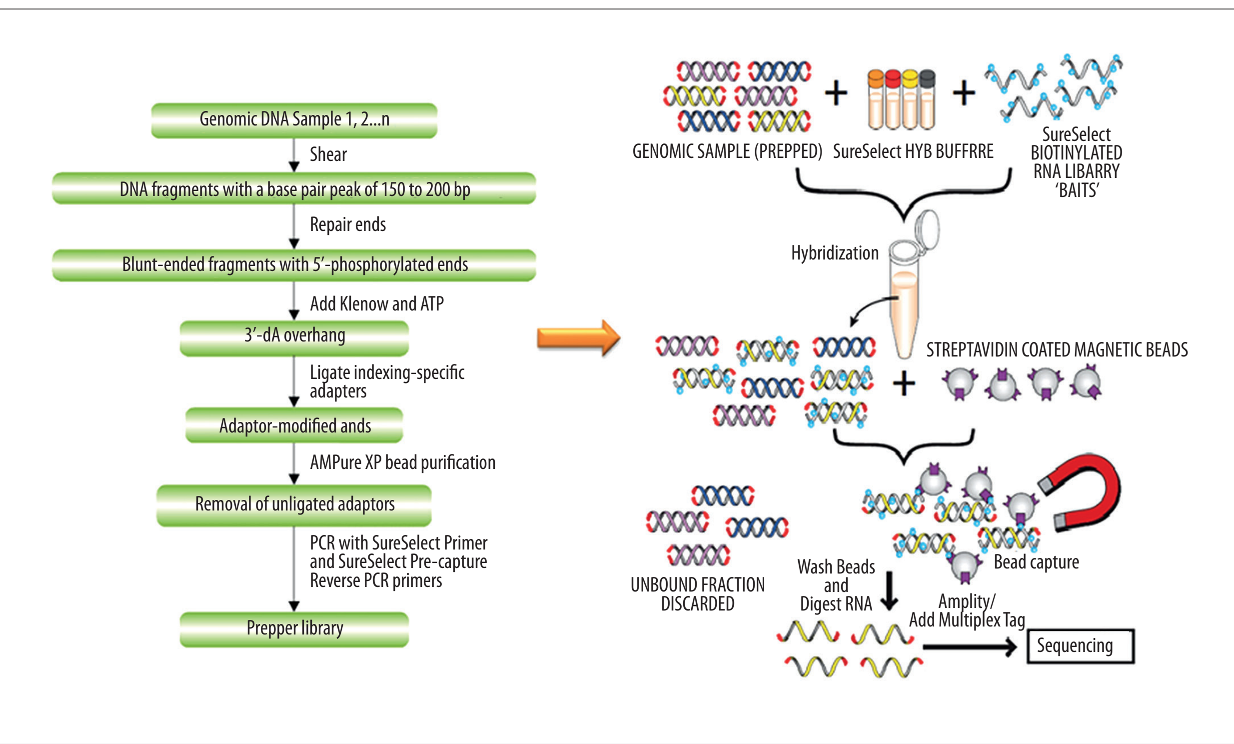
Gene	The information of variants	The position of mutation site on a chromosome	The number of sample	Harmful prediction by SIFT/ Ployphen2/Mutation Taster
ACVR1	NM_001105: c.A275G	158636905	ASD-5	T/P/D
ARL6	NM_001278293: c.G283T	97503827	ASD-5	D/P/D
BBS1	NM_024649: c.G1067A	66291310	ASD-5	T/B/D
VEGFA	NM_001025366: c.C997T	43748503	ASD-5	T/./D
BBS9	NM_014451: c.A2389G	33573776	ASD-1	T/B/D
FMO5	NM_001144830: c.T572A	146684019	ASD-1	D/D/D
FOXL2	NM_023067: c.C155G	138665410	ASD-1	T/P/D
NSD1	NM_022455: c.A2608G	176638008	ASD-1	./D/N
NUP188	NM_015354: c.C3047T	131756681	ASD-1	D/D/D
OBSCN	NM_001098623: c.A7301G	228467050	ASD-1	T/D/D
SOX17	NM_022454: c.A595T	55371905	ASD-1	T/B/N

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Gene	The information of variants	The position of mutation site on a chromosome	The number of sample	Harmful prediction by SIFT/ Ployphen2/Mutation Taster
XPO1	NM_003400: c.G2985C	61708404	ASD-1	T/P/D
BCOR	NM_001123384: c.T4728G	39913231	ASD-4	T/B/N
SGCD	NM_001128209: c.A845G	156186376	ASD-4	T/D/D
EP300	NM_001429: c.G2261A	41545061	ASD-3	T/D/D
FGB	NM_001184741: c.A263C	155487774	ASD-3	T/B/N
KIAA0196	NM_014846: c.A2186C	126062819	ASD-3	T/D/D
USP44	NM_001042403: c.A2134C	95911935	ASD-3	D/B/D
ZNF638	NM_001014972: c.A254G	71576338	ASD-3	D/D/D
FIBP	NM_004214: c.G850A	65652097	ASD-2	T/B/D

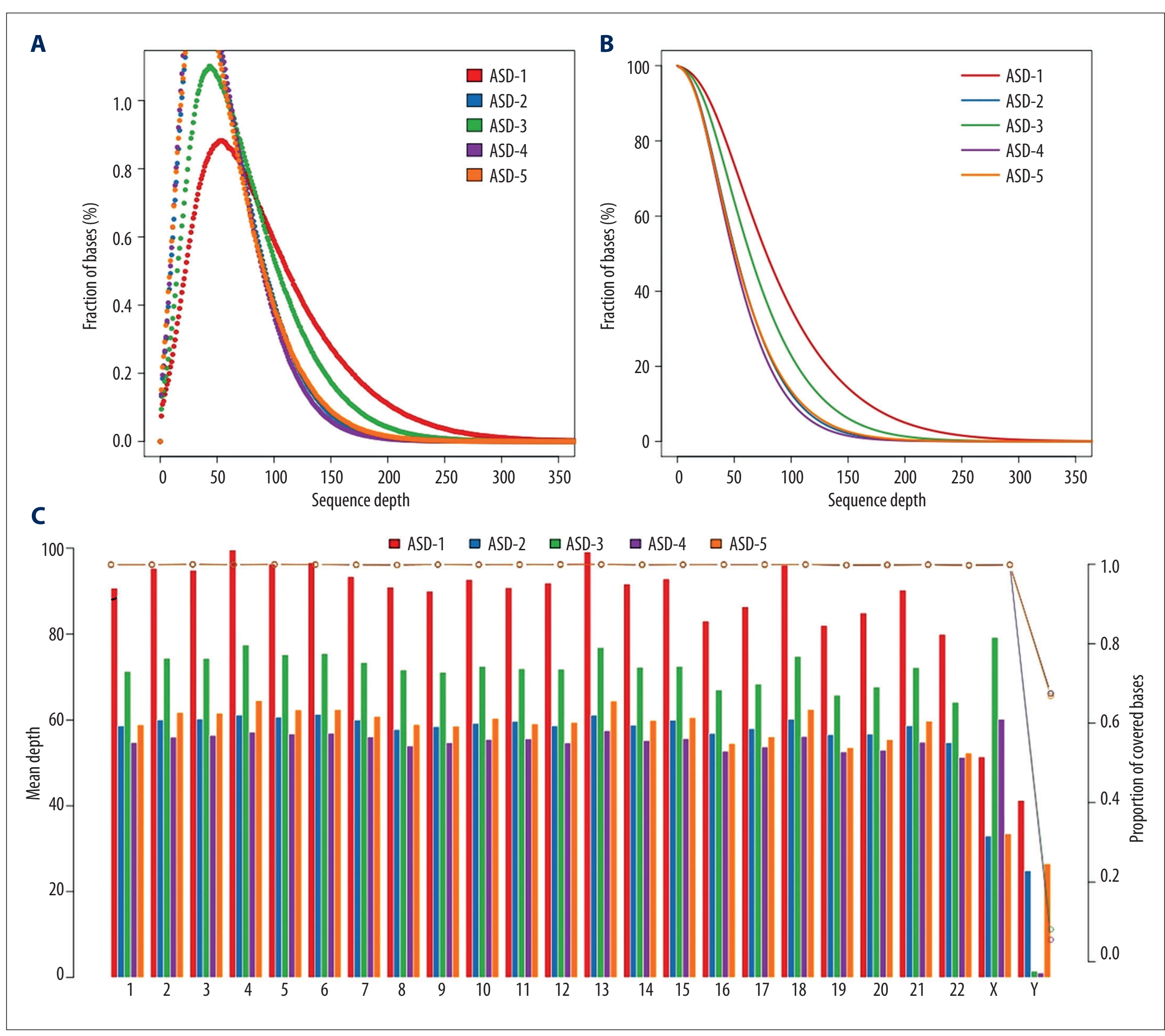
According to SIFT, related gene was noted on tolerated (T, score >0.5) or deleterious (D, score <0.5). According to PolyPhen-2, related gene was noted on probably damaging (D, Polyphen-2 ≥0.909), possibly damaging (P, 0.447 ≤Polyphen-2 <0.909), and benign (B, Polyphen-2 <0.447). According to MutationTaster, related genes were noted on disease-causing automatic (A), disease-causing (D), polymorphism (N), and polymorphism automatic (P) (*http://www.mutationtaster.org/*).



Supplementary Figure 1. Schematic of gene library construction, capture, and sequencing.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

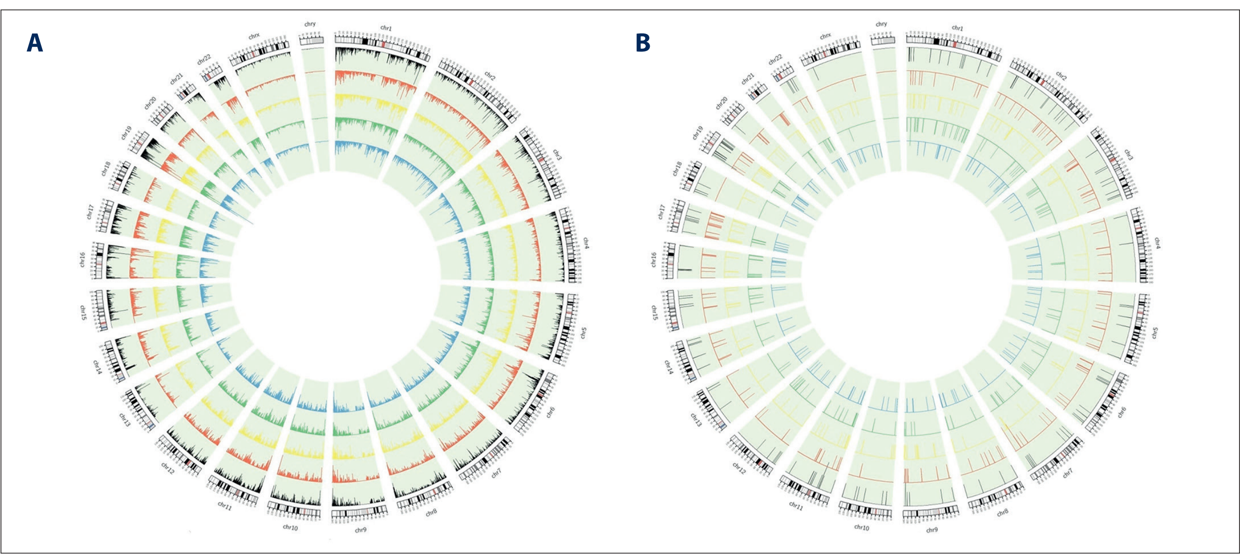
Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



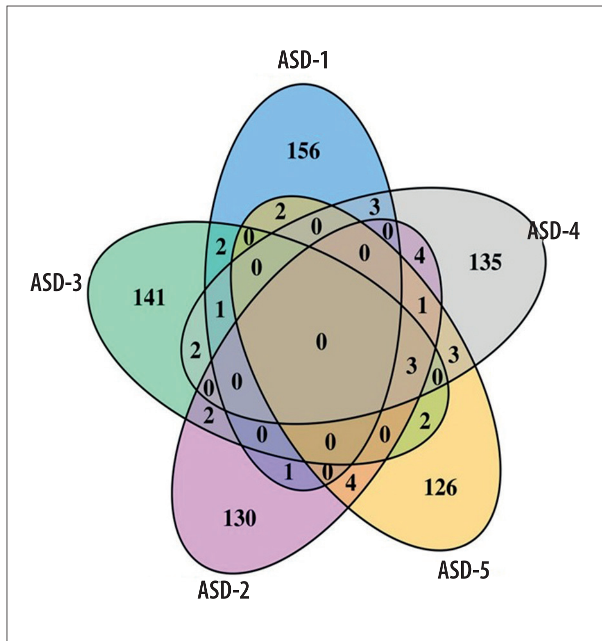
Supplementary Figure 2. The depth of the target areas. (**A**) Sequence depth of the sample. (**B**) Cumulative sequence depth of the sample. (**C**) The depth of coverage (left coordinate) and the ratio of coverage (right coordinate) on the chromosome.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Supplementary Figure 3. The circos map provided to demonstrate the distributions of the variants on the chromosomes. Red – ASD-1; Green – ASD-2; Blue – ASD-3; Yellow – ASD-4; Black – ASD-5. (**A**) Rare variants before filtering. (**B**) Rare variants after filtering.



Supplementary Figure 4. Thirteen variants were obtained by Venn analysis.

FOXL2 NM_023067:exon1:c.C155G:p.A52G 138665410

	46						A52G						56	
Human	- P	E	K	P	D	P	A	Q	K	P	P	Y	S	-
Chimpanzee	- P	E	K	P	D	P	A	Q	K	P	P	Y	S	-
Mouse	- P	E	K	P	D	P	A	Q	K	P	P	Y	S	-
Rat	- P	E	K	P	D	P	A	Q	K	P	P	Y	S	-
Dog	- P	E	K	P	D	P	A	Q	K	P	P	Y	S	-
Cow	- P	E	K	P	D	P	A	Q	K	P	P	Y	S	-

MYH6 NM_002471:exon22:c.G2851T:p.E951X 23862952

	945						E951X						957	
Human	- L	E	D	E	C	S	E	L	K	K	D	I	D	-
Chimpanzee	- L	E	D	E	C	S	E	L	K	K	D	I	D	-
Mouse	- L	E	D	E	C	S	E	L	K	K	D	I	D	-
Rat	- L	E	D	E	C	S	E	L	K	K	D	I	D	-
Dog	- L	E	D	E	C	S	E	L	K	K	D	I	D	-
Cow	- L	E	D	E	C	S	E	L	K	K	D	I	D	-

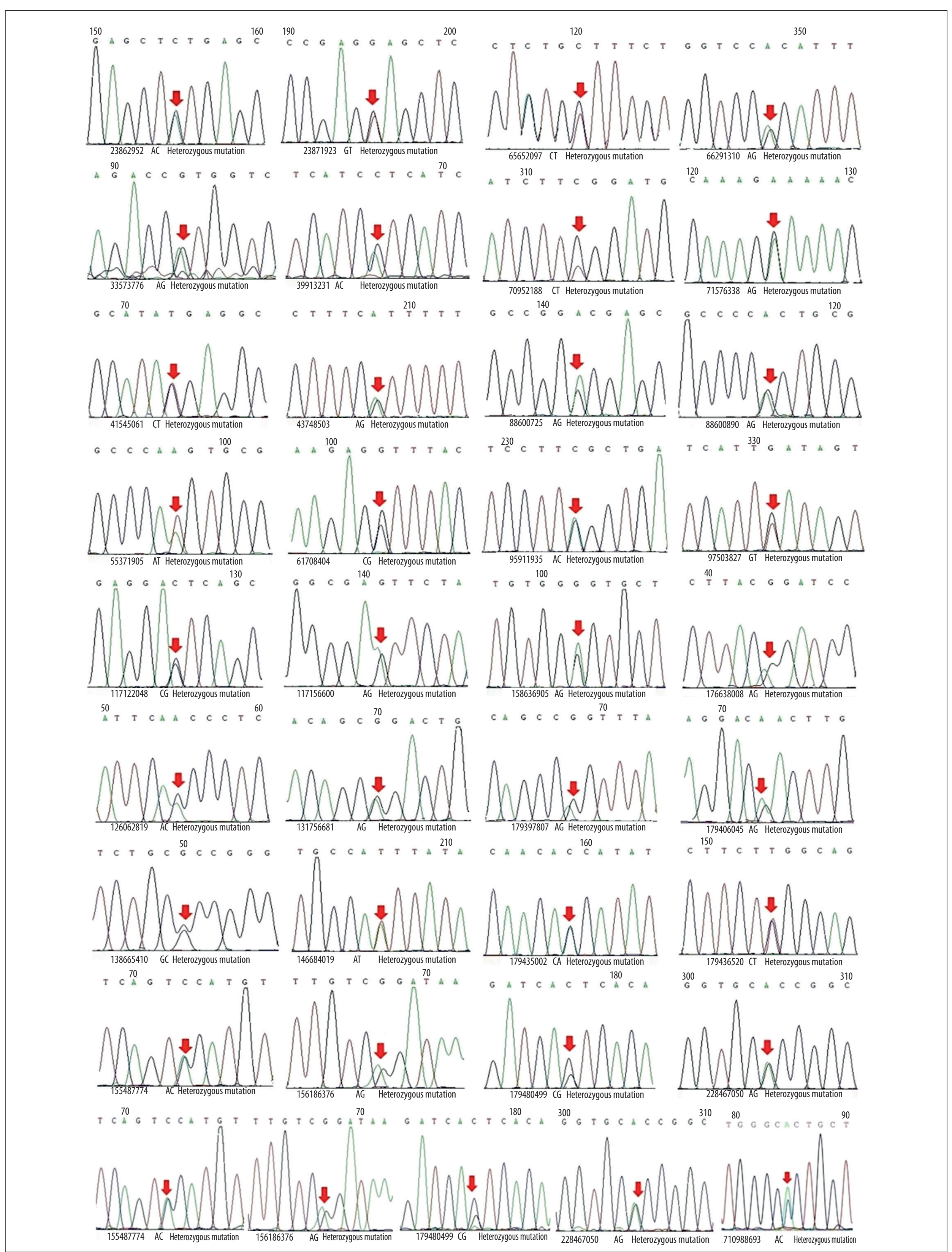
HYDIN NM_001198542:exon16:c.A2207C:p.H736P 71098693

	730						H736P						742	
Human	- T	E	V	D	F	G	H	C	F	L	K	Y	P	-
Chimpanzee	- T	E	V	D	F	G	H	C	F	L	K	Y	P	-
Mouse	- T	E	V	D	F	G	R	C	F	L	K	Y	P	-
Rat	- T	E	V	D	F	G	R	C	F	L	K	Y	P	-
Dog	- T	E	V	D	F	G	H	C	F	L	K	Y	P	-
Cow	- T	E	V	D	F	G	R	C	F	L	K	Y	P	-

Supplementary Figure 6. The conservation of different orthologs presented for the three associated variants in the *HYDIN*, *FOXL2*, and *MYH6* genes

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Supplementary Figure 5. Thirty-three variants were confirmed by Sanger sequencing. Sanger sequencing demonstrates that all of the variants were heterozygous. The variants are marked using red arrows.

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

References:

1. Liu X, Liu G, Wang P et al: Prevalence of congenital heart disease and its related risk indicators among 90,796 Chinese infants aged less than 6 months in Tianjin. Int J Epidemiol, 2015; 44(3): 884–93
2. Sun PF, Ding GC, Zhang MY et al: Prevalence of congenital heart disease among infants from 2012 to 2014 in Langfang, China. Chinese Med J, 2017; 130(9): 1069–73
3. Ottaviani G, Buja LM: Update on congenital heart disease and sudden infant/perinatal death: from history to future trends. J Clin Pathol, 2017; 70(7): 555–62
4. Bruneau BG: The developmental genetics of congenital heart disease. Nature, 2008; 451(7181): 943–48
5. Wang W, Niu Z, Wang Y et al: Comparative transcriptome analysis of atrial septal defect identifies dysregulated genes during heart septum morphogenesis. Gene, 2016; 575(2 Pt 1): 303–12
6. Shendure J, Ji H: Next-generation DNA sequencing. Nature Biotechnol, 2008; 26(10): 1135–45
7. Braenne I, Reiz B, Medack A et al: Whole-exome sequencing in an extended family with myocardial infarction unmasks familial hypercholesterolemia. BMC Cardiovasc Disord, 2014; 14: 108
8. LaHaye S, Corsmeier D, Basu M et al: Utilization of whole-exome sequencing to identify causative mutations in familial congenital heart disease. Circ Cardiovasc Genetics, 2016; 9(4): 320–29
9. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 2009; 25(14): 1754–60
10. Li H, Handsaker B, Wysoker A et al: The sequence alignment/Map format and SAMtools. Bioinformatics, 2009; 25(16): 2078–79
11. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acid Res, 2010; 38(16): e164
12. Ng PC, Henikoff S: SIFT: Predicting amino acid changes that affect protein function. Nucleic Acid Res, 2003; 31(13): 3812–14
13. Adzhubei I, Jordan DM, Sunyaev SR: Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet, 2013; Chapter 7: Unit 7.20
14. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D: MutationTaster evaluates disease-causing potential of sequence alterations. Nature Methods, 2010; 7(8): 575–76
15. Bruneau BG: The developmental genetics of congenital heart disease. Nature, 2008; 451(7181): 943–48
16. Zhu H: Forkhead box transcription factors in embryonic heart development and congenital heart disease. Life Sci, 2016; 144: 194–201
17. Pisarska MD, Kuo FT, Bentsi-Barnes IK et al: LATS1 phosphorylates forkhead L2 and regulates its transcriptional activity. Am J Physiol Endocrinol Metabol, 2010; 299(1): E101–9
18. Caburet S, Georges A, L'Hote D et al: The transcription factor FOXL2: At the crossroads of ovarian physiology and pathology. Mol Cell Endocrinol, 2012; 356(1–2): 55–64
19. Georges A, Auguste A, Bessiere L et al: FOXL2: A central transcription factor of the ovary. J Molec Endocrinol, 2013; 52(1): R17–33
20. Liu CY: Wakayama symposium: Notch-FoxL2-alpha-SMA axis in eyelid levator muscle development and congenital blepharophimosis. Ocul Surf, 2012; 10(4): 221–23
21. Grieskamp T, Rudat C, Ludtke TH et al: Notch signaling regulates smooth muscle differentiation of epicardium-derived cells. Circ Res, 2011; 108(7): 813–23
22. Luxan G, D'Amato G, MacGrogan D, de la Pompa JL: Endocardial notch signaling in cardiac development and disease. Circ Res, 2016; 118(1): e1–e18
23. Ching YH, Ghosh TK, Cross SJ et al: Mutation in myosin heavy chain 6 causes atrial septal defect. Nat Genet, 2005; 37(4): 423–28
24. Jia Y, Louw JJ, Breckpot J et al: The diagnostic value of next generation sequencing in familial nonsyndromic congenital heart defects. Am J Med Genetics Part A, 2015; 167A(8): 1822–29
25. Posch MG, Waldmuller S, Muller M et al: Cardiac alpha-myosin (MYH6) is the predominant sarcomeric disease gene for familial atrial septal defects. PLoS One, 2011; 6(12): e28872
26. Doggett NA, Xie G, Meincke LJ et al: A 360-kb interchromosomal duplication of the human HYDIN locus. Genomics, 2006; 88(6): 762–71
27. Dawe HR, Shaw MK, Farr H, Gull K: The hydrocephalus inducing gene product, Hydin, positions axonemal central pair microtubules. BMC Biology, 2007; 5: 33
28. Lechtreck KF, Delmotte P, Robinson ML et al: Mutations in Hydin impair ciliary motility in mice. J Cell Biol, 2008; 180(3): 633–43
29. Klena NT, Gibbs BC, Lo CW: Cilia and ciliopathies in congenital heart disease. Cold Spring Harb Perspect Biol, 2017; 9(8)pii: a028266

This work is licensed under Creative Common Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]