Shanthithadda and Ramachandra
Egyptian Journal of Medical Human Genetics (2025) 26:72
https://doi.org/10.1186/s43042-025-00708-5

Egyptian Journal of Medical Human Genetics

RESEARCH Open Access

Identification of high confidence genes involved in the manifestation of ventricular septal defect

Chaithra Shanthithadda[1] and Nallur Basappa Ramachandra[1*]

Abstract

Background Ventricular septal defect (VSD) is one of the common congenital heart malformations. Potential risk genes involved in the manifestation of VSD still need to be identified.

Methods The whole exome sequencing was performed on 24 congenital heart disease (CHD) subjects, 19 with VSD, two with VSD-associated atrial septal defects (ASD), and three with tetralogy of fallot (ToF). Besides, data mining using two CHD databases, a knowledgebase for nonsyndromic congenital heart disease-associated risk factors (CHD-RF-KB), and CHDbase and PubMed using medical subject headings were performed to prioritize the genes for further analysis in our study. The functional effects of variations of the selected genes were assessed using prediction tools to check the variant pathogenicity. The protein–protein network was established to identify the interactions among the proteins and their role in heart developmental pathways.

Results The whole exome sequencing analysis identified 628 genes with variants of minor allele frequency < 1%. In addition, the data mining yielded 50 genes, and of these, 34 genes were common in our 24 CHD subjects with new variations. Of the 34 genes, 11 genes, such as *DLC1*, *MTRR, EVC, HOMEZ, EPRS, HOXB1, MTHFD1, MTHFR, NOS3, TRDN,* and *TBX18* recurrently occurred in more than 15 CHD subjects. The variant, c.524C > T in *MTRR*, recurrently occurred in nine CHD subjects, which resulted in the change of an amino acid from serine to leucine, suggesting a change in the functionality of the protein. Besides, four variations in *GATA4* and one in *TBX20* showed high pathogenicity scores. The protein–protein network showed high interactions among 26 transcription factors, and the TNNT2, MYL7, and ZFPM2 were the newly identified proteins in the network.

Conclusions The present study identified 37 potential risk genes through exome sequencing and interaction network analysis based on strong evidence from previously reported studies. From these genes, 176 variations were identified, of which 166 were newly reported from our samples, and 10 were previously reported, indicating that these are novel variants for the Indian population. Therefore, these genes and the variations can be a potential biomarker for creating the gene panel for CHD.

Keywords Ventricular septal defect, Genes, Variations, Genomics, Whole exome sequencing, Network analysis

Background

The cascades of gene regulations control heart development. It is challenging to unravel the specific cause of heart defects due to the complex genetic trait. Congenital heart disease (CHD) is the leading cause of infant death. It is characterized by the abnormality in the function and development of the structure of the heart, affecting ~ 1%

*Correspondence:
Nallur Basappa Ramachandra
nallurbr@gmail.com; ramachandra@genetics.uni-mysore.ac.in
[1] Department of Studies in Genetics and Genomics, University of Mysore, Manasagangotri, Mysuru 570 006, India

of live births [1, 2]. The abnormal structural development of the heart leads to different types of CHD depending on the position of the heart. They are classified under three categories of CHD: left-sided obstruction defects, cyanotic heart disease, and septation defects [3]. The defects in valve formation, patterning of the great vessels, and septation are common in heart development [4]. The ventricular septal defect (VSD), one of the most prevalent septation defects of CHD, accounts for ~20% of the cases [5, 6]. Besides, in India, VSD is the most prominent type, occurring in 30 to 40% of children affected with CHD [7, 8].

Technological advancements and diagnostic techniques have detected disease conditions in the early stages of fetal development. Over the past two decades, next-generation sequencing methods have evolved, allowing the rapid sequencing and analysis of the human genome within hours. Most of the causative variations in genes that code for transcription factors and other types of proteins play a major role in the specific cardiac developmental process, and variants in > 400 genes are estimated to implicate in human CHD [2, 9, 10]. Identifying key genes that regulate the particular stages of heart development is essential to deciphering the molecular architecture of heart development. The interactive networks of these risk genes in the molecular pathways of cardiac development are yet to be studied. Due to the limitations in identifying the genetic etiology of CHD, this study has employed whole exome sequencing (WES) to prioritize the VSD-specific genes. This study also allows for identifying variations occurring in previously implicated nonsyndromic VSD-causing genes using text mining as one of the strategies. The in silico analysis using prediction tools and protein–protein interaction network analysis are utilized in this study to understand the CHD manifestation better. Here, we report the potential risk genes and the variants involved in nonsyndromic VSD-specific CHD.

Methods

Selection of subjects

Twenty-four CHD subjects, including two VSD-associated ASD, three ToF, and 19 VSD, were recruited for this study. The subjects were clinically diagnosed at Jayadeva Hospital and Cheluvamba Hospital, Mysuru. Subjects with a hole size of more than 3 mm were considered for whole exome sequencing. The consent of the subject/parent/guardian was obtained before collecting the blood samples.

Selection criteria

The key inclusion criteria were: (1) clinically diagnosed VSD cases with X-rays, electrocardiograms, and ultrasonic echocardiograms, (2) the subjects including both males and females, and (3) without restriction in the population selection with different ethnic groups. The size of the hole is more than 3 mm. The primary exclusion criteria were: (1) VSD cases without echocardiographic evidence and (2) syndrome-associated VSD cases such as Holt-Oram, DiGeorge, and all other types of associated syndromes.

Whole exome sequencing (WES) and data analysis

The WES of 10 CHD subjects was performed using the NextSeq 500 machine at the coverage of 100X. Later, the WES of 14 CHD subjects was performed using a NextSeq 550 sequencer (Illumina, Inc) with a 150 bp paired-end sequencing kit as per standard manufacturer's guidelines at the coverage of > 70X. The raw data were collected in FASTQ format, which was then aligned, and variant calling was performed using Strand NGS and BaseSpace Variant Interpreter software, which had built-in algorithms with statistical significance. The variant annotation was performed using the wANNOVAR software tool with information on gene variation, protein functional predictions, and allele frequencies. The allele frequency of variation < 0.05% in the population and the synonymous variations were excluded.

Data mining for the VSD reported gene list

The genes reported for nonsyndromic VSD were collected using a knowledgebase for nonsyndromic congenital heart disease-associated risk factors (CHD-RF-KB) [11] and CHDbase [12] databases. Text mining was also performed through PubMed using medical subject headings (MeSH) such as "Heart defects, Congenital". We have considered the studies published from the year 2000 to April 2024. The articles were thoroughly examined to select the genes with their variations. The relevant articles were extracted based on the criteria from the database, and a critical examination of the articles was carried out. Primarily, the titles of each article were assessed to determine their relevance to this study. Later, the abstracts and full texts were retrieved for further examination.

In silico analysis

The reported genes with their variations for the nonsyndromic VSD were analyzed in silico to determine their possible effects on protein function using five prediction tools, sorting intolerant from tolerant (SIFT) [13], https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html;MutationTaster, MutationTaster [14], https://www.genecascade.org/MutationTaster2021/#chrpos; polymorphism phenotyping version 2 (PolyPhen2) [15], http://genetics.bwh.harvard.edu/pph2/; MutationAssessor [16], http://mutationassessor.org/r3/; and Combined

annotation-dependent depletion (CADD) [17], https://cadd.gs.washington.edu/snv. The protein sequences in FASTA format, the position of substitution of amino acids, chromosomal number, and location of the variation were taken as input data. The SIFT score of 0.05 or less is deleterious; greater than 0.05 is signified as a tolerated variation. PolyPhen2 is used to predict the effect on the structure and function of protein due to the amino acid substitution using inbuilt information from different databases and tools. Each amino acid substitution provides the qualitative predictions as 'possibly damaging', 'probably damaging', 'benign', or 'unknown' in the 0 to 1 score range. The cutoff score of 0.5 is considered for PolyPhen2, and scores equal to or greater than 0.5 are considered deleterious. MatationTaster uses different prediction models to categorize nonsynonymous single nucleotide variants (nsSNVs) into binary predictions, such as deleterious or benign, and present the information associated with the given variant. The MutationAssessor helps to predict the impact of amino acid substitutions on protein function, which assesses the effect of amino acids in protein homologs depending on the evolutionary conservation. It indicates the functional impact score of variants as 'high', 'medium', 'low', and neutral. CADD tool predicts the deleteriousness of SNVs by integrating multiple annotations, and Phred style CADD scores are displayed with a cutoff point of 20. The variants showing higher scores (> 20) are predicted as deleterious, whereas the low score (< 20) signifies less pathogenicity.

Protein–protein interaction network analysis

The proteins of the listed genes were considered input for Cytoscape version 3.10.2 using the STRING app. STRING works with an inbuilt curated knowledge base that includes automated scientific literature data mining, co-expression predictions, databases of interaction experiments, and pathways from curated sources [18]. The network was developed for the identified risk genes to check their confidence in their interaction with each other.

Results

Whole exome sequencing data analysis

Identification of genes reported in 24 CHD subjects

The genes having different types of variations, including synonymous, nonsynonymous, frameshift insertion and deletion, nonframeshift insertion and deletion, stop gain, stop loss, and start loss, were identified from the 24 CHD subjects. Initially, we identified about 29,685 genes with overlaps, and after removing the overlaps, 16,369 genes were identified. The commonly used heuristic filtering methods, including allele frequency, prediction tools, and exclusion of synonymous variations, yielded 628 genes. Of the 628 genes, 508 genes were found in VSD subjects, 48 genes were found in ToF subjects, 70 genes were found commonly between VSD and ToF subjects, and two genes were found in VSD-associated ASD subjects. Of the 628 genes, 408 were found with nonsynonymous, 76 with stop gain variations, 38 with nonframeshift deletion, 35 with frameshift deletion, 26 with nonframeshift insertion, 17 with frameshift insertion, seven with stop loss, and five with strat loss variations. The percentage-wise categorization of 628 genes based on their occurrence with different types of variations is shown in Fig. 1.

Identification of VSD candidate gene list and comparison with our WES data

The candidate VSD-specific gene list approach was implemented in this study to identify potentially pathogenic variations in VSD genes that had been previously reported to cause CHD. The text mining using PubMed and two CHD databases using stringent criteria yielded 50 nonsyndromic VSD-specific genes. When compared with our 24 CHD subjects' WES data, 43 common genes were identified. Of these, 34 genes showed different types of variations, which are represented in Table 1. The remaining nine genes had synonymous variations in 24 CHD subjects excluded in this study. Of the 34 genes, 14 coded for transcription factors, seven for enzymes, five for other proteins, four for ligands, and four for receptors. About 15 genes from these were found in > 50% of our samples. Of these, *DLC1* and *MTRR* recurrently occurred in 18 CHD subjects. Of the 34 genes, eight genes, including *EPRS, EVC, GATA4, HOXB1, IL27, MTHFD1, MTRR*, and *NOS3,* with ten variations, were reported in previous studies. The allele frequency of < 1% of variations in eight genes was checked in 1000genome, Exome Aggregation Consortium (ExAC), and Genome Aggregation Database (GnomAD), which are highlighted in Table 1.

Of the 10 common variations, five were found in both VSD and ToF subjects. Of these, the p.D298E in *NOS3* was found in 15 CHD subjects, including 12 VSD and three ToF subjects. Similarly, the p.D308E in *EPRS* was found in 14 CHD subjects, including 11 VSD and three ToF subjects. The p.S175L in *MTRR*, p.R653Q in *MTHFD1*, and p.Q103H in *HOXB1* were found in six, six, and four VSD and three, two, and one ToF subjects, respectively. The remaining four variations were found only in VSD subjects, including p.R576Q in *EVC,* p.S377G, p.P407Q, and p.P408Q in *GATA4*, and p.S59A in *IL27* in three, one, and one subject, respectively.

Table 1 also reveals the genes with variations that are common between VSD and TOF as well as VSD and VSD-associated ASD subjects. Of the 34 genes, 20 genes with variations were common between ToF and isolated VSD subjects. The variations, c.524C > T and c.605C > T

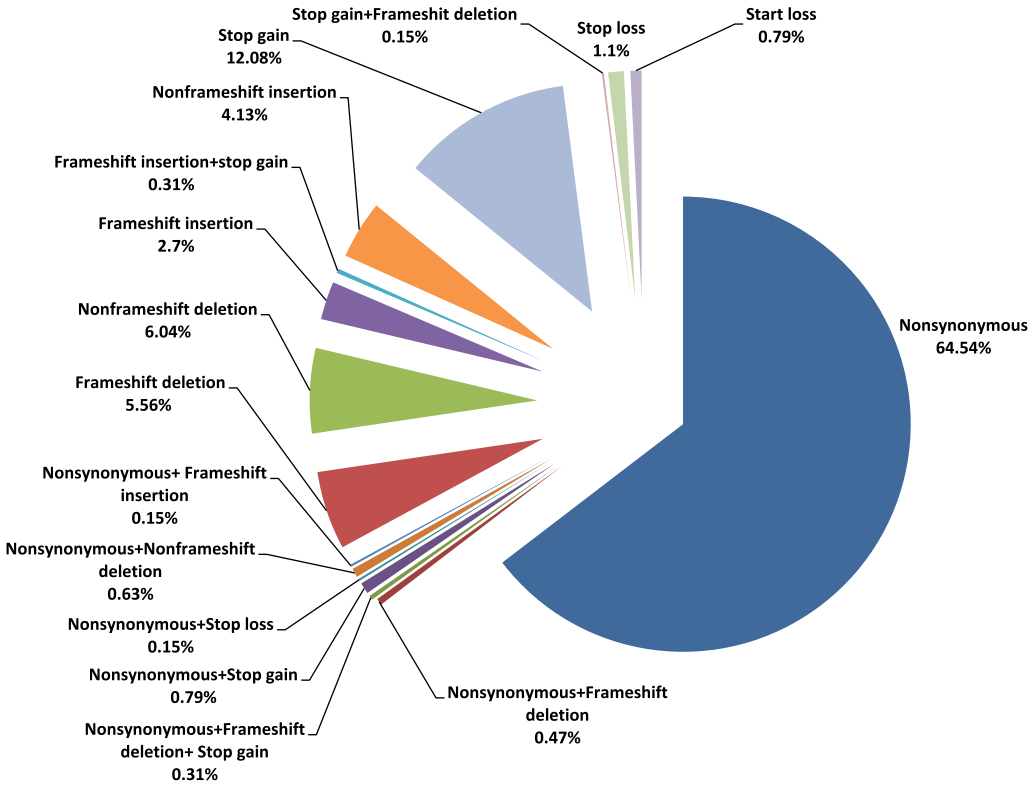

Fig. 1 The percentage-wise categorization of 628 genes based on their occurrence with different types of variations in 24 CHD samples. The 628 genes were found with different types of variations, such as nonsynonymous, frameshift deletion and insertion, nonframeshift deletion and insertion, stop gain, stop loss, and start loss

in *MTRR*, were found in all three ToF subjects and six isolated VSD subjects. Similarly, three genes were found to be shared between VSD-associated ASD subjects and isolated VSD subjects. The variation, c.4G > C in *MESP1*, was found in two VSD-associated ASD subjects and four isolated VSD subjects.

Variations identified in 34 genes

Of the 34 genes, we found nine genes with frameshift and nonframeshift deletion and insertion variations. Of the nine genes, six genes with frameshift deletion, two with nonframeshift deletion, one with frameshift insertion, two with nonframeshift insertion, and five with stop gain variations were observed. This categorization of genes based on the presence of different types of variations is shown in Fig. 2A. The *MESP1* is one of these genes found with three frameshift insertions in six different VSD subjects. The nonsynonymous variations were found in all 34 genes. Besides, *DLC1* and *MTRR* were found to have 20 nonsynonymous variations. In *HOMEZ,* we found four different types of variations, including nonsynonymous, frameshift deletion, nonframeshift deletion, and stop gain, showing 17 variations.

The nonsynonymous variations were found in the highest number of our samples, and some of these variations reoccurred in more than 10 CHD samples. The c.266 T > C, c.455 T > C, and c.596 T > C variants in *BMP4* were found recurrently in 11 CHD subjects, including 10 VSD and one ToF. The *DLC1* with variation c.762A > C was observed in 16 CHD subjects, including 14 VSD and two ToF, and c.779C > T in *DLC1* was observed in 17 CHD subjects, including 14 VSD and three ToF. The variation, c.794A > G in *HOXB1,* was found in 14 CHD subjects, including 11 VSD and three ToF. The variation, c.401A > G in *MTHFD1*, was observed in 15 CHD subjects, including 13 VSD and two ToF. The variation, c.142G > A in *TBX18*, was found in 15 CHD subjects, including 12 VSD and three ToF. The variations, c.17 T > C and c.65 T > C in *TDGF1*, were found in 14 CHD subjects, including 12 VSD and two ToF. Also, the variations c.1316 T > G and c.1313 T > G in *TRDN* occurred in 16 CHD subjects, including 13 VSD and three ToF. These 12 nonsynonymous variations in seven genes were the newly found variations in our analyzed exome data. The distribution of variations in 34 genes and the reoccurrence of these genes across 24 WES CHD subjects are shown in Fig. 2B.

Table 1 Description of 34 genes with variations identified in whole exome sequenced data of 24 CHD subjects

S. no.	Gene symbol	Variation detail	Type of variation	CHD types (No. of samples)	1000genome	ExAC	GnomAD
1	*ACE*	c.38_39insGCTGCC, p.L18_L19insPL	Nonframeshift insertion	VSD (1)	0.02	0.0601	0.0111
		c.1259 T > C, p.I420T	Nonsynonymous	VSD (1)	0.0008	0.0002	0.0002
		c.1136G > A, p.R379Q; c.2858G > A, p.R953Q		VSD (1)	0.0024	0.0013	0.0011
2	*BMP4*	c.30_33del, p.R12Afs*13	Frameshift deletion	VSD (1)	–	–	–
		c.266 T > C, p.V89A; c.455 T > C, p.V152A; c.596 T > C, p.V199A	Nonsynonymous	VSD(10), ToF (1)	0.33	0.457	0.4541
3	*CASZ1*	c.5133delC, p.D1711Efs*131	Frameshift deletion	VSD (1)	–	–	–
		c.5134delG, p.D1712Tfs*130		VSD (1)	–	–	–
		c.5137delG, p.E1713Rfs*129		VSD (1)	–	–	–
		c.5103C > G, p.D1701E	Nonsynonymous	VSD (1)	–	–	0.00001
		c.5118C > G, p.D1706E		VSD (1)	–	–	–
		c.5130C > G, p.D1710E		VSD (1)	–	0.00006	0.0001
		c.5112G > C, p.E1704D		VSD (1)	0.0028	–	0.00006
		c.5124G > C, p.E1708D		VSD (1)	–	–	–
		c.1174A > C, p.T392P		VSD (1)	–	–	–
		c.1180A > C, p.S394R		VSD(3), ToF(2)	–	–	–
		c.1204A > C, p.S402R		VSD (1)	–	–	–
		c.845C > T, p.T282M		VSD (1)	0.0052	0.0082	0.0076
4	*CD80*	c.521C > T, p.S174F	Nonsynonymous	ToF (1)	0.003	0.0016	0.0016
5	*DLC1*	c.29C > A, p.A10E	Nonsynonymous	VSD (2)	0.028	0.0822	0.0498
		c.762A > C, p.Q254H		VSD(14), ToF(2)	0.76	0.8558	0.8513
		c.763A > G, p.N255D		VSD(14), ToF(3)	0.96	0.9673	0.9687
		c.960A > C, p.Q320H		VSD(2), ToF(2)	0.059	0.0219	0.0185
		c.241C > G, p.L81V		VSD(2), ToF(1)	0.085	0.0561	0.0615
		c.779C > T, p.T260I		VSD(14), ToF(3)	0.82	0.8777	0.8699
		c.79C > T, p.R27C		ToF (1)	0.015	0.0247	0.0249
		c.1046G > A, p.R349K		VSD (1)	–	–	–
		c.2579 T > C, p.V860A; c.2801 T > C, p.V934A; c.2903 T > C, p.V968A; c.4112 T > C, p.V1371A		VSD (1)	–	–	–
		c.838G > A, p.V280M; c.1060G > A, p.V354M; c.1162G > A, p.V388M; c.2371G > A, p.V791M		VSD(12), ToF(3)	0.5	0.4935	0.4839
		c.1709C > T, p.T570M; c.1931C > T, p.T644M; c.2033C > T, p.T678M; c.3242C > T, p.T1081M		VSD (1)	–	0.00004	0.00004
6	*EPRS*	c.3550delG, p.E1184Kfs*42	Frameshift deletion	VSD (1)	–	–	–
		c.924C > A, p.D308E**	Nonsynonymous	VSD(11), ToF(3)	**0.9**	**0.845**	**0.8416**
		c.3127A > G, p.I1043V		VSD (1)	0.066	0.0685	0.0662

Table 1 (continued)

S. no.	Gene symbol	Variation detail	Type of variation	CHD types (No. of samples)	1000genome	ExAC	GnomAD
7	*EVC*	c.76G > A, p.A26T	Nonsynonymous	VSD (1)	–	–	–
		c.1346C > A, p.T449K		VSD(9), ToF(2)	0.84	0.7965	0.7901
		c.1428G > C, p.E476D		VSD (1)	0.002	0.0016	0.0015
		c.1727G > A, p.R576Q**		VSD (3)	**0.3**	**0.3472**	**0.3486**
		c.473C > G, p.S158C		VSD (1)	0.0058	0.0048	0.0043
		c.772 T > C, p.Y258H		VSD(9), ToF(3)	0.74	0.7951	0.7967
		c.822G > C, p.K274N		VSD (1)	–	–	–
		c.1055C > T, p.A352V		VSD (1)	–	0.000008	0.00002
		c.1115C > T, p.T372M		VSD (2)	0.027	0.0675	0.0678
8	*GATA4*	c.511A > G, p.S171G; **c.1129A > G, p.S377G****; c.1132A > G, p.S378G	Nonsynonymous	VSD (1)	**0.043**	**0.0962**	**0.0958**
		c.602C > A, p.P201Q;**c.1220C > A, p.P407Q**; c.1223C > A, p.P408Q****		VSD (1)	**0.0012**	**0.0006**	**0.0005**
9	*GATA5*	c.199A > C, p.T67P	Nonsynonymous	VSD (4)	0.14	0.2564	0.1625
10	*GDF3*	c.637G > A, p.G213R	Nonsynonymous	VSD(6), ToF(2)	0.23	0.2642	0.2614
		c.982G > C, p.V328L		VSD(4), ToF(1)	0.038	0.0336	0.0313
		c.147 T > G, p.Y49X	Stopgain	VSD (3)	–	–	–
11	*HOMEZ*	c.432delT, p.H144Qfs*24	Frameshift deletion	VSD (1)	–	–	–
		c.1608delG, p.E537Kfs*9		VSD(2), ToF(1)	–	0.4516	–
		c.1610delA, p.D538Mfs*8		VSD(2), ToF(1)	–	0.2185	–
		c.1614delT, p.D538Efs*9		VSD (1)	–	0.0031	0.001
		c.1631delA, p.D544Vfs*2		VSD (1)	–	0.4888	–
		c.1632delT, p.D544Efs*2		VSD (1)	–	0.0698	–
		c.1633delG, p.D545Mfs*1		VSD (1)	–	0.0699	–
		c.1634delA, p.D545Vfs*1		VSD (2)	–	0.0203	–
		c.1635delT, p.D545Efs*1		VSD (2)	–	0.3461	–
		c.1608_1610del, p.E537del	Nonframeshift deletion	VSD (3)	0.35	0.3729	0.3354
		c.1631_1636del, p.D544_D545del		VSD (2)	–	0.016	0.0195
		c.1634_1636del, p.D545del		VSD (2)	–	0.4332	0.3801
		c.904G > A, p.A302T	Nonsynonymous	VSD(6), ToF(2)	–	0.389	0.3822
		c.1505G > A, p.R502Q		VSD (3)	0.14	0.1605	0.0696
		c.1597G > A, p.E533K		VSD (1)	–	–	–
		c.1608_1612del, p.E537*	Stopgain	VSD (1)	–	0.0031	0.0009
		c.1636delG, p.V546fs*		VSD (2)	–	0.2178	–
12	*HOXB1*	c.82_83insACAGCGCCC, p.A27_P28insHSA	Nonframeshift insertion	VSD (2)	0.13	0.1747	–
		c.309A > T, p.Q103H**	Nonsynonymous	VSD(4), ToF(1)	**0.11**	**0.1622**	**0.1614**
		c.794A > G, p.E265G		VSD(11), ToF(3)	0.99	0.9955	0.9964
13	*IL27*	c.356 T > C, p.L119P	Nonsynonymous	VSD (4)	0.16	0.2851	0.2851
		c.175 T > G, p.S59A**		VSD (1)	**0.067**	**0.062**	**0.0639**
14	*IRX1*	c.185G > T, p.G62V	Nonsynonymous	VSD (1)	–	–	–
15	*IRX4*	c.53 T > C, p.M18T	Nonsynonymous	VSD (2)	–	–	0.00002
		c.355G > A, p.A119T		VSD(1), ToF(1)	0.14	0.1661	0.1665
		c.1545A > T, p.K515N; c.1623A > T, p.K541N		VSD(2), VSD + ASD(1)	–	–	–

Table 1 (continued)

S. no.	Gene symbol	Variation detail	Type of variation	CHD types (No. of samples)	1000genome	ExAC	GnomAD
16	*MESP1*	c.137delC, p.P47Qfs*23	Frameshift deletion	VSD + ASD (1)	–	–	–
		c.137_138insG, p.A48Sfs*24	Frameshift insertion	VSD (1)	–	–	–
		c.156_157insCCGAGCCCCGT, p.A53Pfs*21		VSD (5)	–	0.064	0.0371
		c.158dupC, p.S54Efs*18		VSD (5)	–	0.0957	0.0291
		c.4G > C, p.A2P	Nonsynonymous	VSD(4), VSD + ASD(2)	–	–	–
		c.91G > A, p.D31N		VSD + ASD (1)	–	–	–
		c.157G > C, p.A53P		VSD (4)	0.53	0.4756	0.3396
		c.182 T > G, p.L61R		VSD (5)	0.26	0.0906	0.027
		c.669C > G, p.F223L		VSD (5)	0.2	0.317	0.2193
17	*MSX1*	c.271_272del, p.G91Lfs*83	Frameshift deletion	VSD (1)	–	–	–
		c.119C > G, p.A40G	Nonsynonymous	ToF (1)	0.1	0.152	0.1501
18	*MTHFD1*	c.401A > G, p.K134R	Nonsynonymous	VSD(13), ToF(2)	0.82	0.8331	0.8349
		c.920A > G, p.Y307C		VSD (1)	0.0018	0.0012	0.001
		c.1958G > A, p.R653Q**		VSD(6), ToF(2)	**0.34**	**0.4354**	**0.4453**
19	*MTHFR*	c.89C > T, p.S30F	Nonsynonymous	VSD (1)	0.001	0.0003	0.0002
		c.665C > T, p.A222V; c.788C > T, p.A263V		VSD (3)	0.25	0.3037	0.3143
		c.1286A > C, p.E429A; c.1409A > C, p.E470A		VSD(10), ToF (2)	0.25	0.295	0.2902
		c.1781G > A, p.R594Q; c.1904G > A, p.R635Q		VSD(4), ToF(1)	0.074	0.0563	0.0555
20	*MTRR*	c.66A > G, p.I22M; c.147A > G, p.I49M	Nonsynonymous	VSD (8)	0.36	0.473	0.4678
		c.424 T > C, p.W142R; c.505 T > C, p.W169R		VSD (1)	–	–	–
		c.524C > T, p.S175L**; c.605C > T, p.S202L		VSD(6), ToF(3)	**0.27**	**0.3155**	**0.3115**
		c.1049A > G, p.K350R; c.1130A > G, p.K377R		VSD(4), ToF(1)	0.25	0.1711	0.1748
		c.1243C > T, p.R415C; c.1324C > T, p.R442C		VSD (3)	0.068	0.0566	0.0612
		c.1340A > G, p.K447R; c.1421A > G, p.K474R		VSD (1)	–	–	–
		c.1349C > G, p.P450R; c.1430C > G, p.P477R		VSD (2)	0.065	0.055	0.0586
		c.1498G > C, p.V500L; c.1579G > C, p.V527L		VSD (1)	–	–	–
		c.1783C > T, p.H595Y, c.1864C > T, p.H622Y		VSD (5)	0.22	0.1509	0.1559
		c.1798G > A, p.G600R; c.1879G > A, p.G627R		VSD(1), VSD + ASD(2)	–	–	–
21	*MYBPC3*	c.472G > A, p.V158M	Nonsynonymous	VSD (2)	0.033	0.0904	0.0652
		c.706A > G, p.S236G		VSD (1)	0.067	0.1079	0.0969
		c.1623_1624insT, p.E542*	Stopgain	VSD (2)	–	–	–

Table 1 (continued)

S. no.	Gene symbol	Variation detail	Type of variation	CHD types (No. of samples)	1000genome	ExAC	GnomAD
22	*MYH6*	c.166G > A, p.G56R	Nonsynonymous	VSD(1), ToF(1)	0.057	0.0698	0.0682
		c.3302 T > C, p.V1101A		VSD(6), ToF(1)	0.37	0.3461	0.3362
		c.3388G > A, p.A1130T		VSD (1)	0.073	0.1003	0.0941
		c.3883G > C, p.E1295Q		VSD (1)	0.0042	0.0032	0.0032
		c.4471G > A, p.E1491K		VSD (1)	–	–	0.00001
		c.4778A > T, p.Q1593L		VSD (1)	0.049	0.0147	0.0112
		c.4838 T > C, p.V1613A		VSD (1)	0.049	0.0146	0.0109
23	*NKX2.6*	c.742A > G, p.S248G	Nonsynonymous	VSD + ASD (1)	–	–	–
		c.732C > A, p.Y244X	Stopgain	VSD + ASD (1)	–	–	–
24	*NODAL*	c.95A > G, p.H32R; c.494A > G, p.H165R	Nonsynonymous	VSD(9), ToF(1)	0.67	0.6171	0.6169
25	*NOS3*	**c.894 T > G, p.D298E****	Nonsynonymous	VSD(12), ToF(3)	**0.82**	**0.753**	**0.7497**
		c.985C > T, p.R329C		VSD (1)	–	0.000008	0.000008
		c.1805G > A, p.C602Y		VSD(3), ToF(1)	0.12	0.1305	0.0917
		c.3512G > A, p.S1171N		VSD (1)	–	–	–
		c.2269A > T, p.K757X	Stopgain	VSD (1)	–	–	–
26	*NOTCH1*	c.52G > A, p.A18T	Nonsynonymous	VSD (1)	–	0.0073	0.0015
27	*NTRK3*	c.1788_1790del, p.K596_I597delinsN; c.2058_2060del, p.K686_I687delinsN; c.2082_2084del, p.K694_I695delinsN	Nonframeshift deletion	VSD + ASD (1)	–	–	–
		c.1742A > G, p.D581G; c.2012A > G, p.D671G; c.2036A > G, p.D679G	Nonsynonymous	VSD (1)	–	–	–
28	*SALL4*	c.1520 T > G, p.L507R	Nonsynonymous	VSD(7), ToF(3)	0.36	0.3506	0.3507
29	*TBX1*	c.1025G > A, p.R342Q	Nonsynonymous	VSD (1)	0.0046	0.0219	0.0023
		c.1049C > T, p.T350M		VSD(9), ToF(2)	0.21	0.2111	0.2121
		c.1189A > C, p.N397H		VSD (3)	0.23	0.3115	0.2498
30	*TBX18*	c.142G > A, p.G48R	Nonsynonymous	VSD(12), ToF(3)	0.41	0.5981	0.5041
		c.868G > A, p.G290R		VSD(4), ToF(1)	0.014	0.0081	0.0071
31	*TDGF1*	c.17 T > C, p.V6A; c.65 T > C, p.V22A	Nonsynonymous	VSD(12), ToF(2)	0.51	0.5243	0.5177
		c.79 T > G, p.Y27D; c.127 T > G, p.Y43D		VSD (4)	0.078	0.0579	0.0571
32	*TRDN*	c.383C > G, p.T128S	Nonsynonymous	VSD (5)	0.39	0.4926	0.4419
		c.509G > T, p.G170V		VSD (1)	–	–	–
		c.601C > G, p.L201V		VSD (9)	0.84	0.8259	0.8443
		c.1214 T > G, p.V405G; c.1211 T > G, p.V404G		VSD(5), ToF(1)	0.17	0.2004	0.1535
		c.1316 T > G, p.I439S; c.1313 T > G, p.I438S		VSD(13), ToF(3)	0.94	0.924	0.9256
		c.1620A > G, p.I540M		VSD (1)	0.04	0.0343	0.0151
33	*TWIST1*	c.194A > G, p.E65G	Nonsynonymous	VSD (1)	–	–	–
34	*VEGFA*	c.535G > A, p.E179K	Nonsynonymous	VSD (1)	0.001	0.0011	0.001

The variations highlighted with double asterisks are the common variations found between data mining and WES data of 24 subjects

Data mining

The data mining for nonsyndromic VSD-specific genes yielded 50 potential genes with variations. Of these 50 genes, 26 were coded for transcription factors, 10 were for enzymes, four for ligands, five for receptors, and five for other proteins. A detailed description of these genes

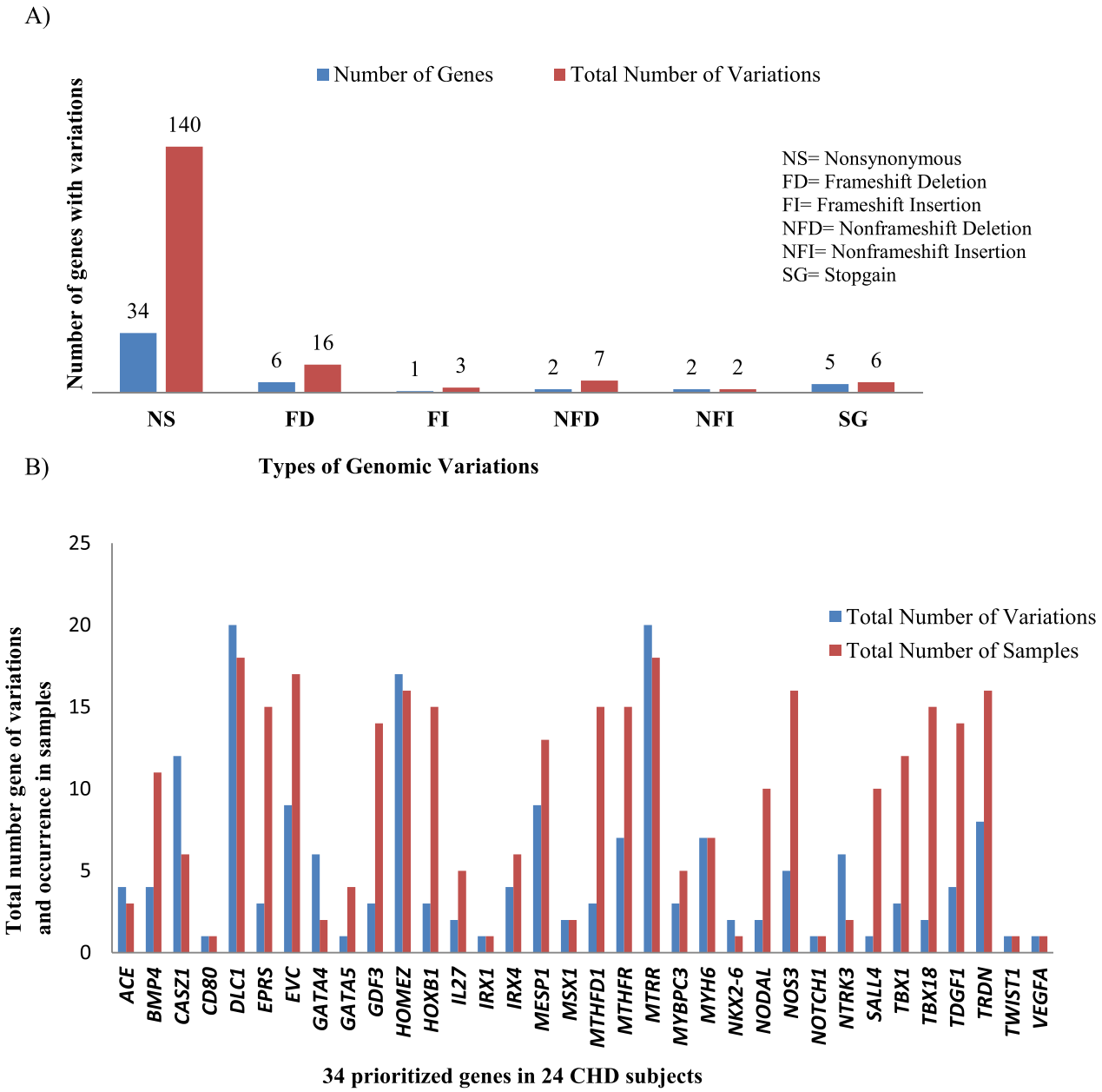

Fig. 2 Types of variations among 34 genes identified through WES in 24 CHD subjects. **A** The bar graph represents quantified different types of variations identified in 34 genes. **B** The bar graph represents the sample-wise distribution and the number of variations identified in the 34 genes

is listed in Tables 2 and 3. These genes were found on different chromosomes. Most case–control studies, i.e., 46 studies on CHD, including 18,304 cases and 22,104 controls in the Chinese population, found variations in 37 genes. Similarly, six Iranian population studies, including 733 cases and 610 controls; six Indian population studies, including 1,326 cases and 1,200 controls; and three American population studies, including 1299 cases and 523 controls, reported gene variations in eight, two, and three genes, respectively. Most of the studies with reported genes and their variations showed heterozygous conditions with dominant mode of inheritance (Tables 2 and 3).

Association of genes reported for VSD with other CHD types

Of the 50 genes, 20 were associated with 15 other CHD types. This information is collected from the same studies that we referred to for VSD-specific genes. The list of

Table 2 Description of the genes with their variations coding for transcription factors involved in VSD

S. no.	Gene	chromosome location (Gene size in kb)	Study population	Sample size Case:control/ familystudy* (VSD cases in %)	Variations	PubMed ID
Transcription Factors						
1	*CASZ1*	1p36.22 (160.047)	Chinese	172:200 (21.51)	c.113 T > C, p.L38P**	27693370
2	*CITED2*	6q24.1 (2.842)	Chinese	700:250 (43.71)	c.548C > T, p.S183L; c.574-579delAGCGGC ,p.S192G193delAGCGGC; c.586A > G, p.S196G	24848765
			Iranian	150:98 (64.66)	c.389A > G, p.N125S**; c.450G > C: p.Q145H**;c. C512-538del27**, c.A-90G; c.716insG**	31515672
			Chinese	200:200 (100)	g.4078A > C**; g.4255C > T**; g.4698A > G**; g.4778G > T**; g.4933 C > A**	33439552
			European	392:192 (45.9)	c.508_534del27 p.Ser170_Gly178del	16287139
			Indian	271:200 (NA)	c.301C > T, p.P101S	33706167
3	*EPAS1*	2p21 (93.031)	Tibetan	286:250 (NA)	c.2170G > T, p.G724W**	30487161
4	*FOXH1*	8q24.3 (2.438)	Chinese	412:250 (NA)	c.659_660insG**	19525021
5	*GATA4*	8p23.1 (83.068)	Chinese	486:486 (65.63)	c.17C > T, p.A6V**; c.1075G > A, p.E359K**; c.487C > T, p.P163S**; c.1286G > C, p.S429T**, c.1325C > T, p.A442V**	18672102
			Chinese	210:100 (100)	c.886G > C, p.G296R	21637914
			Chinese	172:171 (100)	g.4071 T > C**; g.4148 C > A**; g.4566C > T**; g.4653 G > T**; g.4690delG**	22500510
			Chinese	384:760 (42.18)	c.749 T > A, p.I250N; c.1079A > G, p.E360G; c.1325C > T, p.A442V	23626780
			Chinese	224:121 (21.87)	c.788 C > G, p.A263G; c.1329 + 84C > T	22959235
			Indian	100:200 (32)	c.11615928G > A, p.D425N; rs4841587, G > T; rs4841588, G > T; rs804280, A > C	25928801
			Egyptian	165:93 (47.27)	c.578C > A, P193H	28263493
			Iranian	100:50 (57)	c.526G > C, p.A176 P; c.674C > G, p.T225S; c.766A > G,p.K256E; c.958C > T, p.R320W; c.983C > G, p.S328C; c.1325C > T, p.A442V	29377543
			Indian	285:200 (NA)	c.1263C > T; S421S	30590232
			Iranian	66*	c.1220C > A, p.P407Q	31115957
			Chinese	98:200 (21.42)	c.1129A > G, p.S377G	30121862
			Chinese	235:206 (NA)	c.1223C > A, p.P408Q	31513339
			Indian	285:200 (NA)	c.23C > A, p.A8D	30152191
6	*GATA5*	20q13.33 (12.499)	Chinese	320:200 (26.25)	c.569 T > C, p.V190A**; c.821A > G, p.H274R**	23031282
			Chinese	343:348 (100)	g.61051165A > G**; g.61051463delC**	25515806
			Chinese	230:200 (100)	c.164A > G, p.Q55R; c.286G > A, p.G96R; c.590A > G, p.N197S; c.1211A > G, p.K404R	22648249
			European	331:384 (NA)	1692G > A, p.V380M	22011241
			Chinese	62:117 (43.54)	c.1273G > A, p.D425N	19302747
			American	185:150 (10.27)	c.61050385C > A, p.R61S	27066509
7	*GATA6*	18q11.2 (32.940)	Chinese	130:200 (100)	c.658G > A, p.G220S**	22407241
			Chinese	359:365 (100)	g.22169190A > T**; g.22169311C > G**	25036032
8	*HAND1*	5q33.2 (3.256)	Chinese	498:250 (61.04)	c.217G > A, p.G73S; c.456G > T, p.K152N	22032825
9	*HAND2*	4q34.1 (5.261)	Chinese	192:300 (16.14)	c.194G > T, p.S65I**	26865696
10	*HOMEZ*	14q11.2 (27.375)	Chinese	400:400 (100)	c. 116C > T,p.A39V; c. 630 T > A, p.S210R	23574532
11	*HOXB1*	17q21.32 (2.486)	European	57*(33.33)	c.309A > T, p.Q103H	29923154
			Indian	285:200 (NA)	c.1220C > A, p.P407Q	30152191
12	*IRX1*	5p15.33 (5.572)	Chinese	215:249 (24.18)	c.1142C > A, p.A381E	28358424
13	*IRX4*		Chinese	698:250 (48.85)	c.253A > T, p.N85Y; c.275A > G, p.E92G	21544582
14	*ISL1*	5q11.1 (11.283)	Chinese	512:612 (100)	rs3762977; IVS + 17C > T	23572340
			Chinese	210:256 (16.19)	c.409G > T, p.E137X**	30390123

Table 2 (continued)

S. no.	Gene	chromosome location (Gene size in kb)	Study population	Sample size Case:control/ familystudy* (VSD cases in %)	Variations	PubMed ID
15	*MESP1*	15q26.1 (16.478)	American	647:97 (43.27)	c.503A > G, p.D168G; c.436_437delAG,p.L147PfsX9	26694203
16	*MSX1*	4p16.2 (4.272)	Chinese	300:400 (42.66)	rs3821949; rs12532	26556783
17	*NKX2.5*	5q34 (3.213)	Chinese	268:200 (32.09)	c.106C > A, p.R36S**	22179962
			Chinese	391:487 (100)	g.1500G > C	26297999
			Iranian	105:92 (64.76)	c.95A > T, p.E32V	30344277
			Indian	100:200 (32)	c.46G > A, p.D16N	29568389
			Chinese	213:194 (100)	g.4574Cdel	22576768
			Indian	285:200 (NA)	c.284G > T, p.R95L**	34214246
18	*NKX2.6*	8p21.2 (5.017)	Chinese	210:200 (31.42)	c.454A > C, p.K152Q**	25380965
19	*PITX2c*	4q25 (24.701)	Chinese	382:200 (31.15)	c.457A > G, p.N153D**	24083357
20	*SALL4*	20q13.2 (20.191)	Chinese	300:250 (100)	c.586C > T, p.R196W; c.2389A > T, p.S797C	19619907
21	*TBX1*	22q11.21 (26.891)	Chinese	280:267 (100)	g.4199 T > C (rs41260844)	22801995
22	*TBX18*	6q14.3 (77.248)	Chinese	326:327 (100)	g.85473965C > G**; g.85474418C > T**; g.85474871C > T**; g.85474435del**	23749171
23	*TBX20*	7p14.2 (54.165)	Japanese	111:200 (21.62)	c.991A > G, p.T331A	26490186
			Chinese	265:242 (100)	g.4932G > A	22465533
			Australian	353:300 (11.61)	c.456C > G, p.I152M**	17668378
24	*TBX5*	12q24.21 (54.798)	Chinese	192:192 (100)	rsl1067075	19187613
			Chinese	210:205 (100)	g.4303C > G**, g.4900C > T**, g.5126 T > G**	22901678
			Japanese	111:200 (NA)	c.791G > A,p.R264K	26490186
			Chinese	354:341 (100)	c.40C > A, p.P14T**	28434921
25	*TCF21*	6q23.2 (6.441)	Chinese	781:867 (100)	rs12190287	28346832
26	*TWIST1*	7p21.1 (96.646)	Chinese	196:200 (100)	c.247G > A, p.G83S**; c.283A > G, p.S95G**	25981568

NA, not available

'*' represents family studies; '**' represents the gene with variations showing heterozygous condition

genes and their variations identified in VSD associated with other CHD types are shown in Table 4. The VSD-associated double outlet right ventricle (DORV) was the common condition found in 11 genes. Of the 11 genes, the *CAZ1, GATA5*, and *HAND2* were found in VSD-associated DORV, with common variations reported in subjects with only VSD. The diagrammatic representation of genes reported for VSD-associated other CHD types is shown in Fig. 3.

In silico analysis

The nonsynonymous variations identified in 50 genes were subjected to in silico analysis, and the pathogenicity score was calculated (Table 5). The 32 genes of the 50 were observed with nonsynonymous variations. The *GATA4* was found with 19 nonsynonymous variations, of which p.T225S, p.G296R, p.R320W, and p.A442V showed damaging effects on protein function. The p.A442V in *GATA4* recurrently occurred in studies that included two Chinese and one Iranian population. Besides, of the four variations observed in *TBX20*, p.I152M showed damaging effects.

Protein–Protein interaction network analysis

The protein–protein interaction for 50 proteins built a network, as shown in Fig. 4. This network showed the highest interaction of protein families such as NKX, TBX, GATA, and HAND. Proteins such as GATA4, NKX2.5, HAND2, and MYH6 showed darker edges (darker dotted lines) with other proteins in the network, indicating more confident interactions with strong evidence. The transcription factors (diamond shape) showed more interaction with other network proteins. Similarly, the enzymes (hexagonal) such as MTHFD1, MTHFR, SIRT1, ACE, MTRR, and NOS3 showed high confidence interactions with each other by creating a clustered interaction. The BMP4, one of the ligands (circle), showed the highest interaction with transcription factors. Similarly, the receptor NOTCH1 interacted more with transcription

Table 3 Description of genes coding for enzymes, ligands, receptors, and other proteins and the variations involved in VSD

S. no.	Gene	chromosome location (Gene size in kb)	Study population	Sample size Case:control/ familystudy* (VSD cases in %)	Variations	PubMed ID
These genes were reported in previous studies						
1	*ACE*	17q23.3 (21.320)	Iranian	102:98 (61.76)	g.2350A > G	28865601
2	*DARS1*	2q21.3 (80.220)	Chinese	841:2953 (100)	rs2164331, rs6738266, rs309143	27871331
3	*EPRS*	1q41 (77.931)	Chinese	984:2953 (59.45)	c.3405G > A:p.Q1135Q (rs1061160); rs1061248; c.924C > A:p.D308E (rs2230301); c.3102G > A:p.K1034K (rs5030754)	25310850
4	*HAS2*	8q24.13 (29.325)	Chinese	100:250 (100)	c.1496A > T, p.E499V	24558368
5	*MTHFD1*	14q23.3 (75.427)	Iranian	102:98 (61.76)	g.1958G > A	28865601
6	*MTHFR*	1p36.22 (20.733)	Iranian	153:147 (48.36)	g.677C > T	28702146
7	*MTRR*	5p15.31 (54.840)	Iranian	123:125 (100)	c.524C > T; c.66A > G	23358257
			Chinese	183:201 (100)	c.524C > T; c.66A > G	29293099
			Chinese	599:672 (47.07)	c.524C > T; c.66A > G	22057956
8	*NOS3*	7q36.1 (23.572)	Chinese	945:972 (NA)	rs7830	24938467
			Iranian	102:98 (61.76)	g.894G > T	28865601
9	*NTRK3*	15q25.3 (397.048)	American	467:276 (60.81)	c.278C > T, p.T93M; c.488A > T, p.N163I; c.1597A > T, p.I533F	25196463
10	*SIRT1*	10q21.3 (33.735)	Chinese	333:348 (100)	g.69643693A > G; g.69643963A > T; g.69643971G > A; g.69644366Ins	22885181
Ligands						
11	*BMP4*	14q22.2 (9.026)	Chinese	575:844 (71.30)	rs762642	25022354
12	*GDF3*	12p13.31 (5.992)	Chinese	200:202 (47)	c.635C > T, p.S212L	25372014
13	*NODAL*	10q22.1 (16.016)	Chinese	800:250 (55.12)	c.182 T > A, p.L61N	22352765
14	*VEGFA*	6p21.1 (16.304)	Chinese	222:352 (100)	G-634C (rs2010963)	17625508
Receptors						
15	*CD80*	3q13.33 (35.322)	Chinese	3*	g.119537362delT	31440271
16	*IL27*	16p12.1-p11.2 (12.690)	Chinese	150:368 (100)	c.175 T > G, p.S59A	25662568
17	*NOTCH1*	9q34.3 (51.616)	Iranian	4*	c.6797 T > C, p.F2266S**	31867804
18	*TGFBR2*	3p24.1 (87.671)	Chinese	115:188 (100)	rs6785358	26022443
19	*TDGF1*	3p21.31 (7.924)	Chinese	500:250 (NA)	c.110C > G, p.R41G**	19853938
Other proteins						
20	*DLC1*	8p22 (521.260)	Chinese	151:500 (NA)	c.1661A > T, p.D554V; c.1662 T > C, p.D554V; c.2854C > G, p.L952V; c.4111G > C, p.V1371L; c.4533C > G, p.I1511M	24587289
21	*MYBPC3*	11p11.2 (21.297)	Chinese	3*	g.47342683C > T, p.G507R	31440271
22	*MYH6*	14q11.2 (26.292)	European	68:100(NA)	c.831G > T, p.Q277H; c.119C > G, p.P40R; c.4822C > T, p.R1608C	29332214
23	*TRDN*	6q22.31 (420.755)	Chinese	3*	g.123571021C > A, p.S45I	31440271
24	*EVC*	4p16.2 (117.857)	Chinese	65:210 (100)	c.343C > G, p.L115V; c.1727G > A, p.R576Q	29257216

NA, not available

'*' represents family studies; '**' represents the gene with variations showing heterozygous condition

factors, enzymes, and ligands. All these high-interacting proteins are the potential risk genes reported for VSD.

Discussion

In the last twenty years, we have seen tremendous progress in diagnosing genetic diseases with the advent of technology. The VSD is considered one of the common congenital heart malformations, and its frequency varies with age at examination and the sensitivity of the examination technique [19]. VSD is the most prominent CHD in India [8, 20]. In addition, CHD shows heterogeneity; it is important to identify the genes and their variants involved in the manifestation of the disease. The variations in genes encoding transcription

Table 4 Description of genes and the variations involved in VSD and other 15 CHD types

Gene	CHD types	Study type		Variations	PubMed ID
		Population	Case: control/ family study		
Transcription factors					
CASZ1	VSD + DORV	Chinese	172:200	**c.113 T > C, p.L38P**	27693370
CITED2	VSD + ASD	Chinese	120:100	c.573-578del6, p.S192fs	24456003
GATA4	VSD + PA	Japanese	111:200	c.431C > T, p.A144V	26490186
	VSD + PA	Chinese	600: 300	c.1129A > G, p.S377G; c.1138G > A, p.V380M; c.1220C > A, p.P407Q	28161810
	VSD + ASD	Egyptian	165:93	578C > A, p.P193H	28263493
GATA5	VSD + AS	Chinese	320:200	c.821A > G, p.H274R	23031282
	VSD + DORV			**c.569 T > C, p.V190A**	
	VSD + AS + BAV	Chinese	110:200	c.46 T > G, p.Y16D	24638895
GATA6	VSD + ASD	Chinese	130:200	c.658G > A, p.G220S	22407241
HAND1	VSD + OFT	European	12:100	c.376delG, p.A126fs	19586923
	VSD + DORV	Chinese	158:300	c.394A > T, p.K132X	28112363
HAND2	VSD + DORV, VSD + PS	Chinese	192:300	**c.194G > T, p.S65I**	26865696
ISL1	VSD + PDA	Chinese	210:256	c.409G > T, p.E137X	30390123
	VSD + DORV	Chinese	114:218	c.225C > G; p.Y75*	31484864
MEF2C	VSD + PDA	Chinese	200: 300	c.113 T > C, p.L38P	29104469
	VSD + DORV	Chinese	186: 300	c.43C > T, p.R15C	29468350
MESP1	VSD + DORV	Chinese	178: 200	c.352C > T, p.Q118X	28677747
NKX2.5	VSD + TOF	Iranian	105:92	c.95A > T, p.E32V	30344277
	VSD + PS	European	68:100	c.919G > A, p.G307R	29332214
NKX2.6	VSD + DORV	Chinese	320:200	c.529A > T, p.K177X	25195019
PITX2c	VSD + DORV	Chinese	382:200	c.441G > A; p.W147X	24083357
	VSD + TGA	Chinese	170:200	c.272G > A, p.R91Q;	
c.385A > T, p.T129S	24,604,414				
TBX1	VSD + DORV	Chinese	230: 200	c.829C > T, p.Q277X	25860641
	VSD + TGA	Chinese	136: 300	c.698C > A, p.S233Y	29250159
TBX20	VSD + DORV	Chinese	146:200	c.400G > T, p.R134W	25625280
	VSD + PTA	Chinese	175: 400	c.820A > T, p.K274X	28553164
TBX5	VSD + PA	Japanese	111:200	c.791G > A, p.R264K	26490186
ZFPM2	VSD + TGV + PS	European	68:100	c.260C > T, p.P87L	29332214
Ligand					
GDF1	VSD + CoA	European	68:100	c.409G > A, p.D137N	29332214
Other proteins					
DLC1	VSD + PFO	Chinese	151:500	c.797G > A p.G266E	24587289
ELN	VSD + PA +	European	68:100	c.493G > T, p.V165L;	29332214
	MAPCAs			c.892G > A, p.V298I	

PA, pulmonary atresia; DORV, double outlet right ventricle; AS, aortic sclerosis; BAV, bicuspid aortic valve; OFT, outflow tract abnormality; PS, pulmonary stenosis; TGA, transposition of great arteries; PDA, patent ductus arteriosus; PTA, persistent truncus arteriosus; CoA, coarctation of aorta; PFO, patent foramen ovale; MAPCAs, major aorto-pulmonary collateral arteries

factors, chromatin modifiers, and cell signaling transducers affect the development of the heart by interfering in cell type specification, differentiation, and patterning [21]. These encoded proteins work synergistically in heart developmental pathways.

In this study, in the prioritized 34 potential genes, we found different types of variations such as non-synonymous, frameshift deletion and insertions, non-frameshift deletion and insertion, stop gain, stop loss, and start loss variations. Of the 34 genes, the *CASZ1*,

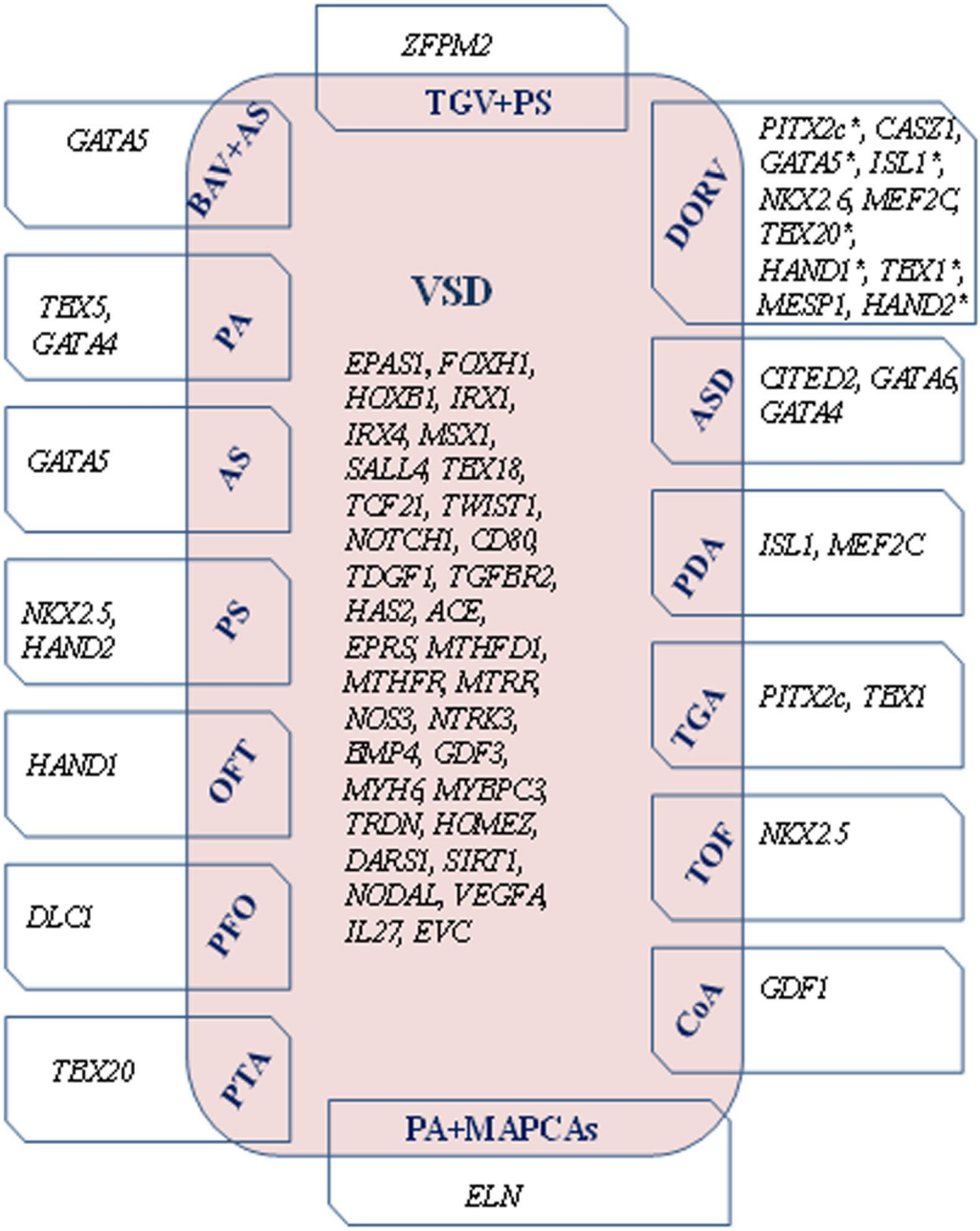

Fig. 3 Genes involved in VSD-associated CHDs. The '*' indicates the common genes found in different VSD-associated CHD phenotypes. DORV, double outlet right ventricle; ASD, atrial septal defect; PA, pulmonary atresia; AS, aortic sclerosis; BAV, bicuspid aortic valve; OFT, outflow tract abnormality; PS, pulmonary stenosis; PDA, patent ductus arteriosus; TOF, tetralogy of fallot; TGA, transposition of great arteries; PTA, persistent truncus arteriosus; CoA, coarctation of aorta; PFO, patent foramen ovale; MAPCAs, major aorto-pulmonary collateral arteries

DLC1, EVC, HOMEZ, MESP1, MTRR, and *TRDN* were found to have more than eight variations. We have also observed the recurrence of nonsynonymous variations in seven genes, such as *BMP4, DLC1, HOXB1, MTHFD1, TBX18, TDGF1*, and *TRDN*, in more than 10 CHD subjects. In addition, we found nine genes with a total of 27 variations with nonframeshift insertions (2), frameshift insertions (3), nonframeshift deletions (6), and frameshift deletions (16) in our samples, indicating that these are more deleterious. If the frameshift variation occurs at the beginning region of the gene, then it affects the protein structure and function more severely than at the end of the gene. The effect of frameshift

Table 5 In silico analysis and pathogenicity score for the gene variations involved in VSD

Gene symbol	Amino acid change	dbSNP	1000Genomes	GnomAD	Single nucleotide variation effect prediction					Total Pathogenicity score (%)
					SIFT	PolyPhen2	MutationTaster	MutationAssessor	CADD	
CASZ1	p.L38P	–	–	–	D	PR	D	L	N	60
CITED2	p.S183L	rs1274837493	–	0.000007	T	B	B	L	NA	0
	p.S196G	rs962878530	–	0.000007	T	B	NA	NA	NA	0
	p.P101S	rs20163924	–	–	T	NA	B	L	NA	0
DLC1	P.D554V	rs773993191	–	0.000014	D	PS	D	M	NA	60
	p.L952V	rs184157214	0.000312	0.000016	D	PR	D	M	NA	80
	p.V1371L	–	–	–	D	B	NA	N	NA	20
	p.I1511M	rs78322853	–	0.000016	D	PR	D	M	NA	80
EPAS1	p.G724W	rs369874127	–	–	D	PS	B	M	N	40
EVC	p.L115V	rs755273705	–	0.000014	B	–	B	U	DM	20
	p.R576Q	rs1383180	0.299969	–	B	PR	B	U	DM	40
GATA4	p.A6V	rs199922907	0.000156	0.000007	D	PR	B	M	DM	80
	p.P163S	rs387906769	0.000468	0.000043	T	T	D	L	DM	40
	p.A176P	–	–	–	D	B	D	L	DM	60
	p.P193H	–	–	–	D	PS	B	L	N	20
	p.T225S	–	–	–	D	PR	D	M	DM	100
	p.I250N	rs1326483649	–	0.000004	D	PS	D	L	DM	60
	p.K256E	–	–	–	D	PS	D	M	DM	80
	p.A263G	–	–	–	D	B	D	N	DM	60
	p.G296R	rs104894073	–	–	D	PR	D	H	DM	100
	p.R320W	–	–	–	D	PR	D	M	DM	100
	p.E359K	rs368489876	–	0.000021	T	PR	D	M	DM	80
	p.E360G	–	–	–	T	PR	D	M	DM	80
	p.S377G	–	–	–	T	B	B	N	N	0
	p.P407Q	–			D	B	D	M	DM	80
	p.P408Q	rs115099192	0.000937	–	D	B	D	M	DM	80
	p.D425N	rs56208331	0.003592	0.001964	D	PS	B	M	DM	60
	p.S429T	–	–		D	PS	D	M	DM	80
	p.A442V	–	–	–	D	PR	D	M	DM	100
	p.A8D	rs864321698	–	–	D	NA	NA	NA	DM	40

Table 5 (continued)

Gene symbol	Amino acid change	dbSNP	1000Genomes	GnomAD	Single nucleotide variation effect prediction					Total Pathogenicity score (%)
					SIFT	PolyPhen2	MutationTaster	MutationAssessor	CADD	
GATA5	p.V190A	rs782051102	–	–	D	PR	B	N	N	40
	p.G96R	–	–	–	B	–	U	L	N	0
	p.H274R	–			D	PR	D	M	N	80
GATA6	p.G220S	–	–	–	T	B	D	N	N	20
GDF3	p.S212L	rs372790667	0.000312	0.0001	T	B	NA	N	NA	0
HAND1	p.G73S	rs756237060	–	0.000005	T	B	D	N	N	20
	p.K152N	–	–	–	D	B	B	M	N	40
HAND2	p.S65I	rs749318216	–	0.000021	T	PS	B	M	DM	40
HOMEZ	p.A39V	rs77990330	0.006402	0.005867	B	–	U	L	N	0
	p.S210R	–	–	–	D	–	B	U	N	20
HAS2	p.E499V	–	–	–	D	B	D	M	NA	60
HOXB1	p.Q103H	rs12939811	0.112586	–	D	B	B	L	NA	20
IL27	p.S59A	rs17855750	0.06574	–	B	B	U	L	N	0
IRX1	p.A381E	rs530506520	0.000156	0.000064	T	B	B	L	N	0
IRX4	p.N85Y	–	–	–	D	PS	D	L	NA	40
	p.E92G	–	–	–	D	PS	B	M	N	40
MESP1	p.D168G	rs200810210	0.009369	0.008563	D	PS	D	M	N	60
MTRR	p.I22M	rs1801394	0.358214	0.449468	D	PR	B	M	NA	60
	p.S175L	rs1532268	0.267645	0.316014	T	B	B	NA	N	0
MYBPC3	p.G507M	rs35736435	0.002967	0.001768	D	PR	D	M	NA	80
MYH6	p.P40R	rs775038703	–	–	T	PR	D	M	N	60
	p.Q277H	rs140660481	–	0.000285	D	B	D	M	DM	80
	p.R1608C	rs201683868	–	0.000036	D	PR	B	M	N	60
NKX2.5	p.E32V	–	–	–	T	B	NA	NA	NA	0
	p.D16N	rs750904275	–	0.000004	D	B	NA	NA	DM	40
NKX2.6	p.K152Q	–	–	–	D	PR	D	N	NA	60
NODAL	p.L61N	–	–	–	D	–	B	U	DM	40
NOTCH1	p.F2266S	–	–	–	T	PR	D	M	NA	60
NTRK3	p.T93M	rs147992979	0.000156	0.000045	T	PS	B	L	N	0
	p.N163I	rs547862658	0.00047	–	T	PS	B	L	DM	20
	p.I153F	rs869112057	–	–	D	PR	NA	M	N	60
PITX2	p.N153D	–	–	–	T	PS	D	M	DM	60

Table 5 (continued)

Gene symbol	Amino acid change	dbSNP	1000Genomes	GnomAD	Single nucleotide variation effect prediction					Total Pathogenicity score (%)
					SIFT	PolyPhen2	MutationTaster	MutationAssessor	CADD	
SALL4	p.R196W	rs151297824	–	0.000021	T	PS	B	M	N	20
	p.S797C	rs1357911800	–	–	D	PR	B	M	DM	80
TBX20	p.P14T	rs773397553	–	0.000021	D	PS	D	L	N	40
	p.I152M	rs137852954	–	0.000014	D	PR	D	M	DM	100
	p.R264K	rs201071418	0.000312	0.000043	D	PS	D	M	DM	80
	p.T331A	rs1420582400	–	0.000008	T	B	B	N	NA	0
TDGF1	p.R41G	rs745367998	–	0.000044	T	B	B	N	N	0
TRDN	p.S45I	rs748395950	–	0.000014	D	PR	D	M	N	80
TWIST1	p.G83S	rs545987863	0.000781	0.000364	T	B	B	N	NA	0
	p.S95G	rs575299986	0.000399	–	T	B	D	L	NA	20

The variations showing high pathogenicity are highlighted in the Table

SIFT score, Tolerated (T), Deleterious (D); PolyPhen2, Possibly damaging (PS), Benign (B), Probably damaging (PR); CADD Neutral (N), Damaging (DM); MutationAssessor, Low (L), Medium (M), High (H), and Neutral (N); Not available (NA); Uncertain (U)

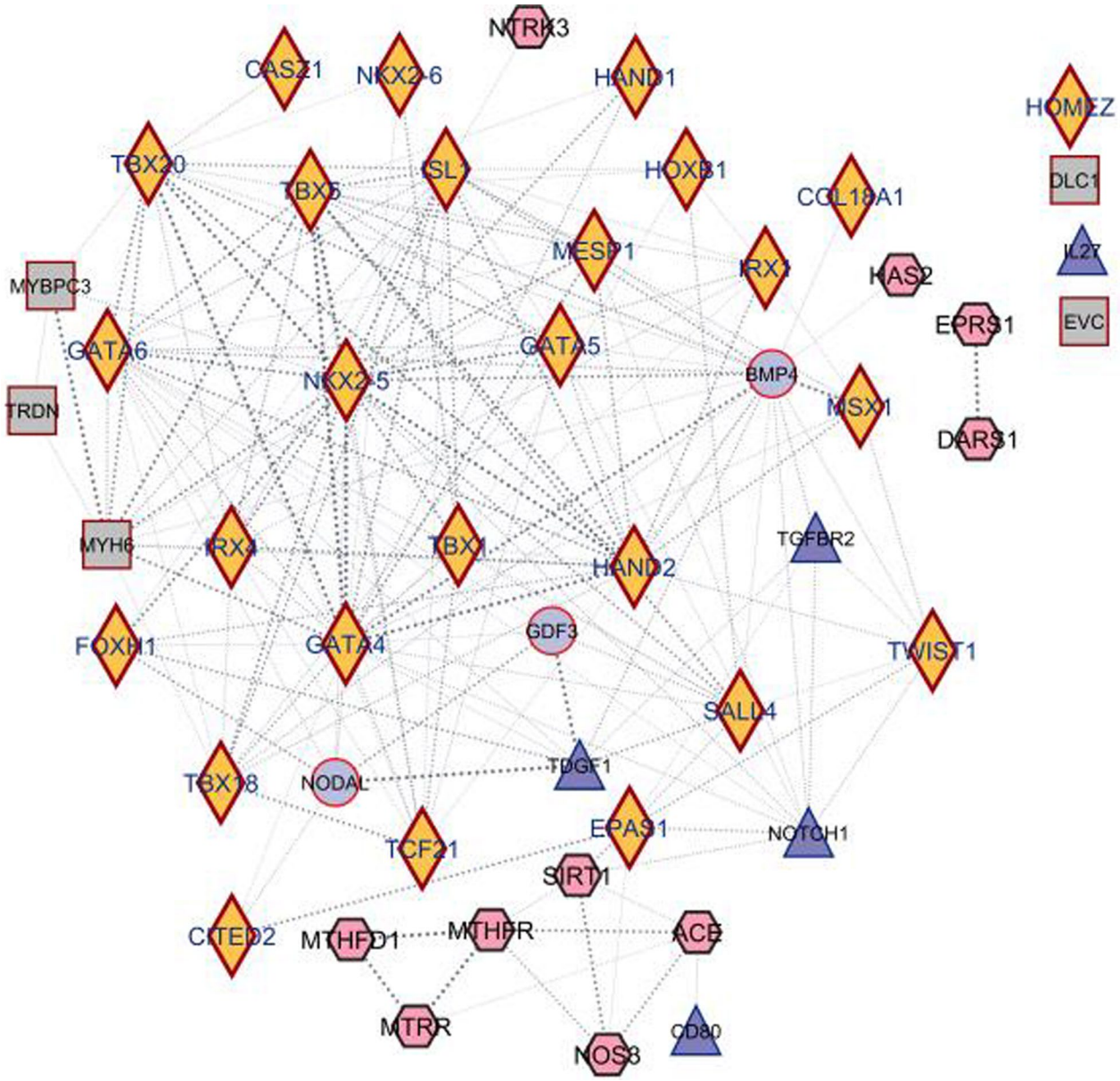

Fig. 4 The protein–protein interaction for 50 genes involved in the manifestation of VSD. This primary network contains 49 nodes with 175 edges. The proteins, based on their functionality, are indicated in different shapes. The transcription factors are shown in diamonds, receptors in triangles, ligands in circles, enzymes in hexagons, and other proteins in square shapes. The dotted edges with different width sizes indicate confidence interactions, whereas the darker edges indicate more confident interactions

variations in the disease manifestation is also greater when it is in a homozygous condition.

In the present study, we found more genes coding for the transcription factors which are involved in the manifestation of nonsyndromic VSD. *NKX*, *GATA*, and *TBX* family of genes are the most studied genes in CHD. Any disruption in the physical interactions between the transcription factors or with other co-factors can impair the transcriptional cooperativity and lineage specification, ultimately resulting in malformation of cardiac formation and function [22, 23].

These genes are studied in vitro and in vivo to unravel their impact on different types of CHD. The mouse models have been used to validate the pathogenic sequence variants in the heart development genes such as *TBX5*, *NKX2.5*, and *GATA4*, which are involved in human septation defects [1, 24]. The genes reported in nonsyndromic CHD, such as *GATA4, GATA6, MYH6, NKX2.5, MYH6*,

and *TBX20*, are studied using Zebrafish models to unravel the cardiac phenotypes association [24]. The *Drosophila* model systems also check the gene–gene interactions during heart development. One example using the *Drosophila* model in vitro study to understand the impact of genomic variation with unknown significance in *NKX2.5* showed a strongly reduced ability to interact with DNA and reduced trans-activation ability [25]. Similarly, subsequent studies have shown that GATA4 and TBX5 cooperatively interact on DNA throughout the genome to regulate the development of the heart [22, 26].

The present study found HOMEZ in 16 CHD subjects, including 14 VSD and 2 ToF. The p.A302T, p.R502Q, and p.E533K are the three nonsynonymous variations newly identified in this study. A genome-wide linkage analysis was conducted by McGregor et al. (2010) [27] in 84 south Indian probands having unaffected consanguineous parents, and this study implicated the linkage of CHD with rs1055061, p.R502Q of *HOMEZ*. Another study reported two variations, p.A39V and p.S210R, in *HOMEZ*, and the position of these variations is found conserved among many species (humans, rats, mice, etc.), which suggested that they might play an important role in maintaining the protein function [28].

Besides, in the present study, the *MTRR* and *DLC1* are found most recurrently in 18 CHD subjects with 20 variations in each gene. The MTRR is required to maintain the MTR in an active state, a central regulatory enzyme in the metabolic pathway of homocysteine/folate. The disrupted function of MTR may be associated with hyperhomocysteinemia, which causes endothelial damage [29]. Earlier studies have reported that c.66A > G and c.524C > T variations in *MTRR* are not associated with an increased risk of CHDs [29, 30]. In contrast, there are associations between c.66A > G and c.524C > T alleles and their effect on the risk of CHD in Chinese and Iranian populations [31, 32]. Similarly, in the current study, we observed c.524C > T in *MTRR* in nine CHD subjects, including six VSD and three ToF, and c.66A > G in eight VSD subjects. Interestingly, in 9 CHD subjects, the variant, c.524C > T in *MTRR*, resulted in the change of amino acid serine to leucine at position 175, suggesting the change of the functionality of this protein.

A study in mice (*Dlc1*-/-) showed embryonic lethal, and histologically, the heart was found incompletely developed with the distorted structure of the heart chambers [33]. Another knockout mouse *DLC1* study demonstrated the abnormalities in the blood vasculature of the yolk sac and heart development [34]. Similarly, we found another recurrent gene, *DLC1*, in 18 CHD subjects, including 15 VSD and three ToF. Hence, these observations indicate that both *MTRR* and *DLC1* are important in the early development of the heart.

Additionally, p.D308E in *EPRS* was the recurrent variation in 14 CHD subjects, including 11 VSD and 3 ToF. This variation was also identified in the previous case–control study on the Chinese population, which included 984 cases and 2953 controls [35]. This study has conferred the susceptibility of this variation to the risk of CHD. This variation is located in exon 23 of *EPRS*, which acts as part of the exonic splicing enhancer according to the SNPinfo tool [36].

One of the interesting observations is that about 59% of the identified genes for VSD in 24 CHD subjects were also found in all three ToF subjects, indicating that VSD-specific genes and variations are common in other septal defects. Similarly, we also found that *MTRR* was common between isolated VSD, ToF, and VSD-associated ASD subjects. Although these findings are encouraging to establish the VSD-specific genes with variations, since the sample size is small, it is premature to make this conclusion.

The in silico analysis of nonsynonymous variations in 32 genes predicted the damaging effect of p.T225S, p.G296R, p.R320W, and p.A442V in *GATA4*; p.I152M in *TBX20* on protein function. As previously described, the GATA and TBX are the major transcription factors that play a vital role in the pathogenesis of CHD. The protein–protein network analysis showed that the GATA, NKX, TBX, and HAND family of proteins interact to control other proteins required for heart septal formation. Among these, most of them are transcription factors, including the zinc finger proteins GATA4, GATA5, and GATA6; homeodomain protein NKX2.5, NKX2.6; and T-box transcription factors TBX1, TBX5, TBX18, and TBX20, and basic helix-loop-helix HAND1 and HAND2 proteins. These major transcription factors have been identified for VSD-specific CHD. Further, the increased nodes in the interaction network analysis identified the other interacting proteins, such as TNNT2, MYL7, and ZFPM2, with functionality as a structural protein, motor protein, and transcription factor, respectively. The pathogenic variation in TNNT2 is reported in primary cardiomyopathy [37]. Another study has shown that the altered expression of *TBX3* and *NOS3* reduces TNNT2 in the developing heart, which affects normal development by disrupting the molecular signals [38]. The *MYL7* is a cardiac development gene that helps in cardiac sarcomere assembly and function. The function of *MYL17* is associated with *SOX9B* expression. A study has shown that the expression of *myl7* has significantly decreased by the cardiomyocyte-specific inhibition of *sox9b* [39]. It indicates that the disruption in *MYL7* could play a substantial role in cardiac malformation. Similarly, the variation p.V763I in *ZFPM2* has been reported previously in CHD, including ToF, interrupted aortic arch, and hypoplastic

left syndrome [40], and the variation p.P87L is reported in a CHD subject with VSD, pulmonary stenosis (PS), and transposition of great vessels (TGV) [10]. Therefore, TNNT2, MYL7, and ZFPM2 show their role in the manifestation of CHD.

Further, the in vivo validations of the variations identified in the present study help better understand the etiology of CHD.

Conclusion

The present study identified 37 potential risk genes through exome sequence analysis and interaction network analysis, with strong evidence from previously reported studies. From these genes, 176 variations were identified, of which 166 were newly reported from our samples, and 10 were previously reported, indicating that these are the novel variants for the Indian population. Further, these variations should be subjected to in vivo and in vitro validations before considering to create a gene panel for CHD.

Acknowledgements
We thank ICMR, the senior research fellowship, DAE, Raja Ramanna, the Chair grant, the Department of Studies in Genetics and Genomics, and the University of Mysore for their help conducting this work. We also thank Dr. Gholipoorfeshkecheh Rahim, Dr. G.K. Chetan, and Dr. Gautham Arunachal Udupi for their support.

Author contributions
Conceptualization was done by Nallur Basappa Ramachandra and Chaithra Shanthithadda; Data curation was performed by Chaithra Shanthithadda; Writing–original draft was carried out by Chaithra Shanthithadda; Writing–review and editing was carried out by Nallur Basappa Ramachandra.

Funding
None.

Availability of data and materials
Not applicable.

Declarations

Ethics approval and consent to participate
The ethical approval was obtained from Institutional Human Ethical Committee, University of Mysore (IHEC-UOM No.77/Res/2021–22) to collect Human blood samples.

Consent for publication
Not applicable.

Competing interest
The authors declare that they have no competing interests.

Received: 15 August 2024 Accepted: 28 March 2025
Published online: 20 April 2025

References

1. Zhu W, Lo CW (2023) Insights into the genetic architecture of congenital heart disease from animal modeling. Zool Res 44(3):577–590. https://doi.org/10.24272/j.issn.2095-8137.2022.463
2. Morton SU, Quiat D, Seidman JG, Seidman CE (2022) Genomic frontiers in congenital heart disease. Nat Rev Cardiol. https://doi.org/10.1038/s41569-021-00587-4
3. Chaithra S, Agarwala S, Ramachandra NB (2022) High-risk genes involved in common septal defects of congenital heart disease. Gene. https://doi.org/10.1016/j.gene.2022.146745
4. Chung IM, Rajakumar G (2016) Genetics of congenital heart defects: the NKX2-5 gene, a key player. Genes. https://doi.org/10.3390/genes7020006
5. Chai H, Yan Z, Huang K, Jiang Y, Zhang L (2018) MicroRNA expression, target genes, and signaling pathways in infants with a ventricular septal defect. Mol Cell Biochem. https://doi.org/10.1007/s11010-017-3146-2
6. An Y, Duan W, Huang G, Chen X, Li L, Nie C et al (2015) Genome-wide copy number variant analysis for congenital ventricular septal defects in Chinese Han population. BMC Med Genomics. https://doi.org/10.1186/s12920-015-0163-4
7. Gholipoorfeshkecheh R, Agarwala S, Krishnappa S, Savitha MR, Ramachandra NB (2020) Whole-exome sequencing and homozygosity mapping identify variants in NCOR1 and MAP2K3 associated with non syndromic congenital heart defects. Egyptian J Med Human Gene. https://doi.org/10.15252/emmm.201809127
8. Bhardwaj R, Rai SK, Yadav AK, Lakhotia S, Agrawal D, Kumar A, Mohapatra B (2015) Epidemiology of congenital heart disease in India. Congenit Heart Dis. https://doi.org/10.1111/chd.12220
9. Nees SN, Chung WK (2020) Genetic basis of human congenital heart disease. Cold Spring Harb Perspect Biol. https://doi.org/10.1101/cshperspect.a036749
10. Pulignani S, Vecoli C, Borghini A, Foffa I, Ait-Alì L, Andreassi MG (2018) Targeted next-generation sequencing in patients with nonsyndromic congenital heart disease. Pediatr Cardiol. https://doi.org/10.1007/s00246-018-1806-y
11. Yang L, Yang Y, Liu X, Chen Y, Chen Y, Lin Y et al (2020) CHDGKB: a knowledgebase for systematic understanding of genetic variations associated with nonsyndromic congenital heart disease. Database. https://doi.org/10.1093/database/baaa048
12 Zhou WZ, Li W, Shen H, Wang RW, Chen W, Zhang Y et al (2023) CHDbase: a comprehensive knowledgebase for congenital heart disease-related genes and clinical manifestations. Genomics, Proteomics Bioinform. https://doi.org/10.1016/j.gpb.2022.08.001
13. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC (2012) SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. https://doi.org/10.1093/nar/gks539
14. Schwarz JM, Cooper DN, Schuelke M, Seelow D (2014) MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods. https://doi.org/10.1038/nmeth.2890
15. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P et al (2010) A method and server for predicting damaging missense mutations. Nat Methods. https://doi.org/10.1038/nmeth0410-248
16. Reva B, Antipin Y, Sander C (2011) Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. https://doi.org/10.1093/nar/gkr407
17. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. https://doi.org/10.1038/ng.2892
18. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R et al (2023) The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. https://doi.org/10.1093/nar/gkac1000
19. Bravo-Valenzuela NJ, Peixoto AB, Júnior EA (2018) Prenatal diagnosis of congenital heart disease: a review of current knowledge. Indian Heart J. https://doi.org/10.1016/j.ihj.2017.12.005
20. Saxena A, Relan J, Agarwal R, Awasthy N, Azad S, Chakrabarty M et al (2019) Indian guidelines for indications and timing of intervention for common congenital heart diseases: revised and updated consensus statement of the working group on management of congenital heart diseases. Ann Pediatr Cardiol. https://doi.org/10.4103/apc.APC_32_19
21. Li Y, Klena NT, Gabriel GC, Liu X, Kim AJ, Lemke K et al (2015) Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature. https://doi.org/10.1038/nature14269
22. Ang YS, Rivas RN, Ribeiro AJ, Srivas R, Rivera J, Stone NR et al (2016) Disease model of GATA4 mutation reveals transcription factor cooperativity in human cardiogenesis. Cell. https://doi.org/10.1016/j.cell.2016.11.033

23. Waldron L, Steimle JD, Greco TM, Gomez NC, Dorr KM, Kweon J et al (2016) The cardiac TBX5 interactome reveals a chromatin remodeling network essential for cardiac septation. Dev Cell. https://doi.org/10.1016/j.devcel.2016.01.009
24. Majumdar U, Yasuhara J, Garg V (2021) In vivo and in vitro genetic models of congenital heart disease. Cold Spring Harb Perspect Biol. https://doi.org/10.1101/cshperspect.a036764
25 Lovato TL, Blotz B, Bileckyj C, Johnston CA, Cripps RM (2023) Modeling a variant of unknown significance in the Drosophila ortholog of the human cardiogenic gene NKX2.5. Disease Models Mech. https://doi.org/10.1242/dmm.050059
26. Luna-Zurita L, Stirnimann CU, Glatt S, Kaynak BL, Thomas S, Baudin F et al (2016) Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. Cell. https://doi.org/10.1016/j.cell.2016.01.004
27. McGregor TL, Misri A, Bartlett J, Orabona G, Friedman RD, Sexton D et al (2010) Consanguinity mapping of congenital heart disease in a South Indian population. PLoS ONE. https://doi.org/10.1371/journal.pone.0010286
28. Xuan C, Jia KG, Wang BB, Bai XY, Gao G, Yang Q et al (2013) Identification of two novel mutations of the HOMEZ gene in Chinese patients with isolated ventricular septal defect. Genet Test Mol Biomarkers. https://doi.org/10.1089/gtmb.2012.0435
29. Verkleij-Hagoort AC, van Driel LM, Lindemans J, Isaacs A, Steegers EA, Helbing WA et al (2008) Genetic and lifestyle factors related to the periconception vitamin B12 status and congenital heart defects: a Dutch case-control study. Mol Genet Metab. https://doi.org/10.1016/j.ymgme.2007.12.002
30. Shaw GM, Lu W, Zhu H, Yang W, Briggs FB, Carmichael SL et al (2009) 118 SNPs of folate-related genes and risks of spina bifida and conotruncal heart defects. BMC Med Genet. https://doi.org/10.1186/1471-2350-10-49
31. Su J, Li Z (2018) Analysis of MTR and MTRR gene polymorphisms in Chinese patients with ventricular septal defect. Appl Immunohistochem Mol Morphol. https://doi.org/10.1097/PAI.0000000000000512
32. Pishva SR, Vasudevan R, Etemad A, Heidari F, Komara M, Ismail P et al (2013) Analysis of MTHFR and MTRR gene polymorphisms in Iranian ventricular septal defect subjects. Int J Mol Sci. https://doi.org/10.3390/ijms14022739
33. Durkin ME, Avner MR, Huh CG, Yuan BZ, Thorgeirsson SS, Popescu NC (2005) DLC-1, a Rho GTPase activating protein with tumor suppressor function, is essential for embryonic development. FEBS Lett. https://doi.org/10.1016/j.febslet.2004.12.090
34. Sabbir MG, Wigle N, Loewen S, Gu Y, Buse C, Hicks GG, Mowat MR (2010) Identification and characterization of Dlc1 isoforms in the mouse and study of the biological function of a single gene trapped isoform. BMC Biol. https://doi.org/10.1186/1741-7007-8-17
35. Da M, Feng Y, Xu J, Hu Y, Lin Y, Ni B et al (2014) Association of aminoacyl-tRNA synthetases gene polymorphisms with the risk of congenital heart disease in the Chinese Han population. PLoS ONE. https://doi.org/10.1371/journal.pone.0110072
36. Xu Z, Taylor JA (2009) SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. Nucleic Acids Res. https://doi.org/10.1093/nar/gkp290
37. Seidel F, Holtgrewe M, Al-Wakeel-Marquard N, Opgen-Rhein B, Dartsch J, Herbst C et al (2021) Pathogenic variants associated with dilated cardiomyopathy predict outcome in pediatric myocarditis. Circ: Genomic Precision Med. https://doi.org/10.1161/CIRCGEN.120.003250
38. England J, Pang KL, Parnall M, Haig MI, Loughna S (2016) Cardiac troponin T is necessary for normal development in the embryonic chick heart. J Anat. https://doi.org/10.1111/joa.12486
39. Gawdzik JC, Yue MS, Martin NR, Elemans LM, Lanham KA, Heideman W et al (2018) sox9b is required in cardiomyocytes for cardiac morphogenesis and function. Sci Rep. https://doi.org/10.1038/s41598-018-32125-7
40. Blue GM, Kirk EP, Giannoulatou E, Dunwoodie SL, Ho JW, Hilton DC et al (2014) Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. J Am Coll Cardiol. https://doi.org/10.1016/j.jacc.2014.09.048

Publisher's Note