Original Article

In silico analysis of GATA4 variants demonstrates main contribution to congenital heart disease

Shiva Abbasi1*, Neda Mohsen-Pour1*, Niloofar Naderi2, Shahin Rahimi3, Majid Maleki4, Samira Kalayinia5

1 Cardiogenetic Research Center, Rajaie Cardiovascular Medical and Research Center, Iran University of Medical Sciences, Tehran, Iran
2 Zanjan Pharmaceutical Biotechnology Research Center, Zanjan University of Medical Sciences, Zanjan, Iran
3 Department of Cardiology, Rajaie Cardiovascular Medical and Research Centre, Iran University of Medical Sciences, Tehran, Iran

Abstract
Introduction: Congenital heart disease (CHD) is the most common congenital abnormality and the main cause of infant mortality worldwide. Some of the mutations that occur in the GATA4 gene region may result in different types of CHD. Here, we report our in silico analysis of gene variants to determine the effects of the GATA4 gene on the development of CHD.

Methods: Online 1000 Genomes Project, ExAC, gnomAD, GO-ESP, TOPMed, Iranome, GME, ClinVar, and HGMD databases were drawn upon to collect information on all the reported GATA4 variations. The functional importance of the genetic variants was assessed by using SIFT, MutationTaster, CADD, PolyPhen-2, PROVEAN, and GERP prediction tools. Thereafter, network analysis of the GATA4 protein via STRING, normal/mutant protein structure prediction via HOPE and I-TASSER, and phylogenetic assessment of the GATA4 sequence alignment via ClustalW were performed.

Results: The most frequent variant was c.874T > C (45.58%), which was reported in Germany. Ventricular septal defect was the most frequent type of CHD. Out of all the reported variants of GATA4, 38 variants were pathogenic. A high level of pathogenicity was shown for p.Gly221Arg (CADD score = 31), which was further analyzed.

Conclusion: The GATA4 gene plays a significant role in CHD; we, therefore, suggest that it be accorded priority in CHD genetic screening.

Introduction
Congenital heart disease (CHD) is the most common congenital malformation and a significant cause of childhood mortality with an estimated prevalence of 1% of infants born each year.1,2 Cardiovascular abnormalities are reported in approximately 29% of dead infants. CHD can be caused by variants in different genes whose roles have evolved. The number of genes and variants thereof involved in the CHD pathogenesis has increased, and an accurate determination of the molecular mechanisms of CHD remains particularly challenging due to genetic heterogeneity and incomplete penetrance.3 Also extremely complex is the differential diagnosis of CHD in that it is a multifactorial disease encompassing both genetic predisposition and environmental components.4 Thus, it is vitally important to identify disease-causing genetic variants.5 Some CHD-associated genes encode transcription factors such as GATA4, NKX2-5, and TBX5, and a number of gene variants identified in these genes have been associated with cardiac structure and functional impairment.6 GATA-binding factor 4 (GATA4) (OMIM: 600576) is one of the 6-member GATA family of transcription factors: GATA1, GATA2, GATA3, GATA4, GATA5, and GATA6. Amongst GATA-binding proteins, GATA1–3 are expressed in hematopoietic stem cells as significant regulators, whereas GATA4–6 are expressed in different mesoderm- and endoderm-derived tissues such as the heart, the lung, the gonad, the gut, and the liver.6 Variants in the GATA4, GATA5, and GATA6 genes have been found in patients with various types of CHD.7-9 GATA proteins comprise 2 conserved zinc finger domains (ZNI and ZNII), which cover various aspects of functions including DNA attachment, GATA4 preservation, and protein-protein and the target DNA sequence interactions. The GATA4 gene consists of 7 exons located on chromosome 8p23.1-p22. The gene encodes one of the earliest-expressed transcription factors with 442 amino acids and is imperative for normal cardiogenesis. GATA4 is significantly expressed in embryonic development, with the expression continuing in the adult myocardium.10-12 A rise has been reported in the number of patients with CHD who reach adulthood.13 This transcription factor...
contains 2 transcriptional activation domains (TAD1 and TAD2); 2 zinc finger domains: 1 at the c-terminal region (CZF) and the other at the n-terminal region (NZF); and 1 nuclear localization signal domain (NLS). Variants in the GATA4 gene are highly associated with different types of CHD, including tetralogy of Fallot, ventricular septal defect, atrial septal defect, atrioventricular septal defect, patent ductus arteriosus, dilated cardiomyopathy, and pulmonary valve stenosis. The current literature lacks in silico analysis on the variants of the GATA4 transcription factor and their critical role in the different levels of cardiovascular development. Accordingly, for the first time, we aimed to conduct a comprehensive in silico analysis of the effects of GATA4 alterations associated with CHD.

Materials and Methods
For the detection of genetic variants in the GATA4 gene, the following methodology was utilized in the present study:

Data Collection
The amino acid sequence of the human GATA4 gene was obtained from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/), based on the human genome assembly GRCh37. Accordingly, the Human Gene Mutation Database (HGMD; http://www.hgmd.cf.ac.uk/ac/index.php), as a strongly reliable database, was employed to identify alterations in the GATA4 gene. Concurrently, all pathogenic/likely pathogenic alterations reported in public access databases were identified. The databases were ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), dbSNP (the NCBI database of genetic variation; https://www.ncbi.nlm.nih.gov/snp/), GeneCards (the human gene database; https://www.genecards.org/), ExAC (the exome aggregation consortium; http://exac.broadinstitute.org/), the 1000 Genomes Project (https://www.internationalgenome.org/), gnomAD (the Genome Aggregation Database; http://gnomad.broadinstitute.org/), GO-ESP (NHBLI “Grand Opportunity” Exome Sequencing Project; http://evs.gs.washington.edu/EVS/), TOPMed (Trans-Omics for Precision Medicine; https://www.nhlbiwgs.org/), Iranome (http://www.iranome.ir/), and the Greater Middle East (GME) Variome Project (http://igm.ucsd.edu/gme/). Moreover, extensive research was carried out through computerized search of PubMed, Scopus, Google Scholar, ScienceDirect, MalaCards (the human disease database), and ResearchGate databases by using the following terms: GATA4 variants, the clinical importance of the GATA4 gene, GATA4-related disorders, CHD, the pathophysiology of CHD, and the incidence of CHD.

Frequency
The frequencies of the selected variants were determined using the aforementioned databases. Furthermore, the number of participants and individuals having variations in the studied populations was reported.

Computational Methods
Given its increasing importance and use to determine the possible effects of genetic variants, computational analysis was employed in the present study. The variants of the GATA4 gene and their correlations with the molecular pathogenesis of CHD were further explored by predicting the pathogenicity/tolerance of the variants through the following bioinformatics tools: SIFT (Sorting Intolerant from Tolerant; https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html), PolyPhen-2 (Polymorphism Phenotyping, version 2; http://genetics.bwh.harvard.edu/pph2), PROVEAN (Protein Variation Effect Analyzer, version 1.1.3; http://provean.jcvi.org/seq_submit.php), CADD (Combined Annotation-Dependent Depletion; https://cadd.gs.washington.edu/), MutationTaster (http://www.mutationtaster.org/), and GERP (Genomic Evolutionary Rate Profiling; http://mendel.stanford.edu/SidowLab/downloads/gerp/). All these bioinformatics tools are capable of distinguishing pathogenic from nonpathogenic alterations. Protein sequences in the FASTA format (NM_002052.5), the positions and substitutions of amino acids, and the positions of chromosomes were used as input data. A SIFT score of 0.05 or less is regarded as deleterious, and a SIFT score of greater than 0.05 is considered to signify a tolerated variant. PolyPhen-2 results are shown with qualitative levels as benign, possibly damaging, and probably damaging. PolyPhen-2 prediction outputs have a numerical score range of 0 to 1. The cutoff score considered for PolyPhen-2 is 0.5, and variants with scores equal to or greater than 0.5 are predicted to be deleterious. The cutoff score for PROVEAN is ~2.5, and variants equal to or greater than ~2.5 are assigned as deleterious. Also calculated in the current investigation was the CADD score. All genomic features used to calculate the CADD score via a machine-learning model are summarized into a Phred score with a cutoff point of 20. Disease-causing variants display a high Phred score (>20), whereas a low score (<20) signifies less pathogenicity. MutationTaster, which was applied for all the detected variants in the present study, considers an alteration to be a polymorphism if it is reported as a single-nucleotide polymorphism (SNP) in the HapMap data and the 1000 Genomes Project. Thus, any alteration that could result in premature termination codon and ultimately lead to nonsense-mediated mRNA decay is considered a disease-causing variant. GERP is an evolutionary measurement tool whose results are based on multi-species sequence alignment by comparison with neutral expectation. GERP scores show a reduction in the number of substitutions. Positive scores indicate a substitution deficit, while negative scores show that a site is probably evolving neutrally.
GATA4 Network Analysis

The functional association between 2 proteins is the primary purpose of the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database. This web-based tool expresses the interaction of proteins in a particular biological function. 41 STRING (version 11.0; https://string-db.org/) is used to recognize the known and predicted interactions between the GATA4 protein and other related proteins in a cell. 42

Prediction of Normal and Mutant Protein Structures

Structural and functional differences between wild-type and mutated GATA4 were anticipated by using HOPE (Have [y]Our Protein Explained; https://www3.cmbi.umcn.nl/hope/input/) and 43 I-TASSER (Iterative Threading ASSEmbly Refinement; https://zhanglab.ccmb.med.umich.edu/I-TASSER/). The objective was to analyze a pathogenic variant with a high CADD score. HOPE shows the 3D structural and functional effects of a point mutation in human proteins. The input for this tool is the amino acid sequence of the GATA4 protein and the specific amino acid alteration of the variant. 43 The I-TASSER server predicts secondary structures and 3D models through various alignment methods. The accuracy of the formed models is evaluated based on a confidence score (C-score). Predicted models with a C-score of greater than −1.5 are considered to possess a correct topology. I-TASSER predicts the template modeling score (Tm-score) and the root mean square deviation (RMSD). The TM-score ranges between 0 and 1, with higher values specifying better structural models. 47

Phylogenetic Analysis

GATA4 protein sequences from 5 different organisms, namely Homo sapiens (humans), Canis lupus familiaris (dogs), Rattus norvegicus (rats), Gallus gallus domesticus (chickens), and Xenopus laevis (African clawed frogs), were retrieved from UniProt (the Universal Protein Resource; https://www.uniprot.org/). Afterward, all the GATA4 protein sequences were aligned via the multiple sequence alignment program ClustalW (version 1.83; https://www.genome.jp/tools-bin/clustalw). Thereafter, a phylogenetic tree was built by using ClustalW via the neighbor-joining method. As a result of the multiple sequence alignment, the tree showed scores that represented a sequence distance measure. These values determine the length of the branches, with the length showing the distance between the sequences.

Results

Literature Analysis

Using online databases and publications, we succeeded in finding 110 reported variations in the GATA4 gene. We also determined the frequency of the gene variants from online resources. The data are depicted in Table 1. The distributions of the reported variants in the different regions of the GATA4 gene are presented in Figure 1.

Frequency of the Variants

A wide range of GATA4 variants has been reported in different countries such as Japan, Australia, the United States, Brazil, Egypt, India, Germany, Lebanon, France, Iran, Italy, and especially China. Precise data on the reported variations and the phenotype condition of the

---

Figure 1. The image shows the distribution of the variations of the GATA4 gene. The figure demonstrates the 3 main parts of the GATA4 protein, the GATA-type transcription activator, and the zinc finger regions. The numbers beneath the protein structure represent the number of amino acids in the GATA4 amino acid sequence. The reported variations are categorized based on their locations.
Table 1. Reported frequency of the variants in online databases

| DNA Change   | Genomic Placement on Chromosome 8 | 1000Genome | ExAC     | GnomAD   | GO-ESP   | TOPMED   | Iranome   | GME      |
|--------------|-----------------------------------|------------|----------|----------|----------|----------|-----------|----------|
| c.17C>T      | 11565838                          | 0.0002     | 0.00002  | 0.00003  | -        | -        | -         | -        |
| c.46G>T      | 1156586                           | 0.0002     | -        | 0.00003  | -        | 0.00016  | -         | -        |
| c.62G>T      | 11565883                          | 0.0002     | 0.00004  | 0.00002588 | 0.000024 | -        | -         | -        |
| c.82C>G      | 11565903                          | -          | -        | -        | -        | 0.00008  | -         | -        |
| c.82C>T      | 11565903                          | -          | -        | -        | -        | 0.00008  | -         | -        |
| c.106C>T     | 11565927                          | -          | -        | -        | -        | -         | -         | -        |
| c.112T>G     | 11565933                          | -          | -        | -        | -        | -         | -         | -        |
| c.115G>T     | 11565936                          | -          | 0.000007541 | 0.000007541 | -      | -        | -         | -        |
| c.127C>T     | 11565948                          | -          | 0.00006  | 0.0000056 | -        | -         | -         | -        |
| c.136-138delTCC | 11565957                       | -          | -        | -        | -        | -         | -         | -        |
| c.151C>G     | 11565972                          | -          | -        | -        | -        | -         | -         | -        |
| c.155C>T     | 11565976                          | -          | -        | -        | -        | -         | -         | -        |
| c.164A>G     | 11565985                          | -          | -        | -        | -        | -         | -         | -        |
| c.191G>A     | 11566012                          | -          | -        | -        | -        | 0.00008  | -         | -        |
| c.196G>A     | 11566017                          | -          | -        | -        | -        | 0.00032  | -         | -        |
| c.206G>A     | 11566027                          | -          | -        | -        | -        | -         | -         | -        |
| c.209G>C     | 11566030                          | -          | -        | -        | -        | -         | -         | -        |
| c.221C>A     | 11566042                          | -          | 0.0003625 | 0.0003625 | -      | 0.00032  | -         | -        |
| c.244A>G     | 11566065                          | -          | 0.00006  | 0.000032 | -        | -         | -         | -        |
| c.259C>T     | 11566080                          | -          | -        | -        | -        | -         | -         | -        |
| c.270C>A     | 11566091                          | -          | -        | -        | -        | -         | -         | -        |
| c.278G>C     | 11566099                          | -          | -        | -        | -        | 0.00016  | -         | -        |
| c.284A>G     | 11566105                          | -          | -        | -        | -        | -         | -         | -        |
| c.286G>A     | 11566107                          | -          | -        | -        | -        | -         | -         | -        |
| c.307C>G     | 11566128                          | -          | -        | -        | -        | -         | -         | -        |
| c.357_359GCC | 11566175-11566176                | -          | -        | -        | -        | -         | -         | -        |
| c.392C>G     | 11566213                          | -          | 0.00003  | 0.000303 | -        | -         | -         | -        |
| c.431C>T     | 11566252                          | -          | 0.0001441 | 0.000016 | -        | -         | -         | -        |
| c.448G>T     | 11566269                          | -          | -        | -        | -        | -         | -         | -        |
| c.479G>C     | 11566300                          | -          | -        | -        | -        | 0.00008  | -         | -        |
| c.487C>A     | 11566308                          | 0.0002     | 0.0002   | 0.00003  | 0.000175 | -         | -         | -        |
| c.488C>G     | 11566309                          | 0.0002     | 0.0003   | -        | 0.00008  | -         | -         | -        |
| c.569A>G     | 11566390                          | -          | -        | -        | -        | -         | -         | -        |
| c.578C>A     | 11566399                          | -          | -        | -        | -        | -         | -         | -        |
| c.590A>G     | 11566411                          | -          | -        | -        | -        | -         | -         | -        |
| c.620C>T     | 11567128                          | 0.2017     | 0.16213  | 0.174583 | -        | -         | -         | -        |
| c.622T>C     | 11606433                          | -          | -        | -        | -        | -         | -         | -        |
| c.628G>A     | 11606439                          | -          | 0.00003  | -        | 0.00008  | 0.00008  | -         | -        |
| c.631T>C     | 11606442                          | -          | -        | -        | -        | -         | -         | -        |
| c.640G>A     | 11606451                          | -          | -        | -        | -        | -         | -         | -        |
| c.640G>A     | 11606451                          | -          | -        | -        | -        | -         | -         | -        |
| c.648C>G     | 11606459                          | -          | -        | -        | -        | -         | -         | -        |
| c.661G>A     | 11606472                          | -          | -        | -        | -        | -         | -         | -        |
| c.668T>C     | 11606479                          | -          | -        | -        | -        | -         | -         | -        |
| c.677C>A     | 11606488                          | -          | -        | -        | -        | -         | -         | -        |
| c.687C>T     | 11606498                          | -          | -        | -        | -        | -         | -         | -        |
| SNP   | Allele 1 | Allele 2 | p-value 1 | p-value 2 | p-value 3 | p-value 4 | p-value 5 |
|-------|----------|----------|-----------|-----------|-----------|-----------|-----------|
| c.700G>A | 11606511 | -        | -         | -         | -         | -         | -         |
| c.715A>G | 11606526 | -        | -         | -         | -         | -         | -         |
| c.716A>G | 11606527 | -        | -         | -         | -         | -         | -         |
| c.731A>G | 11606542 | -        | -         | -         | -         | -         | -         |
| c.740T>C | 11606551 | -        | -         | -         | -         | -         | -         |
| c.743A>G | 11606554 | -        | 0.000016  | 0.000008  | -         | -         | -         |
| c.749T>A | 11606560 | -        | -         | -         | -         | -         | -         |
| c.754C>T | 11606565 | -        | -         | -         | -         | -         | -         |
| c.764T>C | 11606575 | -        | -         | -         | -         | -         | -         |
| c.779G>A | 11606590 | -        | 0.00003   | 0.000008  | -         | -         | -         |
| c.782T>C | 11606593 | -        | -         | -         | -         | -         | -         |
| c.783T>G | 11606594 | -        | -         | -         | -         | -         | -         |
| c.788C>G | 11606575 | -        | -         | -         | -         | -         | -         |
| c.796C>T | 11606590 | -        | -         | -         | -         | -         | -         |
| c.799G>A | 11607635 | 0.0004   | 0.000306  | 0.00029   | -         | 0.000231 | -         |
| c.812G>C | 11607648 | -        | -         | -         | -         | -         | -         |
| c.818A>G | 11607654 | -        | -         | 0.00003   | -         | -         | -         |
| c.819C>A | 11607655 | -        | -         | -         | -         | -         | -         |
| c.822C>T | 11607658 | 0.0018   | 0.002243  | 0.00258   | 0.00308   | 0.003297 | -         |
| c.830G>T | 11607666 | -        | 0.000004  | -         | 0.000008  | -         | -         |
| c.851G>A | 11607687 | -        | -         | -         | -         | -         | -         |
| c.854A>G | 11607690 | -        | -         | -         | -         | -         | -         |
| c.855T>C | 11607691 | -        | -         | -         | -         | -         | -         |
| c.871G>C | 11607707 | -        | -         | -         | -         | -         | -         |
| c.874T>C | 11607710 | -        | -         | -         | -         | -         | -         |
| c.881C>T | 11607717 | -        | -         | -         | -         | -         | -         |
| c.886G>C | 11607722 | -        | -         | -         | -         | -         | -         |
| c.886G>A | 11607722 | -        | -         | -         | -         | -         | -         |
| c.899A>C | 11607735 | -        | -         | -         | -         | -         | -         |
| c.905A>G | 11607741 | -        | -         | -         | -         | -         | -         |
| c.928A>G | 11612573 | -        | -         | -         | 0.000008  | -         | -         |
| c.931C>T | 11612576 | -        | -         | -         | -         | -         | -         |
| c.946C>G | 11612591 | -        | 0.000008  | -         | -         | -         | -         |
| c.955A>G | 11612600 | -        | -         | -         | -         | -         | -         |
| c.958C>T | 11612603 | -        | -         | -         | 0.000008  | -         | -         |
| c.989C>G | 11612634 | -        | -         | -         | -         | -         | -         |
| c.1017C>A | 11614463 | -        | -         | -         | -         | -         | -         |
| c.1037C>T | 11614443 | 0.00060  | 0.00178   | 0.00124   | 0.00238   | 0.00149   | -         |
| c.1060G>A | 11614503 | -        | -         | -         | -         | -         | -         |
| c.1073G>C | 11614520 | -        | -         | -         | -         | -         | -         |
| c.1075G>C | 11614521 | -        | -         | -         | -         | -         | -         |
| c.1079G>C | 11614524 | -        | 0.000074  | 0.000008  | 0.000135  | -         | -         |
| c.1079A>G | 11614525 | -        | -         | -         | -         | -         | -         |

Table 1. Continued.
Table 1. Continued.

| Variant | Allele  | Frequency | Minor Allele  | Minor Allele Freq | Minor Allele Freq 95% CI | Major Allele Freq | Major Allele Freq 95% CI |
|---------|---------|-----------|--------------|-------------------|--------------------------|-------------------|--------------------------|
| c.1081A>G | 11614527 | -         | -            | -                 | -                        | -                 | -                        |
| c.1129A>G | 11614575 | 0.04293   | 0.09621      | 0.10632           | 0.10057                  | 0.08186          | -                        |
| c.1180C>A | 11615835 | 0.0064    | 0.002522     | 0.00003           | -                        | 0.000080         | 0.0025                   |
| c.1196T>G | 11615851 | -         | -            | -                 | -                        | -                 | -                        |
| c.1207C>A | 11615862 | -         | -            | -                 | -                        | -                 | -                        |
| c.1211A>G | 11615866 | -         | -            | -                 | -                        | -                 | -                        |
| c.1220C>A | 11615875 | 0.0012    | -            | 0.00010           | 0.00113                  | 0.000247         | 0.000625                 |
| c.1273G>A | 11615928 | 0.0034    | 0.002117     | 0.00003           | -                        | 0.000239         | 0.01188                  |
| c.1286G>C | 11615941 | -         | -            | -                 | -                        | -                 | -                        |
| c.1288C>G | 11615943 | -         | -            | -                 | -                        | -                 | -                        |
| c.1295T>C | 11615950 | -         | -            | -                 | -                        | -                 | -                        |
| c.1306C>T | 11615961 | -         | -            | -                 | -                        | -                 | -                        |
| c.1310G>C | 11615965 | -         | -            | -                 | -                        | -                 | -                        |
| c.1324G>A | 11615979 | -         | -            | -                 | 0.000008                 | -                 | -                        |
| c.1325C>T | 11615980 | -         | -            | -                 | 0.00008                  | 0.000104         | -                        |

individuals studied in different countries are depicted in Table 2. Genetic alterations in c.1129A>G were reported in 3 countries: China (0.33%), Germany (23%), and Australia (19.04%), with the highest frequency in Germany. Additionally, c.874T>C (45.58%), which was reported in Germany, represented the highest frequency among all the reported variations.

Bioinformatics
The results of the identification and analysis of the variations via online prediction tools are shown in Table 3. Out of the 110 substitutions identified, PROVEAN predicted 55 variations to be deleterious and 50 variations to be neutral. (Five variations were not available.) SIFT predicted 62 alterations to be damaging and 33 variations to be tolerated. (Fourteen variations were not available.) PolyPhen-2 defined 25 variations as benign, 18 as possibly damaging, and 59 as probably damaging. (Eight variations were not available.) MutationTaster predicted 82 disease-causing variations and 11 polymorphisms. (Seventeen variations were not available.) The maximum CADD score (Phred score = 53) was shown by c.796C>T R266X, indicating high pathogenicity, while c.196G>A, A66T showed the lowest CADD score (Phred score = 0.009). As a result, among the 110 substitutions, 38 were predicted to be deleterious by PROVEAN, SIFT, PolyPhen-2, and MutationTaster.

In this study, c.1075G>A indicated the highest GERP score (5.83), which represents 4.83 fewer substitutions than was expected. No negative GERP scores were reported for these variations.

Protein-Protein Interaction Network Analysis
As is illustrated in Figure 2, STRING, version 11.0, demonstrated that 11 proteins (GATA4, NKX2-5, MEF2C, ZFPM2, TBX5, BMP4, SRF, BMP2, HAND2, NPPA, and HEY2) and 41 edges (protein-protein associations) grouped to create a protein network.

Differences Between the Wild-Type GATA4 Protein and the Mutant Model
In this study, the effects of the predicted disease-causing p.Gly221Arg variant in GATA4 with the CADD Phred score of 31 were further analyzed. The variant, p.Gly221Arg, with a high level of pathogenicity is a heterozygous missense variant in the conserved N-terminal zinc finger of GATA4. HOPE results showed the alteration of glycine to arginine at position 221 (G221R, CADD Phred = 31). The size, charge, and hydrophobicity value of the 2 residues, as well as the differences between them, are presented in Figure 3A. The mutant residue showed a larger size, with a positive charge, while the wild-type protein charge was neutral. Furthermore, arginine was more hydrophobic than was glycine. These differences in amino acid features could affect the zinc finger site of the protein and its function. Accordingly, this change in the GATA4 sequence might result in the conformation of the protein and exert negative influences on the structure of the protein in this specific residue (Figure 3B). I-TASSER produced 3D structures of GATA4 in 5 models with different C-scores. A model with a C-score of −0.5, an estimated TM-score of 0.65, and an estimated RMSD of 8.2 Å was selected. Hence, the findings proved that the solubility of the mutant protein was similar to that of the wild-type one, with a score of 3 (Figure 3C).

GATA4 Protein Sequence Alignment and the Phylogenetic Tree
According to the phylogenetic tree generated by ClustalW, the human GATA4 protein had the closest homology with that of Canis lupus familiaris (dogs). Further, the most distant orthologue was Xenopus laevis (African clawed...
Table 2. Frequency of the variants in different populations

| DNA Change   | Condition               | Population (Frequency) | CHD Type | References |
|--------------|-------------------------|------------------------|----------|------------|
| c.17C>T      |                         | China (0.2%)           | VSD      | 48         |
| c.46G>T      |                         | China (0.62%)          | AF       | 49         |
| c.62G>T      | Uncertain significance  | China (1%)             | ASD      | 50, 51     |
| c.82C>G      |                         | China (0.62%)          | AF       | 49         |
| c.82C>T      |                         | China (2%)             | VSD      | 52         |
| c.106C>T     |                         | China (0.45%)          | ASD      | 53         |
| c.112T>G     |                         | China (0.66%)          | AF       | 53         |
| c.115G>T     |                         | China (0.45%)          | DCM      | 51         |
| c.127C>T     | Uncertain significance  | China (0.62%)          | VSD      | 54         |
| c.136-138delICC |                  | China (0.2%)          | VSD      | 48         |
| c.151C>G     |                         | China (1.92%)          | TOF      | 55         |
| c.155C>T     | Pathogenic              | Japan (6.25%)          | ASD      | 54         |
| c.164A>G     |                         | China (0.43%)          | VSD      | 57         |
| c.191G>A     |                         | China (0.38%)          | VSD, CTD | 58, 59     |
| c.196G>A     |                         | China (0.29%)          | VSD, PDA, TOF | 59-61 |
| c.206G>A     |                         | Australia (0.28%)      | VSD      | 62         |
| c.209G>C     |                         | China (0.76%)          | AF       | 63         |
| c.221C>A     |                         | China (0.26%)          | PS       | 63         |
| c.259C>T     |                         | China (0.55%)          | ASD      | 59         |
| c.270C>A     |                         | China (0.83%)          | CHD      | 64         |
| c.278G>C     |                         | America (0.15%)        | ASD      | 65         |
| c.284A>G     |                         | China (0.83%)          | CHD      | 64         |
| c.286G>A     |                         | China (0.43%)          | VSD      | 57         |
| c.307C>G     |                         | China (0.66%)          | AF       | 50         |
| c.357-359CGC | Pathogenic              | China (0.2%)           | VSD      | 48         |
| c.392C>G     | Uncertain significance  | Brazil (3.12%)         | AVSD     | 66         |
| c.431C>T     |                         | Japan (0.9%)           | PA, ASD  | 67         |
| c.448G>T     |                         | China (0.26%)          | TOF      | 61         |
| c.479G>C     |                         | China (0.76%)          | AF       | 61         |
| c.487C>T     | Pathogenic; Uncertain significance | China (0.31%) | AVSD, VSD, SA+SV, TOF, TGA | 48, 50, 51, 52, 54 |
| c.488C>G     | Uncertain significance  | Australia (0.28%)      | VSD      | 62         |
| c.569A>G     |                         | China (0.45%)          | ASD      | 52         |
| c.578C>A     |                         | Egypt (9.09%)          | VSD      | 70         |
| c.590A>G     |                         | China (0.43%)          | VSD      | 57         |
| c.620C>T     |                         | India (3%)             | ASD      | 73         |
| c.622T>C     |                         | Germany (1.47%)        | VSD      | 72         |
| c.628G>A     | Uncertain significance  | China (0.26%)          | AVSD     | 61         |
| c.631T>C     |                         | Germany (2.9%)         | VSD, AVSD | 72         |
| c.640G>A     |                         | Germany (1.47%)        | VSD      | 72         |
| c.640G>A     |                         | India (1%)             | ASD      | 71         |
| c.648C>G     |                         | Lebanon (1.66%)        | TOF      | 73         |
| c.661G>A     | Pathogenic              | France (family-based)  | CHD      | 74         |
| c.668T>C     |                         | Germany (1.47%)        | VSD      | 72         |
| c.677C>A     |                         | China (0.45%)          | DCM      | 53         |
| c.687G>T     |                         | Germany (4.41%)        | VSD      | 73         |
| c.700G>A     |                         | Germany (1.47%)        | AVSD     | 72         |
Table 2. Continued.

| Variant ID | Location | Country | Frequency | Clinical Significance |
|------------|----------|---------|-----------|----------------------|
| GATA4 variants in CHDs
| J Cardiovasc Thorac Res, 2021, 13(4), 336-354 |

| c.715A>G | Germany (1.47%) | VSD |
| c.716A>G | Germany (1.47%) | VSD |
| c.731A>G | Germany (2.94%) | VSD |
| c.740T>C | Germany (1.04%) | AF |
| c.743A>G | Germany (2.94%) | VSD, ASD, AVSD |
| c.749T>A | China (0.26%) | ASD, AVSD |
| c.754C>T | Germany (1.47%) | AVSD |
| c.764T>C | Germany (1.47%) | ASD |
| c.779G>A | Germany (1.47%) | VSD |
| c.782T>C | Germany (2.94%) | VSD, ASD |
| c.783T>G | China (0.45%) | ASD |
| c.788C>G | China (0.44%) | VSD |
| c.796C>T | Germany (2.94%) | ASD, AVSD |
| c.799C>A | China (0.58%) | ASD, CTD |
| c.812G>C | Germany (1.47%) | ASD |
| c.818A>G | Iran (1) | ASD, TOF, VSD |
| c.822C>T | Germany (0.97%) | ASD, ASD, DCM, TOF, VSD |
| c.830C>T | Germany (1.47%) | VSD |
| c.835A>T | China (0.45%) | DCM |
| c.839C>T | Germany (13.33%) | VSD, ASD |
| c.848C>A | Germany (1.47%) | AVSD |
| c.851C>A | France (0.3%) | ASD |
| c.854A>G | China (1.92%) | TOF |
| c.855T>C | Germany (1.47%) | AVSD |
| c.871C>C | China (0.66%) | DCM |
| c.874T>C | Germany (45.58%) | ASD, VSD, AVSD |
| c.881C>T | Germany (1.47%) | ASD, CHD |
| c.886G>C | Pathogenic China (0.47%) | VSD |
| c.886G>A | Pathogenic America (family-based) | ASD, PVS |
| c.886G>T | Pathogenic America (0.93%) | ASD |
| c.899A>C | China (family-based) | ASD |
| c.905A>G | Germany (1.47%) | AVSD |
| c.928A>G | Pathogenic China (family-based) | ASD |
| c.931C>T | China (family-based) | TOF, VSD, ASD, PDA |
| c.946C>G | Pathogenic America (0.31%) | ASD |
| c.958C>T | Likely pathogenic; Uncertain significance Italy (family-based) | ASD |
| c.989C>G | Uncertain significant Japan (0.39%) | PTA, ASD |
| c.1017C>A | Japan (0.39%) | PTA, ASD |
| c.1037C>T | America (0.93%) | ASD |
| c.1060G>A | China (1.17%) | ASD |
| c.1074delC | Japan (family-based) | ASD |
| c.1075G>A | Pathogenic China (0.41%) | VSD |
| c.1075delG | Pathogenic Japan (family-based) | ASD |
| c.1079A>G | China (0.26%) | VSD |
| c.1081A>G | Germany (1.47%) | VSD |
Table 2. Continued.

| SNP     | Ethnicity       | Mutations                | Phenotype                        | Cases (%)       | Controls (%)     | Description                        | Ref. |
|---------|-----------------|--------------------------|----------------------------------|-----------------|------------------|------------------------------------|------|
| c.1129A>G | -               | China (0.33%)            | ASD, VSD, AVSD, TOF, PA          | 17, 30, 62, 80, 95 |                   |                                    |      |
| c.1180C>A | -               | India (1.62%)            | AVSD, VSD                       | 52              |                  |                                    |      |
| c.1196T>G | -               | China (0.45%)            | ASD                           | 53              |                  |                                    |      |
| c.1207G>A | -               | America (0.93%)          | ASD                           | 68              |                  |                                    |      |
| c.1211G>T | -               | China (0.43%)            | VSD                           | 57              |                  |                                    |      |
| c.1220C>A | -               | China (0.59%) (Family-based) | ASD, AVSD, VSD, TOF, VSD      | 17, 40, 49, 77, 91 |                   |                                    |      |
| c.1273G>A | -               | America (0.16%)          | PA, PS, ASD, TOF, AVSD          | 65              |                  |                                    |      |
| c.1286G>C | -               | China (0.2%)             | VSD                           | 48              |                  |                                    |      |
| c.1288C>G | -               | Germany (2.94%)          | ASD                           | 98              |                  |                                    |      |
| c.1295T>C | -               | India (0.32%)            | PDA                           | 52              |                  |                                    |      |
| c.1306C>T | -               | China (8%)               | CSDs                          | 60              |                  |                                    |      |
| c.1310G>C | -               | America (1.28%)          | BAV                           | 94              |                  |                                    |      |
| c.1324G>A | -               | Germany (1.47%)          | VSD                           | 99              |                  |                                    |      |
| c.1325C>T | Pathogenic      | China (0.34%)            | VSD                           | 68              |                  |                                    |      |

Abbreviation: AVSD, atrioventricular septal defect; ASD, atrial septal defects; CDH, congenital diaphragmatic hernia; CTD, conotruncal heart defects; CHD, congenital heart disease; CSDS, cardiac septal defects; DCM, dilated cardiomyopathy; DORV, double-outlet right ventricle; DILV, double-inlet left ventricle; LVHT, left ventricular hypertrophy; LVNC, left ventricular noncompaction; PA, pulmonary atresia; PA + IVS, pulmonary atresia with interventricular septum; PVS, pulmonary valve stenosis; SA+SV, single atrium with single ventricle; TGA, transposition of the great arteries; TOF, tetralogy of Fallot; TGA, transposition of the great arteries; VSD, ventricular septal defect; PTA, persistent truncus arteriosus; BAV, bicuspid aortic valve; AF, atrial fibrillation; PDA, patent ductus arteriosus; PS, pulmonary stenosis

frogs) (Figure 4A). The results of the multiple-alignment sequencing of the species are illustrated in Figure 4B.

Discussion

CHD is the most frequent congenital abnormality and the major cause of infant mortality the world over. GATA4, a transcription factor with 2 zinc finger domains, has been reported to play an essential role in embryogenesis and cardiac development. The GATA4 gene is reported to modulate heart hypertrophy in adults. The number of studies seeking to explicate the correlation between GATA4 variants and CHD occurrence is on the rise. Indeed, recent studies have identified several novel variants in the GATA4 gene with potential roles in CHD development.

CHD is very heterogeneous, and the etiology of the majority of cases remains greatly unknown. Both genetic and environmental factors contribute to CHD. Therefore, the elucidation of the pathogenesis and differential diagnosis of the disease requires the identification of not only the disease-causing or susceptibility genes but also new genetic variants associated with the different types of CHD. Research has linked several genes to CHD, with NKX2-5, TBX5, and GATA4 comprising the most studied transcription factor genes. These genes interact during embryonic development, and they are involved in the regulation of cardiogenesis and embryonic heart development. Protein-protein interactions between transcription factors play a vital role in biological systems. The results concerning GATA4 protein interactions, generated by STRING, showed that 11 proteins (GATA4, NKX2-5, MEF2C, ZFPM2, TBX5, BMP4, SRF, BMP2, HAND2, NPPA, and HEY2) grouped in a network. GATA4 and NKX2-5 transcription factors are critical to cardiomyocyte hypertrophy; thus, single-point variants could create an imbalance in the interaction between these proteins. Remarkably, BMP2 and BMP4 have been shown to interact with HAND2 to modulate the transcription of the downstream gene by binding to the conserved GATA-binding sites of the HAND2 promoter. NKX2-5, as a central regulator of many aspects of heart development, interacts with SRF and GATA4 to promote the expression of the cardiac sarcomeric protein gene. Mutations in the ZFPM2 gene, which encodes the FOG2 protein (a transcription regulator of the GATA family members), disrupt the interaction with GATA4 or the nucleosome remodeling and deacetylation (NuRD) complex and, thus, lead to CHD. Loss-of-function mutation in the MEFS2 gene, which encodes a transcription factor required for normal cardiovascular development, is associated with increased vulnerability to CHD in humans. MEF2C, TBX5, and GATA4 can induce cardiomyocyte differentiation and directly reprogram endogenous cardiac fibroblasts into functional cardiomyocytes. Remarkably, BMP2 and BMP4 are vital for cardiogenesis in that they induce the expression of NKX2-5 and GATA4 transcription factors. These 2 genes play a significant role during the initial induction of cardiogenesis. Nevertheless, no association between BMP2 and BMP4 genetic variations (rs1049007, rs235768, and rs17563) and the risk of CHD
GATA4 variants in CHDs

J Cardiovasc Thorac Res, 2021, 13(4), 336-354

Variations in the NPPA gene, which encodes the ANP precursor, are correlated with hypertension, stroke, coronary artery disease, and heart failure. \(^\text{106}\) The HEY2 transcription factor plays an important function in mammalian heart development. \(^\text{106}\)

Somatic mutations were identified in NKK2-5 and its molecular partners, TBX5 and GATA4, as well as the transcription factor HEY2, in formalin-fixed tissues taken from a collection of hearts with atrial septal defect, \(^\text{107}\) ventricular septal defect, and atrioventricular canal defect. \(^\text{90, 108, 110-112}\)

The GATA4 missense variation (p.G221R), on which we focused in the present study, was identified in three 46, XY DSD patients from a family of French origin. The in vitro assays in that investigation demonstrated the failure of the p.G221R mutant protein to bind to FOG2, which is required for gonad formation. Furthermore, the mutant protein failed to transactivate the anti-Müllerian hormone promoter. \(^\text{74}\)

Some variants of GATA4 investigated in the present study have been previously analyzed for genotype-phenotype correlations. These investigations evaluated families manifesting those variations associated with different CHD types.

Lourenço D et al \(^\text{74}\) reported the G221R variant in 5 members of a family with cardiac anomalies including atrial septal defect, tetralogy of Fallot, and congenital cyanotic heart disease.

In a study conducted by Garg V et al, \(^\text{84}\) the c.886G > A (G296S) variation of GATA4 was stated in 13 affected members with atrial septal defect in a family with 5 generations. The authors also reported the E359del variation of GATA4 in 5 members of another family with

Three non-synonymous variations, namely c.286A > G (p.Thr96Ala), c.293A > C (p.Asp98Ala), and c.299T > C (p.Leu100Ser), were reported to affect the second helix of HEY2 in the diseased cardiac tissues of 2 cases with atrioventricular septal defect, suggesting its possible function in the regulation of ventricular septation in humans. \(^\text{108}\) Somatic mutations were identified in NKK2-5 and its molecular partners, TBX5 and GATA4, as well as the transcription factor HEY2, in formalin-fixed tissues taken from a collection of hearts with atrial septal defect, \(^\text{109}\) ventricular septal defect, and atrioventricular canal defect. \(^\text{90, 108, 110-112}\)

Figure 2: The image presents the STRING protein-protein interaction analysis. The network contains 11 nodes and 41 edges. The edges are represented with various colors, with each color indicating protein-protein associations. The GATA4 protein and its functional interactions with 11 other proteins display possible effects on each other.

NKK2-5: Homeobox protein NKK2-5, MEF2C: Myocyte-specific enhancer factor 2C, ZFPM2: Zinc finger protein ZFPM2, TBX5: T-box transcription factor, BMP4: Bone morphogenetic protein 4, SRF: Serum response factor, BMP2: Bone morphogenetic protein 2, HAND2: Heart- and neural crest derivatives-expressed protein 2, NPPA: Natriuretic peptides A, HEY2: Hairy/enhancer-of-split related with YRPW motif protein 2.

Figure 3: A) The image presents the schematic structure of a normal amino acid on the left (glycine) and a mutant one on the right (arginine) at position 221 of the GATA4 protein. The red parts show the similar parts of the amino acids (the backbone), and the black part shows the unique part of the amino acids (the side chain). This picture illustrates the structural differences between the 2 amino acids. The G221R alteration is shown by HOPE. B) A photograph generated by HOPE shows that the G221R variation affects the structure of the GATA4 protein. The green color shows the wild-type residue (glycine), and the red color represents the mutant residue (arginine). C) I-TASSER shows the secondary and 3D structure, as well as the predicted solvent accessibility, of the normal (left) and G221R mutant (right) of the GATA4 protein.
Table 3. In silico analysis of GATA4 variations

| DNA Change | Protein Change | Variant Type | dbSNP        | HGMD       | CADD^1 | MutationTaster | PolyPhen^2 Score | PROVEAN^3 | SIFT^3 (Score) | GERP       |
|------------|---------------|--------------|--------------|------------|--------|----------------|----------------|------------|---------------|------------|
| c.17C>T    | A6V           | Missense     | rs199922907  | CM086621   | 24.48  | DC             | PRD (0.986)    | NE         | DE (0.01)     | NA         |
| c.46G>T    | G16C          | Missense     | rs33331682   | CM117802   | 23     | POLYMORPHISM   | PRD (1.000)    | NE         | TO (0.1)     | NA         |
| c.62G>T    | G21V          | Missense     | rs202213149  | CM107596   | 24.4   | DC             | PRD (0.972)    | NE         | DE (0.02)     | NA         |
| c.82C>G    | H28D          | Missense     | rs1406275331 | CM117803   | 25     | DC             | PRD (0.993)    | DE         | DE (0)       | NA         |
| c.82C>T    | H28Y          | Missense     | rs1406275331 | CM0910178  | 24     | DC             | PRD (0.993)    | DE         | DE (0)       | NA         |
| c.106C>T   | P36C          | Missense     | CM1313746    | 25.2       | DC     | PRD (1)        | DE (0)        | NA         |               |            |
| c.112T>G   | Y38D          | Missense     | CM123513     | 26         | DC     | PRD (0.998)    | DE (0)        | NA         |               |            |
| c.115G>T   | V39L          | Missense     | rs1139241    | 24         | DC     | PRD (0.950)    | DE (0)        | NA         |               |            |
| c.127C>T   | R43W          | Missense     | CM19519      | 25         | DC     | PRD (1)        | DE (0)        | NA         |               |            |
| c.136-138delITCC | 46delI5 deletion | - | - | - | - | NE | TO (0) | NA |               |            |
| c.151C>G   | L51V          | Missense     | rs11312064   | 23.6       | DC     | PRD (0.977)    | DE (0)        | NA         | ME (0.01)     | NA         |
| c.155C>T   | S52F          | Missense     | rs104894074  | CM1312064  | 25.6   | DC             | PRD (0.975)    | DE         | DE (0)       | NA         |
| c.164A>G   | Q5SR          | Missense     | CM125062     | 22.9       | DC     | POD (0.586)    | DE (0)        | NA         | ME (0.01)     | NA         |
| c.191G>A   | G64E          | Missense     | rs1249347695 | CM107237   | 11.21  | POLYMORPHISM   | BENIGN (0.392) | NE         | TO (0.99)    | NA         |
| c.196G>A   | A66T          | Missense     | rs1139244    | 0.009      | DC     | BENIGN (0)     | TO (0.58)     | NA         |               |            |
| c.206G>A   | G69D          | Missense     | CM109056     | -          | POLYMORPHISM | BENIGN (0.157) | TO (0.46)    | NA         |               |            |
| c.209G>C   | S70T          | Missense     | CM115165     | 10.71      | DC     | BENIGN (0.001) | TO (0.46)     | NA         |               |            |
| c.221C>A   | A74D          | Missense     | rs1258064099 | CM1010265  | 19.96  | POLYMORPHISM   | PRD (0.997)    | NE         | TO (0.14)    | NA         |
| c.244A>G   | T82A          | Missense     | rs961114777  | 12.10      | POLYMORPHISM | BENIGN (0) | - | TO (0.38)     | NA         |
| c.259C>T   | P87S          | Missense     | CM107597     | -          | -      | PRD (0.977)    | TO (0.14)     | NA         |               |            |
| c.270C>A   | S90R          | Missense     | CM104917     | -          | -      | BENIGN (0.440) | DE (0.04)    | NA         |               |            |
| c.278G>C   | G93A          | Missense     | rs56191129   | CM076206   | 19.85  | DC             | POD (0.943)    | NE         | TO (0.09)    | NA         |
| c.284A>G   | D95G          | Missense     | CM104918     | -          | -      | BENIGN (0)     | TO (0.06)     | NA         |               |            |
| c.286G>A   | G96R          | Missense     | CM1213107    | -          | POLYMORPHISM | BENIGN (0.012) | TO (0.06)    | NA         |               |            |
| c.307C>G   | P103A         | Missense     | CM123514     | 19.50      | DC     | BENIGN (0.001) | TO (0.7)     | NA         |               |            |
| c.357,359GCG | A126dup Duplication | rs1142566703 | - | - | - | NE | TO (0) | NA |               |            |
| c.392C>G   | A131G         | Missense     | rs1013983424 | 18.4       | POLYMORPHISM | BENIGN (0.002) | TO (0.66)    | NA         |               |            |
| c.411C>T   | A144V         | Missense     | rs1308945507 | CM161974   | 14.22  | POLYMORPHISM   | POD (0.727)    | NE         | TO (0.16)    | NA         |
| c.448G>T   | G150W         | Missense     | rs1024075653 | CM1010266  | 26.0   | DC             | PRD (0.997)    | DE         | DE (0)       | NA         |
| c.479G>C   | S160T         | Missense     | rs135865879  | CM115166   | 23.5   | DC             | POD (0.891)    | NE         | TO (0.35)    | NA         |
| c.487C>T   | P163S         | Missense     | rs387906769  | CM076201   | 22.1   | DC             | POD (0.669)    | NE         | NA           | NA         |
| c.488C>G   | P163R         | Missense     | rs540578024  | CM109057   | 25.4   | DC             | PRD (0.973)    | DE         | TO (0.42)    | NA         |
Table 3. Continued.

| c.569A > G | H190R | Missense | - | CM1313747 | 24 | DC | PRD (0.988) | DE | DE (0) | NA |
|---|---|---|---|---|---|---|---|---|---|---|
| c.578C > A | P193H | Missense | - | - | 24.2 | DC | POD (0.921) | NE | DE (0.05) | NA |
| c.590A > G | N197S | Missense | - | CM125063 | 15.18 | DC | POLYMORPHISM | BENIGN (0.009) | NE | TO (0.65) | NA |
| c.620C > T | 5'UTR | rs61277615 | - | 8.145 | - | - | - | - | - | NA |
| c.622T > C | F208L | Missense | - | - | 20.6 | DC | BENIGN (0.071) | NE | TO (1) | NA |
| c.628G > A | D210N | Missense | rs377673676 | CM1010267 | 32 | DC | PRD (0.996) | DE | NA | 5.08 |
| c.631T > C | F211L | Missense | - | 22.4 | DC | BENIGN (0.05) | NE | TO (0.47) | NA |
| c.639A > G | N197S | Missense | - | CM125063 | 15.18 | DC | POLYMORPHISM | BENIGN (0.009) | NE | TO (0.65) | NA |
| c.661G > A | G221R | Missense | rs398122402 | CM110562 | 32 | DC | PRD (0.999) | DE | DE (0) | NA |
| c.665T > C | M223T | Missense | - | 23.6 | DC | BENIGN (0.126) | DE | TO (0.32) | NA |
| c.677C > A | P226Q | Missense | - | CM147378 | 25.3 | DC | PRD (1) | DE | DE (0) | NA |
| c.687G > T | R229S | Missense | - | - | 24.2 | DC | PRD (0.998) | DE | DE (0) | 3.83 |
| c.700A > G | G234S | Missense | - | - | 28.1 | DC | PRD (1) | DE | DE (0) | 5.61 |
| c.716T > C | M237T | Missense | - | - | 23.6 | DC | BENIGN (0.05) | NE | TO (0.47) | NA |
| c.718A > G | N239D | Missense | - | - | 25.8 | DC | PRD (1) | DE | DE (0) | NA |
| c.721A > G | N239S | Missense | - | - | 28.9 | DC | PRD (1) | DE | DE (0) | NA |
| c.731A > G | Y244C | Missense | - | - | 25.9 | DC | PRD (0.994) | DE | DE (0) | NA |
| c.749C > T | R252W | Missense | - | CM1010268 | 27.8 | DC | PRD (0.994) | DE | DE (0.01) | NA |
| c.796C > T | R266X | Nonsense | - | - | - | - | - | - | - | NA |
| c.799G > A | V267M | Missense | rs116781972 | CM068343 | 24.9 | DC | BENIGN (0.401) | NE | TO (0.09) | NA |
| c.812G > C | C271S | Missense | - | - | 27.7 | DC | PRD (1) | DE | DE (0) | NA |
| c.818A > G | N273S | Missense | rs1340083717 | - | 25.9 | DC | PRD (1) | DE | DE (0) | NA |
| c.819C > A | N273K | Missense | - | - | 25.3 | DC | PRD (1) | DE | DE (0) | NA |
| SNP       | Allele | p-value | rS | EF | BMI | WC | HC | Blood | 200 | 300 | 400 |
|-----------|--------|---------|----|----|-----|----|----|--------|------|------|------|
| rs12367890 | A      | 0.012   | 0.8 | 0.99 | 0.89 | 0.79 | 0.99 | 0.79 | 0.79 | 0.79 | 0.79 |
| rs12367891 | G      | 0.012   | 0.8 | 0.99 | 0.89 | 0.79 | 0.99 | 0.79 | 0.79 | 0.79 | 0.79 |

Table 1. Continued.
All GATA4 variants are reported based on the NCBI nucleotide (NM_002052.5) and protein (NP_002043.2) sequences (NG_008177.2).

CADD, Phred ≤ 20: Neutral; Phred > 20: Damaging; PolyPhen-2, score = 0-0.15: Benign; score = 0.15-0.85: Possibly damaging; score = 0.85-1: Probably damaging; PROVEAN, score ≤ -2.5: Deleterious; score > -2.5: Neutral; SIFT, score ≤ 0.05: Deleterious; score > 0.05: Tolerable; TO: Tolerable; DE: Deleterious; NE: natural; DC: Disease-causing; NA: Not available; PRD: Probably damaging; POD: Possibly damaging.

| Variant        | Allele | Type | CADD | PolyPhen-2 | PROVEAN | SIFT | DC | NE | TO | DE | Score |
|----------------|--------|------|------|------------|---------|------|----|----|----|----|-------|
| c.1196T>G      | V399G  | Missense | -    | 20.7       | DC      | BENIGN | NE | TO | NA |
| c.1207C>A      | L403M  | Missense | -    | 25         | DC      | PRD    | NE | TO | NA |
| c.1211A>G      | K404R  | Missense | -    | 26.4       | DC      | PRD    | NE | DE | NA |
| c.1220C>A      | P407Q  | Missense | rs115099192 | 25.8    | DC      | POD    | DE | DE | NA |
| c.1273G>A      | D425N  | Missense | rs56208331 | 29       | DC      | PRD    | DE | DE | NA |
| c.1286G>C      | S429T  | Missense | -    | 23.18      | DC      | POD    | NE | DE | NA |
| c.1288C>G      | L430V  | Missense | -    | 24.8       | DC      | PRD    | NE | DE | NA |
| c.1295T>C      | L432S  | Missense | -    | 27.98      | DC      | PRD    | NE | DE | NA |
| c.1306C>T      | H436Y  | Missense | -    | 26.08      | DC      | POD    | NE | DE | NA |
| c.1310G>C      | G437A  | Missense | -    | 23.5       | DC      | POD    | NE | DE | NA |
| c.1324G>A      | A442T  | Missense | rs1270266865 | -       | 26.8    | DC      | PRD | NE | DE | NA |
| c.1325C>T      | A442V  | Missense | rs146017816 | -       | 27.1    | DC      | PRD | NE | DE | 5.18 |

Table 3. Continued.
the autosomal dominant transmission of atrial septal defect in 4 generations, indicating $GATA4$ as a genetic cause of atrial septal defect.

Sarkozy et al. detected the G296S variation of $GATA4$ in 2 members of 1 family and 3 members of another family diagnosed with atrial septal defect.

Chen J et al. recognized the $GATA4$ c.899A > C (K300T) substitution in 10 members of a family: 8 affected members with severe symptoms (7 patients with atrial septal defect and 1 patient with ventricular septal defect) and 2 unaffected members. The K300T substitution lessens the transcription of the $GATA4$ target gene by harming the DNA-binding activity of $GATA4$.

Yu Chen et al. identified the c.928A > G (M310V) variant located in the NLS region of $GATA4$ in all patients of a 3-generation family with atrial septal defect. The variant reduces the transcriptional activity of the $GATA4$ protein and may disturb the interaction between $GATA4$ and TBX5.

A genetic investigation conducted by E. D’Amato et al. reported the R319W variation in 3 members of a family: the proband and the proband’s sister, both diagnosed with atrial septal defect, and the proband’s father, who was considered not affected.

Rajagopal et al. studied 107 probands with cardiac abnormalities and identified the c.886G > T (G296C) variant in a proband with atrial septal defect and pulmonary stenosis. They also reported the substitution in the proband’s father with persistent left superior vena cava to the coronary sinus. The G296S variation resulted in a reduction in $GATA4$ DNA-binding activity and disrupted binding to the transcription factor TBX5. Also in their study, the c.1207C > A (L403M) variant was identified in a proband with a hypoplastic right ventricle and sinus venous atrial septal defect. Their results also demonstrated the c.487C > T (P163S) and c.1037C > T (A346V) variants in probands with endocardial cushion defect. Additionally, a missense variation, c.931C > T (R311W), in $GATA4$ was identified in a pedigree spanning 3 generations with 7 members diagnosed with CHD. All the affected members presented different cardiac phenotypes, including tetralogy of Fallot, ventricular septal defect, atrial septal defect, and patent ductus arteriosus, indicating that the same genetic alteration could lead to different subtypes of CHD.

In the present study, we filtered the literature and online databases for the pathogenic variants of the $GATA4$ gene. Our search yielded 210 variants; nonetheless, we excluded 100 of these variants due to a dearth of information and continued the study with 110 variations. After analyzing the frequency distributions of all the variants, we employed computational tools with different algorithms to predict the pathogenicity of the variants. As is shown in Table 3, our in silico analysis using MutationTaster, PolyPhen,
PROVEAN, and SIFT revealed 38 pathogenic genetic variations. Our findings may broaden the spectrum of the known GATA4 genetic variations associated with different types of CHD.

Conclusions

Several gene deficiencies could contribute to the pathogenesis of CHD. In this study, we drew upon different in silico predictive tools for the analysis of the variants of the GATA4 gene. The most frequent variant was c.874T>C (45.58%), and the most frequent type of CHD was ventricular septal defect. Out of all the reported variants of GATA4, 38 variants were pathogenic. The p.Gly221Arg variant (CADD score = 31) showed a high level of pathogenicity. All the identified pathogenic variations in GATA4 could assist in the rapid identification and better understanding of the mechanisms underlying CHD.

Acknowledgments

The authors wish to thank the Rajaie Cardiovascular Medical and Research Center, Tehran, Iran, and Zanjan University of Medical Sciences, Zanjan, Iran.

Competing Interests

None declared.

Ethical Approval

Not applicable.

Funding

This research was funded by Rajaie Cardiovascular Medical and Research Center, Tehran, Iran, and Zanjan University of Medical Sciences, Zanjan, Iran.

References

1. Williams K, Carson J, Lo C. Genetics of congenital heart disease. Biomolecules. 2019;9(12):879. doi:10.3390/biom9120879
2. Abdul Samad F, Suliman BA, Bashar SH, Manivasagam T, Essa MM. A comprehensive in silico analysis on the structural and functional impact of SNPs in the congenital heart defects associated with NKX2-5 gene-a molecular dynamic simulation approach. PLoS One. 2016;11(5):e0153999. doi:10.1371/journal.pone.0153999
3. Edwards JJ, Gelb BD. Genetics of congenital heart disease. Curr Opin Cardiol. 2016;31(3):235-241. doi:10.1097/hco.0000000000000274
4. Ferencz C, Loffredo CA, Rubin JD, Magee CA. Perspectives in Pediatric Cardiology. Armonk, New York: Futura Publishing Company; 1997.
5. Shabana NA, Shahid SU, Irfan U. Genetic contribution to congenital heart disease (CHD). Pediatr Cardiol. 2020;41(1):12-23. doi:10.1007/s00246-019-02271-4
6. Martinez de LaPiscina I, de Mingo C, Riedl S, Rodriguez A, Martinez de LaPiscina I, de Mingo C, Riedl S, Rodriguez A, Rodriguez A. The functional impact of SNPs in the congenital heart defects by disrupting semaphorin-plexin signaling. Proc Natl Acad Sci U S A. 2009;106(33):13933-13938. doi:10.1073/pnas.0904744106
7. Jiang QJ, Li RG, Wang J, Liu XY, Xu YJ, Fang WY, et al. Prevalence and spectrum of GATA5 mutations associated with congenital heart disease. Int J Cardiol. 2013;165(3):570-573. doi:10.1016/j.ijcard.2012.09.039
8. Granados-Riveron JT, Pope M, Bu’lock FA, Thornborough C, Eason J, Setchfield K, et al. Combined mutation screening of NKX2-5, GATA4, and TBX5 in congenital heart disease: multiple heterozygosity and novel mutations. Congenit Heart Dis. 2012;7(2):151-159. doi:10.1111/j.1747-0803.2011.00573.x
9. Wang J, Sun YM, Yang YQ. Mutation spectrum of the GATA4 gene in patients with idiopathic atrial fibrillation. Mol Biol Rep. 2012;39(6):8127-8135. doi:10.1007/s11033-012-1660-6
10. Li J, Liu WD, Yang ZL, Yuan F, Xu L, Li RG, et al. Prevalence and spectrum of GATA4 mutations associated with sporadic dilated cardiomyopathy. Gene. 2014;548(2):174-181. doi:10.1016/j.gene.2014.07.022
11. Jumpman M, Kinnunen SM, Valimäki MJ, Talman V, Auno S, Bruun T, et al. Synthesis, identification, and structure-activity relationship analysis of GATA4 and NKX2-5 protein-protein interaction modulators. J Med Chem. 2019;62(17):8284-8310. doi:10.1021/acs.jmedchem.9b01086
12. Wang T, Chen L, Yang T, Huang P, Wang L, Zhao L, et al. Congenital heart disease and risk of cardiovascular disease: a meta-analysis of cohort studies. J Am Heart Assoc. 2018;7(10):e012030. doi:10.1161/jaha.118.012030
13. Chen J, Qin B, Zhao J, Liu W, Duan R, Zhang M. A novel mutation of GATA4 (K300T) associated with familial atrial septal defect. Gene. 2016;575(2 Pt 2):473-477. doi:10.1016/j. gene.2015.09.021
14. Suluba E, Shuwei L, Xia Q, Mwanga A. Congenital heart diseases: genetics, non-inherited risk factors, and signaling pathways. Egypt J Med Hum Genet. 2020;21(1):11. doi:10.1186/s43042-020-0050-1
15. Su W, Zhu P, Wang R, Wu Q, Wang M, Zhang X, et al. Congenital heart diseases and their association with the variant distribution features on susceptibility genes. Clin Genet. 2017;91(3):349-354. doi:10.1111/cge.12835
16. Kalayvinia S, Maleki M, Rokni-Zadeh H, Changi-Ashtiani M, Ahanhar H, Biglari A, et al. GATA4 screening in Iranian patients of various ethnicities affected with congenital heart disease: co-occurrence of a novel de novo translocation (5;7) and a likely pathogenic heterozygous GATA4 mutation in a family with autosomal dominant congenital heart disease. J Clin Genet. 2019;57(7):20-29. doi:10.1016/j.jcg.2019.07.006
17. Clark KL, Utzey KE, Benson DW. Transcription factors and congenital heart defects. Annu Rev Physiol. 2006;68:97-121. doi:10.1146/annurev.physiol.68.041014.113828
18. El Bouchikhi I, Bouguenouch L, Moufid FZ, Belhassan K, Samri I, Chaouti A, et al. Absence of GATA4 mutations in Moroccan patients with atrial septal defect (ASD) provides further evidence of limited involvement of GATA4 in major congenital heart defects. Eur J Med Genet. 2015;58(1):32-35. doi:10.1016/j.ejmg.2014.12.003
19. Misra C, Sachan N, McNally CR, Koenig SN, Nichols HA, Guggilam A, et al. Congenital heart disease-causing GATA4 mutation displays functional deficits in vivo. PLoS Genet. 2012;8(5):e1002690. doi:10.1371/journal.pgen.1002690
20. Maitra M, Schluterman MK, Nichols HA, Richardson JA, Lo CW, Srivastava D, et al. Interaction of GATA4 and GATA6 with Tbx5 is critical for normal cardiac development. Dev Biol. 2009;326(2):368-377. doi:10.1016/j.ydbio.2008.11.004
21. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas...
effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;Chapter 7:Unit7.20.
doi:10.1002/humu.20112

39. Niroula A, Vihinen M. How good are pathogenicity predictors in detecting benign variants? *PloS Comput Biol*. 2019;15(2):e1006481. doi:10.1371/journal.pcbi.1006481

40. Huber CD, Kim BY, Lohmueller KE. Population genetic models of GERP scores suggest pervasive turnover of constrained sites across mammalian evolution. *PloS Genet*. 2020;16(5):e1008827. doi:10.1371/journal.pgen.1008827

41. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Mingeu P, et al. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res*. 2011;39(Database issue):D651-656. doi:10.1093/nar/gkq797

42. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(Database issue):D607-D613. doi:10.1093/nar/gky1131

43. Venselaar H, Te Beek TA, Kuipers RK, Heekelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 2010;11:548. doi:10.1186/1471-2105-11-548

44. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 2015;10(5):725-738. doi:10.1038/nprot.2015.5

45. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*. 2015;12(1):7-8. doi:10.1038/nmeth.3213

46. Yang J, Zhang Y. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res*. 2015;43(W1):W174-181. doi:10.1093/nar/gkv342

47. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9:40. doi:10.1186/1471-2105-9-40

48. Zhang W, Li X, Shen A, Jiao W, Guan X, Li Z. GATA4 mutations in 486 Chinese patients with congenital heart disease. *Eur J Med Genet*. 2008;51(6):527-535. doi:10.1016/j.ejmg.2008.06.003

49. Jiang QQ, Shen FF, Fang WY, Liu X, Yang YQ. Novel GATA4 mutations in lone atrial fibrillation. *Int J Mol Med*. 2011;28(6):1025-1032. doi:10.3892/ijmm.2011.783

50. Liu XY, Yang YQ, Ma J, Lin XP, Zheng JH, Bai K, et al. [Novel GATA4 mutations identified in patients with congenital atrial septal defects]. *Zhonghua Xue Guan Bing Za Zhi*. 2010;38(8):724-727. [Chinese].

51. Liu XY, Wang J, Zheng JH, Bai K, Liu ZM, Wang XZ, et al. Involvement of a novel GATA4 mutation in atrial septal defects. *Int J Mol Med*. 2011;28(1):17-23. doi:10.3892/ijmm.2011.638

52. Chen MW, Pang VS, Guo Y, Liu BL, Shen J, Song HD, et al. [Association between GATA-4 mutations and congenital cardiac septal defects in Han Chinese patients]. *Zhonghua Xue Guan Bing Za Zhi*. 2009;37(5):409-412. [Chinese].

53. Yang YQ, Wang J, Liu XY, Chen XZ, Zhang W, Wang XZ. Mutation spectrum of GATA4 associated with congenital atrial septal defects. *Arch Med Sci*. 2013;9(6):976-983. doi:10.5114/ams.2013.39788

54. Yang YQ, Li L, Wang J, Liu XY, Chen XZ, Zhang W, et al. A novel GATA4 loss-of-function mutation associated with congenital ventricular septal defect. *Pediatr Cardiol*. 2012;33(4):539-546. doi:10.1007/s00246-011-0146-y

55. Yang YQ, Gharibeh L, Li RG, Xin YF, Wang J, Liu ZM, et al.

Abbasi et al.

NS, et al. Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat*. 2003;21(6):577-581. doi:10.1002/humu.10212

23. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipirlla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D1067. doi:10.1093/nar/gkx1153

24. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics*. 2016;54:1.30.31-31.30.33. doi:10.1002/cpbi.5

25. Karczewski KJ, Weisburd B, Thomas B, Solomonson M, Ruderfer DM, Kavanagh D, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45(D1):D840-D845. doi:10.1093/nar/gkw971

26. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

27. Karczewski KJ, Franciolli LC, Tao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443. doi:10.1038/s41586-020-2308-7

28. Auer PL, Reiner AP, Wang G, Kang HM, Abecasis GR, Altshuler D, et al. Guidelines for large-scale sequence-based complex trait association studies: lessons learned from the NHLBI exome sequencing project. *Am J Hum Genet*. 2016;99(4):791-801. doi:10.1016/j.ajhg.2016.08.012

29. Brody JA, Morrison AC, Bis JC, O’Connell JR, Brown MR, Huffman JE, et al. Analysis commons, a team approach to discovery in a big-data environment for genetic epidemiology. *Nat Genet*. 2017;49(11):1560-1563. doi:10.1038/ng.3968

30. Fattahi Z, Beheshtian M, Mohseni M, Poustiti H, Sellas E, Nezhadi SH, et al. Iranome: a catalog of genomic variations in the Iranian population. *Hum Mutat*. 2019;40(10):1968-1984. doi:10.1002/humu.23880

31. Scott EM, Hallees A, Iltan Y, Spencer EG, He Y, Azab MA, et al. Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. *Nat Genet*. 2016;48(9):1071-1076. doi:10.1038/ng.3592

32. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*. 2012;40(Web Server issue):W452-457. doi:10.1093/nar/gks539

33. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248

34. Choi Y, Chan AP, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*. 2015;31(16):2745-2747. doi:10.1093/bioinformatics/btv195

35. Rentseas P, Widen D, Cooper GM, Shendure J, Kircher M, CAD2: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016

36. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361-362. doi:10.1038/nmeth.2890

37. Cooper GM, Stone EA, Asimenos G, Green ED, Batzoglou S, Sidow A. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15(7):901-913. doi:10.1101/gr.3577405

38. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional
GATA4 variants in CHDs

GATA4 loss-of-function mutations underlie familial tetralogy of Fallot. *Hum Mutat.* 2013;34(12):1662-1671. doi:10.1002/humu.22434

56. Hirayama-Yamada K, Kamisago M, Akimoto K, Aotsuka H, Nakamura Y, Tomita H, et al. Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. *Am J Med Genet A.* 2005;135(1):47-52. doi:10.1002/ajmg.a.30684

57. Yang YQ, Wang J, Liu XY, Chen XZ, Zhang W, Wang XZ, et al. Novel GATA4 mutations in patients with congenital ventricular septal defects. *Med Sci Monit.* 2012;18(6):CR344-350. doi:10.12659/msm.882877

58. Yang YQ, Tang YQ, Liu XY, Lin XP, Chen YH. [A novel GATA4 mutation leading to congenital ventricular septal defect]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi.* 2010;27(5):512-516. doi:10.3760/cma.j.issn.1003-9406.2010.05.008

59. Liu Y, Li B, Xu Y, Sun K. Mutation screening of GATA4 gene in CTD patients within Chinese Han population. *Pediatr Cardiol.* 2017;38(3):506-512. doi:10.1007/s00246-016-1542-0

60. Chen MW, Pang YS, Guo Y, Pan JH, Liu BL, Shen J, et al. GATA4 mutations in Chinese patients with congenital cardiac septal defects. *Pediatr Cardiol.* 2010;31(1):85-89. doi:10.1007/s00246-009-9576-1

61. Wang E, Sun S, Qiao B, Duan W, Huang G, An Y, et al. Identification of functional mutations in GATA4 in patients with congenital heart disease. *PloS One.* 2013;8(4):e26138. doi:10.1371/journal.pone.0062138

62. Butler TL, Esposito G, Blue GM, Cole AD, Costa MW, Waddell LB, et al. GATA4 mutations in 357 unrelated patients with congenital heart malformation. *Genet Test Mol Biomarkers.* 2010;14(6):797-802. doi:10.1089/gtm.2010.0028

63. Yang YQ, Wang MY, Zhang X, Tan HW, Shi HF, Jiang WF, et al. GATA4 loss-of-function mutations in familial atrial fibrillation. *Clin Chim Acta.* 2011;412(19-20):1825-1830. doi:10.1016/j.cca.2011.06.017

64. Wang J, Hu DY, Li XM, Xin YF, Wang LJ, Hu DY. [Genetic screening for novel GATA4 mutations associated with congenital atrial septal defect]. *Zhonghua Xin Xue Guan Bing Za Zhi.* 2010;38(5):429-434. [Chinese].

65. Li RG, Li L, Qiu XB, Yuan F, Xu L, Li X, et al. GATA4 loss-of-function mutation underlies familial dilated cardiomyopathy. *Biochem Biophys Res Commun.* 2013;439(4):591-596. doi:10.1016/j.bbr.2013.09.023

66. Sohelli J, Jallili Z, Rahbar M, Khatoooni Z, Mashayekhi A, Jafari H. Novel mutation of GATA4 gene in Kurdish population with nonsyndromic congenital heart septal defects. *Congent Heart Dis.* 2018;13(2):295-304. doi:10.1111/chd.12571

67. Posch MG, Perrot A, Schmitt K, Mittelmann S, Esseen EM, Stiller B, et al. Mutations in GATA4, NKX2.5, CRELID1, and BMP4 are infrequently found in patients with congenital cardiac septal defects. *Am J Med Genet A.* 2008;146A(2):251-253. doi:10.1002/ajmg.a.32042

68. Chen Y, Mao J, Sun Y, Zhang Q, Cheng HB, Yan WH, et al. A novel mutation of GATA4 in a familial atrial septal defect. *Clin Chim Acta.* 2010;411(21-22):1741-1745. doi:10.1016/j.cca.2010.07.021

69. El Malti R, Liu H, Doray B, Thuvin C, Malfret A, Dauphin C, et al. A systematic variant screening in familial cases of congenital heart defects demonstrates the usefulness of molecular genetics in this field. *Eur J Hum Genet.* 2016;24(2):228-236. doi:10.1038/ejhg.2015.105

70. Wang J, Fang M, Liu XY, Xin YF, Liu ZM, Chen XZ, et al. A novel GATA4 mutation responsible for congenital ventricular septal defects. *Int J Mol Med.* 2011;28(4):557-564. doi:10.3892/ijmm.2011.715

71. Garg V, Katriya IS, Barnes R, Schluterman MK, King IN, Butler CA, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature.* 2003;424(6947):443-447. doi:10.1038/nature01827

72. Sarkozy A, Conti E, Neri C, D’Agostino R, Digilio MC, Esposito G, et al. Spectrum of atrial septal defects associated with mutations of NKX2.5 and GATA4 transcription factors. *J Med Genet.* 2005;42(2):e16. doi:10.1136/jmg.2004.026740

73. Chen Y, Han ZQ, Yan WD, Tang CZ, Xie JY, Chen H, et al. A novel mutation in GATA4 gene associated with dominant inherited familial atrial septal defect. *J Thorac Cardiovasc Surg.* 2010;140(3):684-687. doi:10.1016/j.jtcvs.2010.01.013

74. Zhang X, Wang J, Wang B, Chen S, Fu Q, Sun K. A novel missense mutation of GATA4 in a Chinese family with...
Abbasi et al
J Cardiovasc Thorac Res, 2021, 13(4), 336-354

congenital heart disease. PLoS One. 2016;11(7):e0158904. doi:10.1371/journal.pone.0158904

88. D’Amato E, Giacopelli F, Giannattasio A, D’Annunzio G, Bocciardi R, Musso M, et al. Genetic investigation in an Italian child with an unusual association of atrial septal defect, attributable to a new familial GATA4 gene mutation, and neonatal diabetes due to pancreatic agenesis. Diabet Med. 2010;27(10):1195-1200. doi:10.1111/j.1464-5491.2010.03046.x

99. Pizzuti A, Sarkozy A, Newton AL, Conti E, Flex E, Digilio MC, et al. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. Hum Mutat. 2003;22(5):372-377. doi:10.1002/humu.10261

100. Garnatz AS, Gao Z, Broman M, Martinsen S, Earley JU, Svensson EC. FOG-2 mediated recruitment of the NuRD complex regulates cardiomyocyte proliferation during heart development. Dev Biol. 2014;395(1):50-61. doi:10.1016/j.ydbio.2014.08.030

101. Dinesh SM, Lingaiah K, Savitha MR, Krishnamurthy B, Narayannappa D, Ramachandra NB. GATA4 specific nonsynonymous single-nucleotide polymorphisms in congenital heart diseases patients of Mysore, India. Genet Test Mol Biomarkers. 2011;15(10):715-720. doi:10.1089/gtmb.2010.0278

102. De Luca A, Sarkozy A, Ferese R, Consoli F, Lepri F, Dentici ML, et al. New mutations in ZFPM2/FOG2 gene in tetralogy of Fallot and double outlet right ventricle. Clin Genet. 2011;80(2):184-190. doi:10.1111/j.1399-0004.2010.01523.x

103. Tan ZP, Huang C, Xu ZB, Yang JF, Yang YF. Novel ZFPM2/FOG2 variants in patients with double outlet right ventricle. Clin Genet. 2012;82(5):466-471. doi:10.1111/j.1399-0004.2011.01787.x

104. Qiao XH, Wang F, Zhang XL, Huang RT, Xue S, Wang J, et al. MEF2C loss-of-function mutation contributes to congenital heart defects. Int J Med Sci. 2017;14(11):1143-1153. doi:10.7150/ijms.21353

105. leda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell. 2010;142(3):375-386. doi:10.1016/j.cell.2010.07.002

106. Li FF, Deng X, Zhou J, Yan P, Zhao EY, Liu SL. Characterization of human bone morphogenetic protein gene variants for possible roles in congenital heart disease. Mol Med Rep. 2016;14(2):1459-1464. doi:10.3892/mmr.2016.5428

107. Song W, Wang H, Wu Q. Atrial natriuretic peptide in cardiovascular biology and disease (NPAA). Gene. 2015;569(1):1-6. doi:10.1016/j.gene.2015.06.029

108. Reamont-Buettner SM, Borlak J. HEY2 mutations in malformed hearts. Hum Mutat. 2006;27(1):118. doi:10.1002/humu.9390

109. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing hearts. Mol Biol Cell. 2011;22(5):532-546. doi:10.1091/mbc.e10-06-0483

110. Reamont-Buettner SM, Borlak J. TBX5 mutations in non-Holt-Oram syndrome (HOS) malformed hearts. Hum Mutat. 2004;24(1):104. doi:10.1002/humu.2952

111. Reamont-Buettner SM, Hecker H, Spanel-Borowski K, Craatz S, Kuenzel E, Borlak J. Novel NKX2-5 mutations in diseased hearts of patients with cardiac malformations. Am J Pathol. 2004;164(6):2117-2125. doi:10.1016/j.ajpath.2004.09.005

112. Reamont-Buettner SM, Borlak J. Somatic NKX2-5 mutations as a novel mechanism of disease in complex congenital heart disease. J Med Genet. 2004;41(9):684-690. doi:10.1136/jmg.2003.017483