An update on genetic variants of the NKX2-5

Jorge E. Kolomenski1,2 | Marisol Delea3 | Leandro Simonetti4 | Mónica C. Fabbro5 | Lucía D. Espeche3 | Melisa Taboas3 | Alejandro D. Nadra1,2 | Carlos D. Bruque3,6 | Liliana Dain2,3,6

1Departamento de Química Biológica Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, IQUIBICEN-CONICET, Buenos Aires, Argentina
2Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Instituto de Biociencias, Biotecnología y Biología Traslacional, IB3, Universidad de Buenos Aires, Buenos Aires, Argentina
3Centro Nacional de Genética Médica, ANLIS, Buenos Aires, Argentina
4Department of Chemistry-Biomedical Centre, Uppsala University, Uppsala, Sweden
5Laboratorio Novagen, Buenos Aires, Argentina
6Instituto de Biología y Medicina Experimental, (IBYME-CONICET), Buenos Aires, Argentina

Correspondence
Carlos D. Bruque and Liliana Dain, Centro Nacional de Genética Médica, ANLIS, Buenos Aires, Argentina. Avda. Las Heras 2670 3er piso, C.A.B.A. (1425), Argentina. Email: bruquecarlos@gmail.com (C. D. B.) and ldain@fbmc.fcen.uba.ar (L. D.)

Funding information
National Secretary of Health, Grant/Award Number: Health Scholarships Research “Dr. Abraam Sonis”; Secretaría de Ciencia y Tecnica, Universidad de Buenos Aires, Grant/Award Number: UBACYT 20020170100592BA; Fondo para la Investigación Científica y Tecnológica, Grant/Award Number: PID 2012-0060

Abstract
NKX2-5 is a homeodomain transcription factor that plays a crucial role in heart development. It is the first gene where a single genetic variant (GV) was found to be associated with congenital heart diseases in humans. In this study, we carried out a comprehensive survey of NKX2-5 GVs to build a unified, curated, and updated compilation of all available GVs. We retrieved a total of 1,380 unique GVs. From these, 970 had information on their frequency in the general population and 143 have been linked to pathogenic phenotypes in humans. In vitro effect was ascertained for 38 GVs. The homeodomain had the biggest cluster of pathogenic variants in the protein: 49 GVs in 60 residues, 23 in its third α-helix, where 11 missense variants may affect protein–DNA interaction or the hydrophobic core. We also pinpointed the likely location of pathogenic GVs in four linear motifs. These analyses allowed us to assign a putative explanation for the effect of 90 GVs. This study pointed to reliable pathogenicity for GVs in helix 3 of the homeodomain and may broaden the scope of functional and structural studies that can be done to better understand the effect of GVs in NKX2-5 function.

KEYWORDS
associated phenotypes, curated database, genetic variant evaluation, linear motif, NKX2-5

1 | BACKGROUND

NKX2-5 belongs to the NK2 subfamily of the NK homeobox gene family. It is a conserved homeodomain cardiac transcription factor that is present in animals from flies to humans (Bodmer, 1993; Harvey, 1996). It is one of the earliest cardiac transcription factors expressed in the heart and its expression is maintained through adulthood (Harvey 1996; Kasahara, Bartunkova, Schinke, Tanaka, & Izumo, 1998). It plays a crucial role in the development of the heart, regulating the proliferation, differentiation, and electrophysiological properties of cardiac cells. It is also expressed during thyroid (Fagman & Nilsson, 2011) and spleen development (Brendolan et al., 2005; Burn et al., 2008).

The canonical sequence of the NKX2-5 gene encodes the NKX2-5 protein of 324 residues. The three most conserved domains in the NK2 subfamily of proteins are the tinman (TN) domain, the homeobox domain (HD), and the NK2-specific domain (NK2-SD;
Chung & Rajakumar, 2016; Su et al., 2017). The TN domain was suggested to mediate the repressive activity of NKX2-5 (Elliott, Kirk, Schaft, & Harvey, 2010). The HD is a conserved helix-loop-helix domain with three α helices (Pradhan, Gopal, & Nam, 2014) that recognizes and binds to DNA and can homo- or hetero-dimerize (Elliott et al., 2010). The NK2-SD was suggested to function as an intramolecular transactivation regulator (Watada, Mirmira, Kalamaras, & German, 2000).

A number of other regions have been shown to have specific functions, like a tyrosine-rich region (YRR; also known as the tyrosine-rich domain or YRD), which serves as a dimerization interface (Bouveret et al., 2015; Elliott et al., 2006; Harvey, 1996; Liu et al., 2015), two putative nuclear localization signals (NLS; Kasahara & Izumo, 1999; Ouyang et al., 2016), a SUMOylation motif (Kim et al., 2011; Wang, Zhang, Iyer, Feng, & Schwartz, 2008), a phosphorylation site and an acetylation site, both located in the HD (p.S164 and p.K183, respectively) (Kasahara & Izumo, 1999; Li et al., 2007; Tang et al., 2016), an NKX2-5 box motif, which seems to be important for the transcriptional effect mediated by the C-terminal region of the protein (Elliott et al., 2006; Evans, 1999), and a GIREA motif, which is believed to play a role in protein interactions (Elliott et al., 2010; Evans, 1999).

NKX2-5 mostly binds to DNA as a homodimer or paired with TBX5 as a heterodimer (Luna-Zurita et al., 2016; Pradhan et al., 2016). In particular, it binds to a 5’-TNAAGTG-3’ motif (Chen & Schwartz, 1995; Dupays et al., 2015), as shown in its interaction with the NPPA gene promoter (Pradhan et al., 2012, 2016).

There are three reported alternative transcripts for this gene: two have shorter splice variants of exon 2 (NM_001166176.2 and NM_001166175.2) and the other one is a predicted isoform with no evidence of being transcribed (XM_017009071.2). There seems to be no further information on these alternative transcripts besides their first description by Shiojima et al. (1996).

Genetic defects in murine Nkx2-5 result in embryonic death and abnormal structure and function of the heart (Biben et al., 2000; Briggs et al., 2008; Choquet et al., 2018; Lyons et al., 1995; Terada et al., 2011; Wakimoto et al., 2003; Zakariyah et al., 2018). Moreover, NKX2-5 is the first gene where a single genetic variant (GV) was found to be associated with congenital heart disease (CHD) in humans (Muntean, Togânel, & Benedek, 2017; Schott et al., 1998). CHD encompasses a broad spectrum of anomalies and it is estimated to affect around 0.6–0.9% of all live births worldwide (Hoffman, 2013; Hoffman & Kaplan, 2002; van der Linde et al., 2011). It is the leading noninfectious cause of death in the first year of life (Huang et al., 2010; Zaidi & Brueckner, 2017) and the most frequent type of congenital disease (Huang et al., 2010; Su et al., 2017; Zaidi & Brueckner, 2017). Pathogenic GVs in NKX2-5 have been described in patients with atrioventricular conduction blocks (AVB), atrial septal defect (ASD), ventricular septal defect (VSD), tetrology of Fallot (ToF), hypoplastic left heart syndrome, double outlet right ventricle, arrhythmia, and sudden death, among others (Ellesøe et al., 2016; Reamon-Buettner & Borlak, 2010; Zakariyah, Raigara, Veinot, Skerjanc, & Burgon, 2017). In addition, NKX2-5 pathogenic GVs were detected in patients with thyroid ectopia or athyreosis (Dentice et al., 2006).

Despite several works reviewing GVs for NKX2-5 have been conducted (Reamon-Buettner & Borlak, 2010; Su et al., 2017), unaccounted information regarding pathogenicity can still be found distributed among several sources. Given the importance of the NKX2-5 gene, we decided to carry out a comprehensive survey of all available GVs, both in public databases and from the bibliography, to build a unified, curated, and updated compilation of the GVs of this gene.

2 | NKX2-5 DATABASE OF GVS

We compiled the information of GVs in NKX2-5 from six public databases, 381 published articles from scientific literature, and a cohort of 64 CHD patients from the Centro Nacional de Genética Médica (CNGM; Figure 1). The analysis was limited to the region encompassing the whole canonical transcript of NKX2-5 (chr5: 173, 232, 109–173, 235, 206; GRCh38/hg38). All of the compiled GVs were uploaded to the LOVD database.

The consulted databases were: NCBI’s dbSNP (https://www.ncbi.nlm.nih.gov/snp/) (Sherry et al., 2001), NHLBI-ESP’s EVS (https://evs.gs.washington.edu/EVS/), ExPASy’s SwissVar (https://swissvar.expasy.org/cgi-bin/swissvar/home/) (Mottaz, David, Veuthey, & Yip, 2010), ExAC

![FIGURE 1](https://www.wiley.com/legacy/legacy/1188-KOLOMENSKI-ET-AL.png)
The methods with which data was retrieved from the different databases depended on the options offered by each one of them. All of the databases were then formatted and stored into the same fields to make them easily comparable.

The 381 published articles were obtained from PubMed on August 2018 by searching for the different names for the gene in GeneCards, the HGVS identification standard (den Dunnen et al., 2016), and LOVD (Fokkema et al., 2011). All of the variants were gathered from the public databases on July 2018 and a second batch of variant was added from LOVD and ClinVar on July 2019. One specific CHD database was also evaluated (ACGV; www.cardiobdb.org/acgv/) (Walsh et al., 2017), but GVs were not found for NKX2-5.

The identifier was built in three parts: a reference sequence or scaffold, the position of the GV within that scaffold, and the variation as defined by the HGVS convention system. Our scaffold was the build 38, patch 12 of the Genome Reference Consortium (GRC) human genome (GRCh38.p12). As of July of 2018, this is the standard reference assembly sequence used by the NCBI. All but two of the databases already had their GVs with positions relative to it. Lastly, some variations were manually curated because they did not follow HGVS standards, as was the case for some deletions, insertions, and duplications and some of the data from the scientific literature (Supporting Information Text S1). In parallel, the complementary DNA and protein references had the same scaffold through all sources (NM_004387.3 and NP_004378.1, respectively), so they only needed minor adjustments to be consistent with HGVS standards (Supporting Information Text S1).

We extracted information on frequencies of the variants from general databases (GDBs): dbSNP, EVS, and ExAC. In parallel, we used the VarAFT system (http://varaft.eu) (Desvignes et al., 2018) for variant annotation of all GVs in the compiled database (Supporting Information Text S1).

We further compiled information regarding the relationship to a disease-associated phenotype, allele origin, and experimental evidence for each variant, either from the scientific literature or from LOVD, ClinVar, and SwissVar (Supporting Information Text S1).

In silico analysis were performed using PDBsum (Laskowski et al., 1997) to obtain information on predicted protein–DNA interactions within the HD of NKX2-5. All of this information was studied and presented using UCSF Chimera (Pettersen et al., 2004).

Lastly, information about functional sites in the NKX2-5 protein (Uniprot ID P52952) was retrieved from the scientific literature (posttranslational modification sites, NLS, etc.). This information was matched to predicted linear motifs from the eukaryotic linear motif resource (http://elm.eu.org) (Gouw et al., 2018) to better define their exact location in the protein.

3 | VARIANTS OF NKX2-5

Each of the consulted sources contained different numbers of, and information about, NKX2-5 GVs. A vast majority of the variants were found in the dbSNP database, but every single one of the sources had at least one GV that was unique to them (data not shown).

The compiled database has a total of 1,380 unique GVs (Supporting Information Table S1); 1,314 of these variants come from the different databases, which accumulated 1,706 GVs before curating repeated variants. In addition, 170 GVs were found among 97 of the 381 publications reviewed for the present work; 64 of these GVs were not found in any of the online databases. Lastly, the data obtained from the cohort from CNGM provided two novel GVs out of the 11 found (Supporting Information Table S2).

Frequency information was retrieved for 970 of the variants. Frequencies analysis showed that only 26 GVs are present in more than 1/100 individuals, and 131 between 1/100 and 1/10,000. Therefore, 83.8% of these GVs are present in <1/10,000 individuals.

3.1 | Types of GV in NKX2-5

From a total of 1,380 unique variants, 626 were found in the translated region (TR) and 754 in the nontranslated region (NTR; including the 5′ and 3′ untranslated regions and introns; Table 1).

A total of 389 out of 626 (62.1%) GVs located in the TR were missense variants, of which 100 were classified as pathogenic. We also found 169 (27%) synonymous variants, none of which were linked with human disease. In addition, we found 20 (3.2%) nonsense, 33 (5.3%) frameshift, and 1 (0.2%) stop loss variations, 41 of which were reported with putative effects in the function of the protein and thus with a possible implication in human health. Furthermore, we counted a total of 14 (2.2%) in-frame variants: seven deletions, six duplications, and one insertion-deletion. No in-frame variants have been implicated with disease nor predicted to cause an effect on the protein.

From the 754 GVs located in the NTR, two were classified as pathogenic: c.(334+1G>T) and c.(335-1G>T). Both variants were located in consensus splicing sites.

3.2 | Distribution of GVs in NKX2-5

The distribution of all variants compiled for the NKX2-5 gene was slightly denser in the TR of both exons of the canonical isoform of the
protein (Figure 2a, blue lines). This may correspond with the fact that most of the GVs come from studies focusing on coding regions.

There was also an accumulation of variants in a small region within the intron (chr5: 173, 233, 490–173, 233, 550). This is located close to the TR in exon 2 of one of the isoforms (NM_001166175.2) and consists mostly of GVs from GDBs. Other than that, GVs retrieved from GDBs do not seem to have a preferential distribution in the gene (Figure 2a, green lines).

Finally, there was an accumulation of variants classified as pathogenic in the TR of exon 2 but there are no confirmed pathogenic variants within the NTR except for those in the canonical splice acceptor/donor sites (Figures 2a,b, red lines).

### 4 | EVALUATION OF PATHOGENICITY IN NKX2-5 GVs

Information on the pathogenicity of the variants was obtained from the LOVD, ClinVar, and SwissVar databases and the scientific literature for a total of 163 GVs. Variants classified as “Pathogenic,” “Likely benign,” or “Conflicting evidence,” as well as the two novel variants found in patients from the cohort at CNGM, were also classified according to the American College of Medical Genetics and Genomics’ (ACMG) standards (Richards et al., 2015).

Taking into account scientific reports and clinical databases, a total of 143 variants have been grouped as pathogenic (Table 2 and Figure 2), six as conflicting evidence and 14 as likely benign in the compiled database (Table S1). From the 143 pathogenic GVs, 126 have been implicated in human disease in the scientific literature and 17 others have been predicted to be pathological according to LOVD/ClinVar. Of note, 42 of these GVs have not been deposited in any of the public databases consulted (Table 2, in bold). Following the ACMG recommendations, 107 variants were classified as pathogenic or likely pathogenic, 15 were classified as benign or likely benign, and 43 were classified as variants of uncertain significance (VUS), including the two novel variants from the CNGM cohort, one synonymous change and one intronic change (Tables 2 and S1). Allele origin was ascertained for 108 GVs, from which 65 were familial and 10 de novo (Table S1).

In total, for 118 GVs the effects in pathogenicity were coincident, as either benign/likely benign or pathogenic/likely pathogenic. Only two GVs, p.(Pro275Thr) and p.(Ala42Pro), have discordant classifications, both being classified as pathogenic in the scientific literature and predicted to be likely benign following ACMG standards. Of note, other two GVs, p.(Lys183Asn) and p.Cys270Tyr, which have conflicting evidence in the scientific literature, were classified as likely pathogenic and likely benign, respectively, using the ACMG criteria. Lastly, 35 GVs classified as pathogenic in the scientific literature were classified as VUS according to the ACMG criteria.

For seven pathogenic GVs in the NKX2-5 gene, one or more variants in other relevant genes were concomitantly found in patients (Table S3). One of the concomitant variants (c.1349G>A, p.Arg450His in TSHR found with c.872A>T, p.(Asn291Ile) in NKX2-5) was observed in a patient with congenital hypothyroidism and predicted to be likely pathogenic by the Varsome online tool (https://varsome.com/). Other two were predicted to be VUS (for two different NKX2-5 GVs) and six were predicted to be benign/likely benign (for four different NKX2-5 GVs).

It should be noted that 32 out of the 143 pathogenic GVs have also been found in GDBs. Nonetheless, 24 of these GVs had a frequency of <1/10,000 and 7 between 1/10,000 and 1/100, leaving only one GV with a frequency of 1.07/100 (Table 2). Considering the classification following the ACMG guidelines, four out of these last eight variants were classified as VUS and two as likely benign.

### Table 1 | Types of genetic variants in different regions/domains of the NKX2-5 gene

| Region       | Missense | Synonymous | Nonsense | Inframe | Frameshift | Stop loss | Total |
|--------------|----------|------------|----------|---------|------------|-----------|-------|
| Translated region |          |            |          |         |            |           |       |
| Tinman       | 10 (4)   | 6 (0)      | 0 (0)    | 0 (0)   | 1 (0)      | 0 (0)     | 17 (4) |
| HD Helix 1   | 18 (4)   | 11 (0)     | 1 (1)    | 0 (0)   | 2 (2)      | 0 (0)     | 32 (7) |
| Helix 2      | 13 (4)   | 6 (0)      | 2 (2)    | 1 (0)   | 2 (2)      | 0 (0)     | 24 (8) |
| Helix 3      | 28 (18)  | 11 (0)     | 4 (4)    | 0 (0)   | 1 (1)      | 0 (0)     | 44 (23)|
| Loops        | 26 (9)   | 13 (0)     | 2 (1)    | 0 (0)   | 1 (1)      | 0 (0)     | 42 (11)|
| Total        | 85 (35)  | 41 (0)     | 9 (8)    | 1 (0)   | 6 (6)      | 0 (0)     | 142 (49)|
| NK2-SD       | 27 (4)   | 9 (0)      | 1 (0)    | 4 (0)   | 2 (1)      | 0 (0)     | 43 (5) |
| YRR          | 44 (5)   | 10 (0)     | 7 (7)    | 0 (0)   | 5 (4)      | 0 (0)     | 66 (16)|
| NKX2-5 box   | 19 (3)   | 9 (0)      | 0 (0)    | 1 (0)   | 1 (0)      | 0 (0)     | 30 (3) |
| GIRAW motif  | 6 (3)    | 2 (0)      | 0 (0)    | 0 (0)   | 0 (0)      | 0 (0)     | 8 (3)  |
| Others       | 198 (46) | 92 (0)     | 3 (2)    | 8 (0)   | 18 (12)    | 1 (1)     | 320 (61)|
| Total        | 389 (100)| 169 (0)    | 20 (17)  | 14 (0)  | 33 (23)    | 1 (1)     | 626 (141)|

Untranslated region | 754 (2) | 0 (0) |
Total: 1,380 (143)

Note: Number and type of unique genetic variants in the compiled database, grouped by protein region/domain. In parenthesis, number of pathogenic variants.

Abbreviations: HD, homeodomain; NK2-SD, NK2-specific domain; YRR, tyrosine-rich region.
Lastly, cardiac diseases are the most frequent patient phenotypes, present in 134 of the pathogenic GVs. Within cardiac phenotypes, the most common subtypes are ASD (78), followed by VSD (55), AVB (47), atrioventricular septal defect (AVSD) (18), and ToF (14) (Table 2). Other extracardiac phenotypes include thyroid dysgenesis, heterotaxy, asplenia, and polysplenia.

5 | LOCATION OF PATHOGENIC GVS

Among the pathogenic GVs, 141 are located within the canonical TR: 23 frameshift, 17 nonsense, 100 missense, and 1 stop loss. The two remaining GVs occurring in the NTR correspond to the two splice site variants. Figure 2b depicts the 141 pathogenic variants located within the TR of the NKX2.5 protein.

No variants in the canonical NTR except for GVs on the splice donor/acceptor sites were found that had been classified as pathogenic. Thus, no pathogenic variants were found in the alternative exons of any of the putative isoforms. Moreover, the fact that no synonymous nor in-frame variants have been found to be pathogenic restricts all pathogenic GVs to being missense, frameshift, nonsense, stop loss, or splice acceptor/donor variants (Table 1).

5.1 | Pathogenic GVs in different protein motifs

The analysis of the distribution of pathogenic variants along different identified regions of the protein shows that the TN domain (residues, 10–21) has four variants, the HD (residues, 138–197) has 49 GVs, and the NK2-SD (residues, 212–234) has five GVs. These account for 58 pathogenic GVs from a total of 141 in the TR. In addition, 16 pathogenic GVs are located in the YRR (residues, 237–274), three in the NKX2-5 box (residues, 291–304), and three in the GIRAW motif (residues, 320–324), leaving 61 GVs in the rest of the protein (Table 2 and Figure 2b).
| cDNA variant | Protein variant | Affected region | Frequency (per 1,000) | Associated phenotype | ACMG classification | In vitro effect | References |
|--------------|----------------|----------------|-----------------------|----------------------|---------------------|----------------|------------|
| c.973T>C     | p.(*325Ginext*18) | GIRAW         |                       | ASD, VSD, AVSD       | VUS                 | impaired       | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.967G>A     | p.(Ala323Thr)   | GIRAW         |                       | ToF                  | LP                  |                | McElhinney et al. (2003) |
| c.965G>A     | p.(Arg322Gln)   | GIRAW         |                       | VSD                  | LP                  |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.958G>A     | p.(Gly320Ser)   | GIRAW         |                       | ASD, VSD, AVSD       | LP                  |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.943G>T     | p.Val315Leu     | 0.035         | ToF                   | VUS                  | No effect           |                | Rauch et al. (2016) |
| c.919G>A     | p.(Gly307Arg)   | VSD, PS       |                       | VUS                  |                     |                | Pulignani et al. (2018) |
| c.913A>G     | p.(Ser305Gly)   | VSD           |                       | VUS                  |                     |                |                       |
| c.896A>G     | p.(Asp299Gly)   | NKX2-5 box    |                       | ASD, VSD, AVSD       | LP                  |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.880A>C     | p.(Asn294His)   | NKX2-5 box    |                       | AVSD                 | LP                  |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.872A>T     | p.(Asn291Ile)   | NKX2-5 box    | TD                    | VUS                  |                     |                | Wang et al. (2017) |
| c.857C>T     | p.(Ala286Val)   | ASD, VSD, AVSD |                       | VUS                  |                     |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.848C>A     | p.(Pro283Gln)   | 0.06          | VSD, PDA, CoA         | VUS                  |                     |                | Peng et al. (2010) |
| c.842C>T     | p.(Ala281Val)   | 4.15          | ASD, VSD, AVSD        | VUS                  |                     |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.839C>T     | p.(Pro280Leu)   | 0.059         | IAVC                  | VUS                  |                     |                | Esposito et al. (2009) |
| c.836C>T     | p.(Ser279Phe)   | 0.004         | VSD, AVSD             | LP                   |                     |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.837T>C     | p.(Ser279Pro)   | 10.74         | VSD                   | VUS                  |                     |                | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.823C>A     | p.(Pro275Thr)   | 0.181         | CoA                   | LB                   |                     |                | McElhinney et al. (2003) |
| c.795C>A     | p.(Ser265Arg)   | Tyr-Rich      | TD                    | P                    | TA                  |                | Hermanns et al. (2011) |
| c.792C>A     | p.(Cys264*)     | Tyr-Rich      |                       | ASD, AVB             | P                   |                | Ikeda et al. (2002) |
| c.783del     | p.(Ala262Argfs*32) | Tyr-Rich    |                       | AVB                  | P                   |                | Guntheroth et al. (2012) |
| c.777C>G     | p.(Tyr259*)     | Tyr-Rich      |                       | P                    |                     |                | Predicted by ClinVar |
| cDNA variant | Protein variant | Affected region | Frequency (per 1000) | Associated phenotype | ACMG classification | In vitro effect impaired | References |
|--------------|-----------------|-----------------|---------------------|----------------------|---------------------|------------------------|------------|
| c.777C>A     | p.Tyr259*       | Tyr-Rich        |                     | ASD, VSD, DORV, AVB  | P                   | TA                     | Benson et al. (1999); Kasahara et al. (2000) |
| c.769C>G     | p.(Pro257Ala)   | Tyr-Rich        |                     |                       |                     |                        | Chen et al. (2010) |
| c.768T>G     | p.(Tyr256*)     | Tyr-Rich        |                     | ASD                   | P                   |                        | Predicted by ClinVar |
| c.768T>A     | p.(Tyr256*)     | Tyr-Rich        |                     | ASD, AVB, MVP        | P                   |                        | Gutierrez-Roelens et al. (2004) |
| c.762del     | p.(Ala255Profs*39) | Tyr-Rich      |                     | ASD, AVB             | P                   |                        | Stallmeyer et al. (2010) |
| c.742T>C     | p.(Tyr248His)   | Tyr-Rich        |                     | VSD                  | VUS                 |                        | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.738T>A     | p.(Asn246Lys)   | Tyr-Rich        |                     | CHD                   | LP                  |                        | Tong (2016) |
| c.721,728del | p.(Tyr241Glyfs*8) | Tyr-Rich         |                     | ASD, AVB             | P                   |                        | Abou Hassan et al. (2015) |
| c.720,726del | p.(Tyr241Trps*51) | Tyr-Rich       |                     | ASD, AVB, VT, DCM, SD | P                   |                        | El Malti et al. (2016) |
| c.723C>G     | p.(Tyr241*)     | Tyr-Rich        |                     |                       |                     |                        | Predicted by LOVD |
| c.711C>A     | p.(Tyr237*)     | Tyr-Rich        |                     | ASD, VF, DCM, NCC    | P                   |                        | Predicted by ClinVar |
| c.709T>C     | p.(Tyr237His)   | Tyr-Rich        |                     | ASD, AVB, VT, DCM, SD | VUS                 |                        | El Malti et al. (2016) |
| c.707C>A     | p.Pro236His     | 0.009           | AP                  | AP                   | P                   | TA                     | Koss et al. (2012) |
| c.701,702insT-CCCT | p.(Ala235Profs*61) | ASD, AVB     |                     |                       |                     |                        | McElhinney et al. (2003) |
| c.694G>A     | p.(Gly232Arg)   | NK2-SD          | 0.016               | PS                   | VUS                 |                        | Granados-Riveron et al. (2012) |
| c.685_686dup | p.(Cys230Hisfs*3) | NK2-SD         |                     |                       |                     |                        | Predicted by ClinVar |
| c.676G>A     | p.(Asp226Asn)   | NK2-SD          | 0.006               | VSD                  | LP                  |                        | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.656C>T     | p.(Ala219Val)   | NK2-SD          | 0.008               | VSD, ToF             | P                   |                        | Goldmuntz, Geiger, and Benson (2001); McElhinney et al. (2003); Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.646C>T     | p.(Arg216Cys)   | NK2-SD          | 0.004               | ToF                  | P                   |                        | Goldmuntz et al. (2001); McElhinney et al. (2003) |
| c.626C>T     | p.(Pro209Leu)   | NK2-SD          |                     | ASD                  | LP                  |                        | Wang et al. (2011a) |

(Continues)
| cDNA variant | Protein variant | Affected region | Frequency (per 1,000) | Associated phenotype | ACMG classification | In vitro effect impaired | References |
|--------------|-----------------|----------------|----------------------|----------------------|---------------------|------------------------|------------|
| c.618del     | p.(Leu207Cysfs*25) | ASD            | P                    |                      |                     |                        | Abou Hassan et al. (2015) |
| c.614T>A     | p.(Val205Glu)    | VSD            | VUS                  |                      |                     |                        | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.592C>T     | p.Gln198*        | ASD, AVB, SD   | P                    |                      | TA*, DB             |                        | Hosoda et al. (1999); Kasahara et al. (2003); Schott et al. (1998); Zhu et al. (2000) |
| c.581A>G     | p.(Lys194Arg)    | HD (H3)        | VSD                  | LP                   |                     |                        | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.575A>G     | p.(Lys192Arg)    | HD (H3)        | VSD                  | LP                   |                     |                        | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.575A>C     | p.(Lys192Thr)    | HD (H3)        | VSD                  | LP                   |                     |                        | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.574A>T     | p.Lys192*        | HD (H3)        | ASD, AVS, AVB, AF, BAV | P                    | TA                  | Qu et al. (2014)       |
| c.572A>G     | p.Tyr191Cys      | HD (H3)        | ASD, VSD, AVB        | P                    | TA, DB              | Benson et al. (1999); Costa et al. (2013); Kasahara and Benson (2004); Kasahara et al. (2000) |
| c.569G-T     | p.(Arg190Leu)    | HD (H3)        | ASD, AVB             | LP                   |                     | Stallmeyer et al. (2010) |
| c.569G>A     | p.Arg190His      | HD (H3)        | ASD, VSD, AVB        | P                    | TA, DB, D*          | Kasahara and Benson (2004) |
| c.568C>T     | p.(Arg190Cys)    | HD (H3)        | 0.008                | ASD, AVB             | P                   | Hirayama-Yamada et al. (2005); Matsuoka (2005) |
| c.566G>C     | p.(Arg189Pro)    | HD (H3)        | ASD                  | LP                   |                     | Predicted by ClinVar   |
| c.565C>G     | p.Arg189Gly      | HD (H3)        | ASD, TVA, AVB, AF    | P                    | TA, DB              | Benson et al. (1999); Kasahara and Benson (2004); Kasahara et al. (2000) |
| c.564C>A     | p.Asn188Lys      | HD (H3)        | ASD, EA, TVA, AVB    | P                    | TA, DB              | Benson et al. (1999); Kasahara and Benson (2004); Kasahara et al. (2000) |
| c.561G>C     | p.Gln187His      | HD (H3)        | ASD, AVB, AVR        | P                    | TA, DB, D**         | Gutierrez-Roelens et al. (2002); Kasahara and Benson (2004) |
| c.559C>T     | p.(Gln187*)      | HD (H3)        | VSD                  | P                    |                     | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.557T>C     | p.Phe186Ser      | HD (H3)        | ASD, AVB, AF         | P                    | TA                  | Xie et al. (2013)      |
| cDNA variant | Protein variant | Affected region | Frequency (per 1000) | Associated phenotype | ACMG classification | In vitro effect impaired | References |
|-------------|-----------------|-----------------|----------------------|----------------------|---------------------|-------------------------|------------|
| c.554_555insC | p.(Trp185Cysfs*67) | HD (H3) | | | P | TA, DB stability<sup>a</sup> | Costa et al. (2013); Hanley et al. (2016) |
| c.555G>A | p.(Trp185*) | HD (H3) | | ASD, VSD, AVB | P | | El Malti et al. (2016) |
| c.554G>T | p.(Trp185Leu) | HD (H3) | | ASD, VSD, AVB, VNC, MVP | P | | Sarkozy et al. (2005) |
| c.552C>G | p.Ile184Met | HD (H3) | | ASD, TCA, CD, VNC, DCM, PFO | P | | |
| c.547A>G | p.(Lys183Glu) | HD (H3) | | ASD, AVSD | LP | | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004); Tang et al. (2016) |
| c.543G>C | p.(Gln181His) | HD (H3) | | ASD, AVSD, CoA, AVB, SD | LP | | |
| c.541C>T | p.Gln181* | HD (H3) | | ASD, AVB | P | | TA | Xu et al. (2017b) |
| c.538A>G | p.Thr180Ala | HD (H3) | | AF | P | | TA | Xu et al. (2017b) |
| c.536C>T | p.(Ser179Phe) | HD (H3) | | ASD, VSD, AVB | P | | Liu et al. (2009a); Liu et al. (2009b, 2011) |
| c.533C>T | p.Thr178Met | HD | | ASD, VSD, HLHS, AVB, SSS | P | | TA, DB | Elliott et al. (2003); Kasahara and Benson (2004, 2006); Schott et al. (1998, 2000); Zhu and Li (2000); Reamon-Buettnet and Borlak (2004); Reamon-Buettner et al. (2004); Hirayama-Yamada et al. (2003); Matsuoka (2005) |
| c.512T>G | p.(Leu171Arg) | HD (H2) | | CHD | LP | | Predicted by ClinVar |
| c.512T>C | p.Leu171Pro | HD (H2) | | ASD, VSD, TVA, AVB | P | | TA, DB, D<sup>+</sup> | Kasahara and Benson (2004) |
| c.510_511dup | p.Leu171Argfs*6 | HD (H2) | | ASD, AVB, SD | P | | TA, localization<sup>c</sup> | Ouyang et al. (2011) |
| c.509A>C | p.(Gln170Pro) | HD (H2) | | ASD, VSD, AVB, MVP | LP | | | El Malti et al. (2016) |
| c.508C>T | p.Gln170* | HD (H2) | | ASD, AVB, SD | P | | TA, DB, D<sup>**</sup> | Hatemi et al. (2011); Kasahara et al. (2003); Schott et al. (1998); Zhu et al. (2000) |
| c.499G>T | p.Glu167* | HD (H2) | | ASD, AVB, DCM | P | | | Xu et al. (2017a) |
| c.499G>A | p.(Glu167Lys) | HD (H2) | | ASD, VSD, PA, VNC | LP | | | Bermúdez-Jiménez, Jiménez-Jáime, and López-Fernández (2017) |
| cDNA variant | Protein variant | Affected region | Frequency (per 1,000) | Associated phenotype | ACMG classification | In vitro effect impaired | References |
|--------------|-----------------|-----------------|------------------------|----------------------|---------------------|------------------------|------------|
| c.498dup     | p.(Glu167Argfs*85) | HD (H2)         |                        | ASD, AVB             | P                   |                        | Sarkozy et al. (2005) |
| c.491C>A     | p.(Ser164*)     | HD              |                        | CHD                  | P                   |                        | Predicted by ClinVar |
| c.488T>G     | p.(Leu163Arg)   | HD              | 4.43                   | ASD, AVB, AF         | LP                  |                        | El Malti et al. (2016) |
| c.482G>C     | p.Arg161Pro     | HD              | 0.01                   | TD                   | P                   | TA, DB                 | Dentice et al. (2006) |
| c.478del     | c.470del       | HD              |                        | CHD                  | P                   |                        | Predicted by ClinVar |
| c.479A>C     | p.(Gln160Pro)   | HD              |                        | ASD, AVB             | P                   |                        | Rifai, Maazouzi, and Sefiani (2007) |
| c.471del     | p.(Phe157Leufs*94) | HD (H1)        |                        | CHD                  | P                   |                        | Predicted by ClinVar |
| c.461A>G     | p.(Glu154Gly)   | HD (H1)         |                        | ASD                  | LP                  |                        | Abou Hassan et al. (2015) |
| c.448G>A     | p.(Val150Ile)   | HD (H1)         | 0.043                  | VSD                  | VUS                 |                        | De Luca et al. (2010) |
| c.445C>T     | p.Gln149*       | HD (H1)         | 0.003                  | ASD, VSD, ToF, AVB   | P                   | TA, DB, D              | Benson et al. (1999); Bjørnstad and Leren (2009); Kasahara et al. (2000) |
| c.443del     | p.(Ala148Glyfs*28) | HD (H1)        |                        | P                    |                     |                        | Predicted by ClinVar |
| c.443C>A     | p.(Ala148Glu)   | HD (H1)         |                        | ASD, VSD             | LP                  |                        | Predicted by ClinVar |
| c.437C>G     | p.Ser146Trp    | HD (H1)         |                        | AVB, AF, DCM, SD     | P                   | TA                     | Yuan et al. (2015) |
| c.434T>C     | p.(Phe145Ser)  | HD              | 0.013                  | AF                   | P                   | TA                     | Huang et al. (2013); Ritchie et al. (2012) |
| c.431T>C     | p.(Leu144Pro)  | HD              | 4.65                   | ASD, AVSD            | VUS                 |                        | Reamond-Buettner and Borlak (2004); Reamond-Buettner et al. (2004) |
| c.424C>T     | p.Arg142Cys    | HD              |                        | ASD, VSD, ToF, PDA, PS, AVB | P | TA, DB, D*  | Gutierrez-Roelens et al. (2002); Kasahara and Benson (2004) |
| c.421C>G     | p.(Pro141Ala)  | HD              | 0.007                  | ASD, VSD, AVSD       | VUS                 |                        | El-Bouchikhi et al. (2017) |
| c.415A>T     | p.Arg139Trp    | HD              |                        | ASD, AVB, DCM        | P                   | TA                     | Xu et al. (2017a) |
| c.403G>A     | p.(Ala135Thr)  | HD              |                        | ASD, AVSD            | VUS                 |                        | Reamond-Buettner and Borlak (2004); Reamond-Buettner et al. (2004) |
| c.397_400del | p.(Pro133Glyfs*42) | AVSD, DORV, AVR, HT, AP | LP |                        |                    |                        | Reamond-Buettner and Borlak (2004); Reamond-Buettner et al. (2004) |
| c.397C>T     | p.(Pro133Ser)  | HD              | VSD                   | VUS                 |                     |                        |                        |
| cDNA variant     | Protein variant       | Affected region | Frequency (per 1,000) | Associated phenotype | ACMG classification | In vitro effect impaired | References                                |
|------------------|-----------------------|-----------------|-----------------------|----------------------|---------------------|-------------------------|------------------------------------------|
| c.391G>A         | p.(Glu131Lys)         | ASD             |                       | VUS                  |                     |                         | Predicted by ClinVar                   |
| c.380C>A         | p.(Ala127Glu)         | ASD             |                       | VUS                  |                     |                         | McElhinney et al. (2003)                |
| c.377A>T         | p.(Glu126Val)         | ASD, VSD, AVSD  |                       | VUS                  |                     |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.375dup         | p.(Glu126Argfs*27)    | ASD             |                       | P                    |                     |                         | Predicted by ClinVar                   |
| c.371A>G         | p.(Lys124Arg)         | VSD             |                       | VUS                  |                     |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.365T>C         | p.(Leu122Pro)         | 0.008           | ASD                   | VUS                  |                     |                         | Granados-Riveron et al. (2012)          |
| c.356C>A         | p.Ala119Glu           | 0.025           | AVSD                  | VUS                  | TA                  |                         | Reamon-Buettner et al. (2013)           |
| c.353A>G         | p.(Lys118Arg)         | ASD, VSD        |                       | VUS                  |                     |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.351G>C         | p.(Gln117His)         | ToF             |                       | LP                   |                     |                         | Pulignani et al. (2018)                |
| c.340T>C         | p.(Cys114Arg)         | ASD, VSD, AVSD  |                       | LP                   |                     |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.340T>A         | p.(Cys114Ser)         | ASD, AVSD       |                       | LP                   |                     |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.335-1G>T       | p.?                   | ASD, VSD        |                       | P                    |                     |                         | Predicted by ClinVar                   |
| c.334+1G>T       | p.?                   | AVB             |                       | P                    |                     | No accumulation in the cell | Benson et al. (1999); Kasahara et al. (2000) |
| c.326A>T         | p.(Glu109Val)         | VSD             |                       | LP                   |                     |                         | Wang et al. (2011a)                    |
| c.325G>T         | p.(Glu109*)           | 5.84            | ASD, VSD, PS, AVB, PFO| P                   |                     |                         | Akçaboy et al. (2008)                  |
| c.313del         | p.(Asp105Thrfs*71)    | ASD, AVB        |                       | P                    |                     |                         | König et al. (2006)                   |
| c.262del         | p.(Ala88Profs*88)     | ASD, AVB        |                       | P                    |                     |                         | Hirayama-Yamada et al. (2005); Matsuoka (2003) |
| c.244T>A         | p.Cys82Ser            | 0.017           | IAVC                  | VUS                  | TA\(^d\)           |                         | Esposito et al. (2009)                 |
| c.230C>T         | p.(Pro77Leu)          | VSD             |                       | VUS                  |                     |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.228_229del     | p.(Pro77Phefs*30)     | ASD, AVB        |                       | P                    |                     |                         | Watanabe et al. (2002)                |

(Continues)
| cDNA variant | Protein variant | Affected region | Frequency (per 1,000) | Associated phenotype | ACMG classification | In vitro effect impaired | References |
|-------------|----------------|----------------|----------------------|----------------------|---------------------|--------------------------|------------|
| c.215_221del | p.(Glu72Alafs*102) | ASD, AVB, AF, HT, PP | | | P | | Watanabe et al. (2002) |
| c.214G>A | p.(Glu72Lys) | ASD | VUS | | | | Mattapally et al. (2018); Tian et al. (2008) |
| c.206T>C | p.(Leu69Pro) | VSD | VUS | | | | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.202G>A | p.(Glu68Lys) | ASD | P | | | | Tian et al. (2008) |
| c.188C>T | p.(Ala63Val) | 0.03 | L-TGA | VUS | | | McElhinney et al. (2003) |
| c.175C>G | p.Pro59Ala | VSD | P | TA | | | Wang et al. (2011c) |
| c.147_163delAGCT | p.(Ala50Profs*123) | ASD | P | | | | Predicted by ClinVar |
| c.160G>A | p.(Glu54Lys) | ToF | LP | | | | Wang, Liu, and Yang (2011b) |
| c.151T>C | p.(Phe51Leu) | 0.006 | VSD | VUS | | | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.126_142del | p.(Pro43Glyfs*59) | ASD, AVB | P | | | | Liu et al. (2011) |
| c.138C>G | p.(Cys46Trp) | ASD, VSD, AVB | LP | | | | Liu et al. (2011) |
| c.133T>C | p.(Ser45Pro) | 0.009 | VSD | VUS | | | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.124G>C | p.(Ala42Pro) | 0.33 | EA | LB | | | Gioli-Pereira et al. (2010) |
| c.112del | p.(Glu38Argfs*138) | ASD, VSD, AVB | P | | | | Ellesøe et al. (2016) |
| c.106C>A | p.(Arg36Ser) | VSD | LP | | | | Liu et al. (2009b); Wang et al. (2011a) |
| c.95A>T | p.(Glu32Val) | VSD, ToF | P | | | | Khatami et al. (2018) |
| c.94G>A | p.(Glu32Lys) | 0.017 | ASD | P | | | Tian et al. (2008) |
| c.65A>G | p.(Gln22Arg) | 0.159 | ASD | VUS | | | Draus et al. (2009) |
| c.65A>C | p.(Gln22Pro) | 0.039 | ToF | LP | | | McElhinney et al. (2003) |
| c.64C>A | p.(Gln22Lys) | ASD | P | | | | Wang et al. (2011b) |
| c.56A>G | p.(Asn19Ser) | TN | VSD | LP | | | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.55A>G | p.Asn19Asp | TN | AF | P | TA | | Xie et al. (2013) |
| cDNA variant | Protein variant | Affected region | Frequency (per 1000) | Associated phenotype | ACMG classification | In vitro effect impaired | References |
|--------------|-----------------|-----------------|----------------------|----------------------|---------------------|-------------------------|------------|
| c.46G>A      | p.(Asp16Asn)    | TN              | 0.006                | VSD                  | VUS                 |                         | Mattapally et al. (2018) |
| c.44A>T      | p.(Lys15Ile)    | TN              | 0.013                | ASD                  | LP                  |                         | McElhinney et al. (2003) |
| c.20T>C      | p.(Leu7Pro)     | AVSD            |                      | VUS                  |                      |                         | Reamon-Buettner and Borlak (2004); Reamon-Buettner et al. (2004) |
| c.17C>T      | p.Ala6Val       | ToF             |                      |                      | P                   | TA                      | Kodo et al. (2012) |

Note: Each variant is noted with their "c." and "p." descriptors (Ref-seq: NM_004378.1 and NP_004378.1, respectively), along with its location in a particular protein region/domain. Genetic variants with no associated phenotype were predicted by either ClinVar or LOVD to be pathogenic but no patient phenotype was reported. In bold, genetic variants do not present in any of the consulted databases. Abbreviations: ACMG, American College of Medical Genetics; AF, atrial fibrillation; AP, asplenia; ASD, atrial septal defect; AVB, atrioventricular block; AVR, anomalous venous return; AVS, aortic valve stenosis; AVSD, atroventricular septal defect; BAV, bicuspid aortic valve; CD, conduction defects; CHD, congenital heart disease; CoA, coarctation of the aorta; DCM, dilated cardiomyopathy; D, dimerization (tested with the wild-type counterpart); DB, DNA binding; DORV, double outlet right ventricle; EA, Ebstein's anomaly; GV, genetic variant; HD, homeodomain; H1-3, α-helices 1–3; HLHS, hypoplastic left heart syndrome; HT, heterotaxy; IAVC, isolated accessory atrioventricular connection; L-TGA, levo-transposition of the great arteries; LB, likely benign; LP, likely pathogenic; MVP, mitral valve prolapse; NCC, noncompaction cardiomyopathy; NK2-SD, NK2-specific domain; P, pathogenic; PA, pulmonary atresia; PDA, patent ductus arteriosus; PFO, patent foramen ovale; PP, polysplenia; PS, pulmonary stenosis; SD, sudden death; SSS, sick sinus syndrome; TA, transactivation; TCA, tricuspid atresia; TD, thyroid dysgenesis; TN, tinman; ToF, tetralogy of Fallot; TVA, tricuspid valve anomaly; Tyr-rich, tyrosine-rich region/domain; VF, ventricular fibrillation; VNC, ventricular noncompaction; VSD, ventricular septal defect; VT, ventricular tachycardia; VUS, variant of uncertain significance.

* Decreased interaction also with GATA4 and TBX5.
** Decreased interaction with GATA4.
1 Increased protein stability.
1 Protein in nucleus and cytoplasm.
1 Small effect.
Searching for functional explanations for these 61 variants, we found four different linear motifs in the scientific literature, three of which were predicted by the ELM resource. These linear motifs and their corresponding pathogenic GVs are summarized in Table 3. Excluding 14 GVs that are also located in the HD, these linear motifs accounted for five pathogenic GVs: three in the SUMOylation motif and one in each of the two predicted NLS motifs. Therefore, 56 pathogenic GVs are located in regions with barely any information regarding their function.

To further study the pattern distribution of pathogenic GVs, we plotted the number of missense variants (considering those classified as pathogenic and also those found in GDBs with a frequency over 1/10,000) in every window of three residues along the protein (Figure 3). This analysis shows that there is a high number of pathogenic clusters in the third helix of the HD (23 GVs), which, interestingly, is also devoid of missense GVs from the GDBs with higher frequencies than 1/10,000.

### 5.2 Missense pathogenic variants in the homeodomain

The HD has 35 missense pathogenic GVs distributed in 30 of its 60 residues (Table 2). These GVs can be found on the three α helices, although 18 of them are located in the third one (residues 179-194) where all but two of the residues had missense pathogenic GVs (Figure 4a, in red).

The in silico analyses showed that eight residues in the HD point out to the hydrophobic core (Figure 4b), while 12 are predicted to interact with the DNA (Figure 4c-g). Among them, four of the residues pointing into the hydrophobic core and 10 of the residues predicted to interact with the DNA have missense pathogenic GVs (Table 2). Moreover, from the 16 residues with pathogenic GVs in the third α-helix, seven were predicted to interact with the DNA or to be part of the hydrophobic core of the HD. These residues concentrate 11 of the 18 missense pathogenic GVs of the third helix.

### 6 FUNCTIONAL IMPLICATIONS OF NKX2-5 GVS

To further understand the biological implications of the NKX2-5 GVs, we also compiled information of functional assays available for 38 GVs, 32 of them classified as pathogenic (Tables 2 and S1). A total of 22 GVs were located in the HD, three in the YRR, and three in the TN domain.

From the 22 GVs in the HD, 15 were missense, five nonsense, one frameshift, and one synonymous. All of the different in vitro studies performed for nonsynonymous GVs in the HD confirm a severe reduction in transactivation (Table 2). For the synonymous variant c.543G>A (glutamine 181) a synergistic effect was demonstrated when combined with p.Ala119Glu and c.63A>G (glutamic acid 21).

| Table 3 | Pathogenic genetic variants in functional and linear motifs |
|---------|----------------------------------------------------------|
| Description | Residues | Number of pathogenic GVs | Pathogenic GVs in the region | Sources and references |
| SUMOylation motif | 51–54 | 3 | p.(F51L), p.(E54K), p.(A50fs) | ELM; Kim et al. (2013); Wang et al. (2008); Kim et al. (2011); Wang et al. (2008); Li et al. (2000); Tang et al. (2006) |
| Acetylation site | 183 | 1 | p.(K183E) | Li et al. (2007); Tang et al. (2016) |
| CKII phosphorylation motif | 161–167 | 6 | p.R161P, p.(L163R), p.(S164*), p.(E167fs), p.(E167K), p.E167* | ELM; Kasahara and Izumo (1999) |
| NLS motifs | 136–143 | 4 | p.(A135T), p.R139W, p.(P141A), p.R142C | ELM; Kasahara and Izumo (1999) |
| NLS motifs | 192–199 | 5 | p.K192*, p.(K192T), p.(K192R), p.Q198*, p.(K194R), p.K194R, p.Q198* | ELM; Ouyang et al. (2016) |

Note: For each functional or linear motif, their positions in NKX2-5 are listed as a range of residues, along with the pathogenic genetic variants found, represented by their "p." descriptors. Genetic variants in bold are unique to the linear motif (not present in other known regions). ELM: Predicted by the Eukaryotic Linear Motif resource (http://elm.eu.org/). (Gouw et al., 2018).
There were 11 GVs from the HD in which the effects on dimerization were tested and six showed a clear impairment of dimerization. Five of these studies also showed decreased interaction with GATA4 and three of them reduced interaction with TBX5, as well. Furthermore, 12 of the 22 GVs (six in the third α helix) included studies of DNA binding, all of which showed a reduction in comparison with wild-type NKX2-5. From these, six were located in residues predicted to interact with the DNA or to be part of the hydrophobic core of the HD (Figure 4b-g).

From the three GVs with in vitro studies located in the YRR, only one (p.Cys270Tyr) showed no change in transactivation compared with the wild type. Of note, p.Cys270Tyr is the variant classified as likely benign following the ACMG guidelines. Additionally, two GVs (p.Glu21Gln and p.Asn19Asp) in the TN domain showed diminished transactivation across different experiments. The other one, c.63A>G (glutamic acid 21), was a synonymous change that was found to cause a small reduction in transactivation and a synergistic effect when combined with the variants p.Ala119Glu and the synonymous change c.543G>A (glutamine 181).

Lastly, from the remaining 10 variants, the most severe effect was seen for the splice donor GV c.334+1G>T, which was shown to not accumulate in the cell, while only one variant (p.Val315Leu) showed no change compared with wild type transactivation. Of note, there were two variants (p.Arg25Cys and p.Gln198*), which showed no change compared with wild type transactivation. Of note, p.Cys270Tyr is the variant classified as likely benign following the ACMG guidelines. Additionally, two GVs (p.Glu21Gln and p.Asn19Asp) in the TN domain showed diminished transactivation across different experiments. The other one, c.63A>G (glutamic acid 21), was a synonymous change that was found to cause a small reduction in transactivation and a synergistic effect when combined with the variants p.Ala119Glu and the synonymous change c.543G>A (glutamine 181).

Among the GVs in the scientific literature, more than a third were not present in any of the public databases. Additionally, the two novel variants from CNGM were obtained in a cohort of 64 patients. This points out the importance of compiling information from different sources when characterizing GVs on a gene. Moreover, the analysis of patients in populations often underrepresented in the databases, like ours, reinforces the notion that novel variants can still be found and for which biological implications could be studied. In that sense, although neither of the novel variants from the CNGM cohort affect the protein sequence, we cannot rule out that these variants could have an effect on protein expression and/or contribute with other concomitant GVs to the causality of the disease. Indeed, synonymous variants have been found to modulate transactivation both by themselves (Ouyang et al., 2011) and/or when associated with a pathogenic variant (Reamon-Buettner et al., 2013). Additional studies searching for GVs in NKX2-5 gene in Latin American countries would also be of interest to increase our knowledge of their roles in pathogenicity among different populations.

Around 10% of the total GVs compiled represents pathogenic GVs and a high degree of concordance was observed when applying ACMG guidelines for prediction of pathogenicity. For the variants

7 | DISCUSSION AND FUTURE PROSPECTS

NKX2-5 is a homeobox protein that plays an important role in the formation of the early heart (Ellesèe et al., 2016; Shiojima et al., 1995). Starting with Schott et al. (1998), several pathological variants have been reported for this gene over the years, mostly in patients with CHD (Ellesèe et al., 2016; Reamon-Buettner & Borlak, 2010; Su et al., 2017).

The present update compiles 1,380 GVs in NKX2-5 retrieved from different sources, including six public databases, 97 scientific publications, and a cohort of CHD patients from Argentina. It contains a comprehensive list of all pathogenic GVs, along with their phenotypes, and variants found in the GDBs, with their respective frequencies. The retrieved GVs were evaluated in relation to their location in domains, regions, motifs, and sites with relevance in the gene. In addition, when available, data of in vitro effect of the variants was collected. The integration of this information allows us to analyze the data in ways that would otherwise be hard to ascertain when individually evaluating GVs.

Among the GVs in the scientific literature, more than a third were not present in any of the public databases. Additionally, the two novel variants from CNGM were obtained in a cohort of 64 patients. This points out the importance of compiling information from different sources when characterizing GVs on a gene. Moreover, the analysis of patients in populations often underrepresented in the databases, like ours, reinforces the notion that novel variants can still be found and for which biological implications could be studied. In that sense, although neither of the novel variants from the CNGM cohort affect the protein sequence, we cannot rule out that these variants could have an effect on protein expression and/or contribute with other concomitant GVs to the causality of the disease. Indeed, synonymous variants have been found to modulate transactivation both by themselves (Ouyang et al., 2011) and/or when associated with a pathogenic variant (Reamon-Buettner et al., 2013). Additional studies searching for GVs in NKX2-5 gene in Latin American countries would also be of interest to increase our knowledge of their roles in pathogenicity among different populations.

Around 10% of the total GVs compiled represents pathogenic GVs and a high degree of concordance was observed when applying ACMG guidelines for prediction of pathogenicity. For the variants
FIGURE 4  Tridimensional structure of the NKX2-5 homeodomain and its interaction with DNA. The image was generated with help from UCSF Chimera (http://www.rbvi.ucsf.edu/chimera). (a) Schematic 3D representation of the NKX2-5 homeodomain along with a segment of the ANF promoter (PDB ID 3RKQ, B chain). In red, residues that contain genetic variants classified as pathogenic. (b) Pipe and planks representation of the NKX2-5 homeodomain. Helices 1–3 are represented as yellow pipes (H1–3). Loops between helices are shown in blue. The hydrophobic residues that point toward the inner face or “hydrophobic core” of the α helices are shown in beige. (c) Schematic representation of the ANF sequence of DNA to which NKX2-5 binds. Phosphate groups, ribose, nitrogenous bases, and interactions with NKX2-5 residues are indicated. Contacts between the DNA and NKX2-5 residues indicate water molecules mediating the interaction with the residue. (d–g) Close inspection of residues, which interact with different parts of the ANF promoter sequence. Contacts with the DNA and the water molecules that could mediate the interaction are shown. H1–3, helixes 1–3
showing discordant results, the differences could be explained considering the classification criteria used. Any variant classified as pathogenic in the literature or predicted as such in clinical databases was classified as "Pathogenic" in the compiled database and therefore considered as having relevance in human health. The ACMG criteria, on the other hand, includes population frequencies, third party in silico predictive tools, allele origin, and in vitro results, among others. It is important to note, however, that the finding of healthy carriers of NKK2-5 GVs in some families and GDBs could skew predictions of pathogenicity. Incomplete penetrance has been repeatedly observed in cardiopathies (Benson, 2002), which could explain the reported healthy carriers of pathogenic variants in the NKK2-5 gene (Benson et al., 1999; De Luca et al., 2010; Kasahara & Benson, 2004; Liu et al., 2011; Reamon-Buettnner & Borlak, 2010; Stallmeyer, Fenge, Nowak-Göttl, & Schulze-Bahr, 2010).

Although a detailed analysis of genotype-phenotype correlation is beyond the scope of the current update, it is important to note that the most common pathologies associated with NKK2-5 GVs in our database are ASD, followed by VSD, AV block, AVSD, and ToF. This observation reinforces similar data from the literature (Ellesøe et al., 2010; Reamon-Buettnner & Borlak, 2010; Su et al., 2017). We also found other noncardiac phenotypes besides the already known thyroid-related ones: two GVs in patients with heterotaxy and asplenia or polysplenia (Izumi, Noon, Wilkins, & Krantz, 2014; Watanabe et al., 2002) and one GV in patients with isolated congenital asplenia (Koss et al., 2012).

In the final compiled database, GVs have been found along the NKK2-5 gene, but all pathogenic variants have been found in sites where they directly affect the protein sequence. As noted in other studies (Elliott et al., 2010; Su et al., 2017) and in this update, the HD has the biggest cluster of pathogenic variants in the protein. Our study further reinforces that there is a cluster of GVs in the third α-helix of the HD, supported by the fact that we observed no high-frequency missense GVs in this helix. We also made more detailed in silico predictions of residues that might be involved in protein–DNA or hydrophobic core interactions in the HD. These predictions suggest that around 61% of missense pathogenic GVs in the third α-helix are located in residues that may either interact with the DNA or be part of the hydrophobic core.

In summary, these studies highlight an important role of the third helix of the homeodomain, which is supported by tridimensional structures showing it is the part of the HD that is inserted in the major groove of the DNA (Luna-Zurita et al., 2016; Pradhan et al., 2012, 2016). Nevertheless, the fact that there are pathogenic variants in helix 3 not pointed to the DNA nor its hydrophobic core suggests that other factors may be playing an important role in the physiological function of these residues, like interaction with other proteins. For example, residue C193 may interact with TBX5 based on 3D structure observations (Pradhan et al., 2016). It is plausible to predict that any missense GV in the third helix of the HD might have a high risk of being pathogenic.

The possible functional effect of GVs as interpreted with in vitro assays can be roughly divided in two groups. On one hand, missense GVs and inframe insertions/deletions most likely affect the residues that are being modified by the change in protein sequence, so their location in a functional region of a protein would hint that their effect is related to that region. Therefore, in vitro studies of missense GVs can help illustrate the function of both a region and a GV. Examples are those missense GVs related to DNA binding and/or dimerization. Frameshift, nonsense, and splice site GVs, on the other hand, most often have an effect on the entire protein, either by the deletion of residues downstream of their location leading to a truncated protein, to a protein with a different coding frame or even to the absence of the protein.

It is important to note that some of these variants might cause functional haploinsufficiency, like the splice donor site c.334+1G>T found in a patient with AV block (Kasahara et al., 2000). Haploinsufficiency has been demonstrated to be related to cardiac defects in animal models (Azhari, Shihata, & Al-Fatani, 2004; Biben et al., 2000; Tanaka et al., 2002; Winston et al., 2010). Moreover, deletions encompassing NKK2-5 have been described in patients presenting ASD, VSD, and ventricular myocardial noncompaction (Joseph, Kimm, Kalyan-Raman, Nixon, & Hiller, 1990; Kleczkowska, Fryns, & van den Bergh, 1993; Pauli, Scheib-Wixted, Cripe, Izumo, & Sekhon, 1999). It should be remembered that, as our database does not include GVs that directly changed over 50 base pairs, some partial or complete deletions of the gene were not compiled, and information should be expanded upon when studying NKK2-5 haploinsufficiency.

While functional assays are one of the main sources of confirmation for functional effect on GVs, we have observed that most of the compiled GVs retrieved from the literate or clinical databases do not have functional studies associated with them. It would be of interest to fill this gap, especially in regions outside of the homeodomain, which could contribute to shed light into the function of less studied regions of the NKK2-5 protein. In this regard, variants classified as VUS would be of particular interest. Alternatively, further studies finding these variants in either families with heart disease or in individuals in the general population could clarify their role in pathogenesis.

In summary, around 40% of the pathogenic GVs in our compiled database are in the most conserved domains of the NK2 subfamily of proteins (TN, HD, NK2-SD). When accounting for other regions (YRR, NLS motif, Nkx2-5 box, GIARAW motif, SUMOylation site, phosphorylation site, and acetylation site), we were able to assign a possible functional effect on protein motifs of approximately 60% of the pathogenic GVs. Functional studies confirm an effect for 26 of the GVs in the abovementioned regions and further add information on five GVs out of them. Therefore, a total of 63% of the pathogenic GVs have a putative explanation for their assigned pathogenicity.

Finally, it was suggested that oligogenic combinations of inherited GVs could explain the majority of CHD that lack a detectable monogenic basis (D’Alessandro et al., 2016; De Luca et al., 2010; Li et al., 2016; Töpf et al., 2014). In addition, genetic modifiers may contribute to modulate or even to abolish or promote the effect of a given GV (Winston et al., 2010). Although physical and functional...
interactions between NKX2-5 and GATA4 and NKX2-5 and TBX5 are well-documented (Durocher, Charron, Warren, Schwartz, & Nemer, 1997; Hiroi et al., 2001), it has been reported that NKX2-5 may interact with other proteins like KDM6 (Lee, Lee, & Lee, 2012), JARID2 (Kim, Chen, Sadoshima, & Lee, 2004), CAMTA2 (Song et al., 2006), and Fbxo25 (Jeong et al., 2015). Also, high-throughput experiments identified interactions of NKX2-5 with RBPJ, FOXA1, FOXE1, SMAD4, and GRB2 (Huttlin et al., 2017; Li et al., 2015). Altogether this data could provide a polygenic explanation for both the incomplete penetrance and phenotypic variability seen for NKX2-5 GVs. In line with this concept, we have documented in our database GVs in NKX2-5 seen concomitantly with variants in other relevant genes related to diseases. Even though some of these GVs are classified as likely pathogenic, others classified as likely benign or VUS, could still possibly impact the development of the disease in combination with the NKX2-5 variants.

8 | CONCLUDING REMARKS

This study was designed to build an exhaustive database of NKX2-5 genetic variants. All of the compiled information pointed to reliable pathogenicity for GVs in helix 3 of the homeodomain. In addition, the compiled data may broaden the scope of functional and structural studies that can be done to better understand the effect of pathogenic GVs in NKX2-5 function.

ACKNOWLEDGMENTS

We want to thank Isabel Lüthy, PhD, for her collaboration and support in this study at the IBYME-CONICET and Gustavo Ontiveros, MD, for his clinical support in the cardiogenic field. This study was supported by Health Scholarships Research “Dr. Abraam Sonis” 2017 from the National Secretary of Health (Carlos D. Bruque). Grants from Agencia Nacional de Promoción Científica y Tecnológica PID 2012-0060 (Liliana Dain) and Universidad Nacional de Buenos Aires: UBACyT 20020170100592BA (Alejandro D. Nadra and Liliana Dain).

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Conceived and designed the methodology: JEK, ADN, CDB, and LD. Compilation of reported genetic variants: JEK, LS, and MF. Analyzed patients’ samples: MD, JDE, and MT. Analyzed and discussed the data: JEK, MD, ADN, CDB, and LD. Contributed reagents/materials/analysis tools: ADN, CDB, and LD. Wrote the paper: JEK, CDB, and LD. All authors reviewed the manuscript.

DATA AVAILABILITY STATEMENT

The genetic variant data from this study have been submitted to LOVD database (https://www.lovd.nl/NKX2-5).

ORCID

Jorge E. Kolomenski http://orcid.org/0000-0002-1299-3017
Marisol Delen http://orcid.org/0000-0001-7380-9968
Leandro Simonetti http://orcid.org/0000-0003-1283-9770
Mónica C. Fabbro http://orcid.org/0000-0003-0192-4777
Lucia D. Espeche http://orcid.org/0000-0002-9935-6925
Melisa Taboos http://orcid.org/0000-0002-9488-901X
Alejandro D. Nadra http://orcid.org/0000-0002-7860-3245
Carlos D. Bruque http://orcid.org/0000-0003-0726-5418
Liliana Dain http://orcid.org/0000-0002-2155-4999

REFERENCES

Abou Hassan, O. K., Fahed, A. C., Batrawi, M., Arabi, M., Refaat, M. M., DePalma, S. R., ... Nemer, G. M. (2015). NKX2-5 mutations in an inbred consanguineous population: Genetic and phenotypic diversity. Scientific Reports, 5, 8848.

Akçaboy, M. I., Cengiz, F. B., Inceoğlu, B., Uçar, T., Atalay, S., Tutar, E., & Tekin, M. (2008). The Effect of p.Arg25Cys Alteration in NKX2-5 on Conotruncal Heart Anomalies: Mutation or Polymorphism? Pediatric Cardiology, 29(1), 126–129. http://dx.doi.org/10.1007/s00246-007-9058-2

Azhari, N., Shihata, M. S., & Al-Fatani, A. (2004). Spontaneous closure of atrial septal defects within the oval fossa. Cardiology in the Young, 14(2), 148–155.

Benson, D. W. (2000). The genetics of congenital heart disease: A point in the revolution. Cardiology Clinics, 20(3), 385–394. vi.

Benson, D. W., Silberbach, G. M., Kavanaugh-McHugh, A., Cottrill, C., Zhang, Y., Riggs, S., ... Kugler, J. D. (1999). Mutations in the cardiac transcription factor NKX2.5 affect diverse cardiac developmental pathways. The Journal of Clinical Investigation, 104(11), 1567–1573.

Bermúdez-Jiménez, F. J., Jiménez-Jáimez, J., & López-Fernández, S. (2017). Overlap of arrhythmogenic cardiomyopathy, spongiform cardiomyopathy, and congenital heart disease. Revista Española de Cardiología, 70(1), 51.

Biben, C., Weber, R., Kesteven, S., Stanley, E., McDonald, L., Elliott, D. A., ... Harvey, R. P. (2000). Cardiac septal and valvular dysmorphogenesis in mice heterozygous for mutations in the homeobox gene Nkx2-5. Circulation Research, 87(10), 888–895.

Bjørnstad, P. G., & Leren, T. P. (2009). Familial atrial septal defect in the oval fossa with progressive prolongation of the atrioventricular conduction caused by mutations in the NKX2.5 gene. Cardiology in the Young, 19(1), 40–44.

Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in Drosophila. Development, 118(3), 719–729.

Bouveret, R., Waardenberg, A. J., Schonrock, N., Ramalison, M., Doan, T., deJong, D., ... Harvey, R. P. (2015). NKX2-5 mutations causative for congenital heart disease retain functionality and are directed to hundreds of targets. eLife, 4, 4. https://doi.org/10.7554/eLife.06942

Brendolan, A., Ferretti, E., Salsi, V., Moses, K., Quagggin, S., Blasi, F., ... Selleri, L. (2005). A Pbx1-dependent genetic and transcriptional network regulates spleen ontogeny. Development, 132(13), 3113–3126.

Briggs, L. E., Takeda, M., Cuadra, A. E., Wakimoto, H., Marks, M. H., Walker, A. J., ... Kasahara, H. (2008). Perinatal loss of Nkx2-5 results in rapid conduction and contraction defects. Circulation Research, 103(6), 580–590.

Burn, S. F., Boot, M. J., deAngelis, C., Doothan, R., Arques, C. G., Torres, M., & Hill, R. E. (2008). The dynamics of spleen morphogenesis. Developmental Biology, 318(2), 303–311.

Chen, Y., Mao, J., Sun, Y., Zhang, Q., Cheng, H.-B., Yan, W.-H., ... Li, H. (2010). A novel mutation of GATA4 in a familial atrial septal
HAND1 genes in patients with atrial isomerism. Anadolu Kardiyojeli Dergisi, 11(4), 319–328.

Hermanns, P., Grasberger, H., Refetoff, S., & Pohlenz, J. (2011). Mutations in the NKX2.5 gene and the PAC8 promoter in a girl with thyroid dysgenesis. The Journal of Clinical Endocrinology and Metabolism, 96(6), E977–E981.

Hirayama-Yamada, K., Kamisago, M., Akimoto, K., Aotsuka, H., Nakamura, Y., Tomita, H., & Matsuoka, R. (2005). Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. American Journal of Medical Genetics Part A, 135A(1), 47–52. http://dx.doi.org/10.1002/ajmg.a30684

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R., & Komuro, I. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. Nature Genetics, 28(3), 276–280. https://doi.org/10.1038/ng10123

Hoffman, J. I. (2013). The global burden of congenital heart disease. Cardiovascular Journal of Africa, 24(4), 141–145.

Hoffman, J. I. E., & Kaplan, S. (2002). The incidence of congenital heart disease. Journal of the American College of Cardiology, 39(12), 1890–1900.

Homsy, J., Zaidi, S., Shen, Y., Ware, J. S., Samocha, K. E., Karczewski, K. J., ... Chung, W. K. (2015). De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. Science, 350(6265), 1262–1266.

Hosoda, T., Komuro, I., Shiiojima, I., Hiroi, Y., Harada, M., Murakawa, Y., ... Yazaki, Y. (1999). Familial atrial septal defect and atrioventricular conduction disturbance associated with a point mutation in the cardiac homeobox gene Csx/NKX2-5 in a Japanese patient. Japanese Circulation Journal, 63(5), 425–426.

Huang, J.-B., Liu, Y.-L., Sun, P.-W., Lu, X.-D., Du, M., & Fan, X.-M. (2010). Molecular mechanisms of congenital heart disease. Cardiovascular Pathology, 19(5), e183–e192.

Huang, R-T., Xue, S., Xu, Y.-J., Zhou, M., & Yang, Y.-Q. (2013). A novel NKX2.5 loss-of-function mutation responsible for familial atrial fibrillation. International Journal of Molecular Medicine, 31(5), 1119–1126.

Huttlin, E. L., Bruckner, R. J., Paulo, J. A., Cannon, J. R., Ting, L., Baltier, K., ... Krantz, I. D. (2014). NKX2.5 mutation with GATA4 or NKX2.5 mutations in familial atrial septal defect. Annales de Genetique, 36(2), 126–128.

Kado, K., Nishizawa, T., Furutani, M., Arai, S., Ishihara, K., Oda, M., ... Yamagishi, H. (2012). Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects. Circulation Journal, 76(7), 1703–1711.

König, K., Will, J. C., Berger, F., Müller, D., & Benson, D. W. (2006). Familial congenital heart disease, progressive atrioventricular block and the cardiac homebox transcription factor gene NKX2.5: Identification of a novel mutation. Clinical Research in Cardiology, 95(9), 499–503.

Koss, M., Bolse, A., Brendolan, A., Saggese, M., Capellini, T. D., Bojilova, E., ... Selleri, L. (2012). Congenital asplenia in mice and humans with mutations in a Pbx/Nkx2-5/p15 module. Developmental Cell, 22(5), 913–926.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., ... Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. Nucleic Acids Research, 46(D1), D1062–D1067.

Laskowski, R. A., Hutchinson, E. G., Michie, A. D., Wallace, A. C., Jones, M. L., & Thornton, J. M. (1997). PDbsum: A Web-based database of summaries and analyses of all PDB structures. Trends in Biochemical Sciences, 22(12), 488–490.

Lee, S., Lee, J. W., & Lee, S.-K. (2012). UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. Developmental Cell, 22(1), 25–37.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... Exome Aggregation Consortium. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536(7616), 285–291.

Li, T., Li, Y.-M., Jia, Z.-Q., Chen, P., Ma, K.-T., & Zhou, C.-Y. (2007). Carboxyl terminus of Nkx2.5 impairs its interaction with p300. Journal of Molecular Biology, 370(5), 976–992.

Li, X., Wang, W., Wang, J., Malovannaya, A., Xi, Y., Li, W., ... Chen, J. (2015). Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. Molecular Systems Biology, 11(1), 775.

Li, Y., Yagi, H., Ono, H., Damerla, R. R., Francis, R., Furutani, Y., ... Lo, C. W. (2016). DNA-H6 and its interactions with PCD genes in heterotaxy and primary ciliary dyskinesia. PLOS Genetics, 12(2), e1005821.

Liu, X.-Y., Wang, J., Yang, Y.-Q., Zhang, Y.-Y., Chen, X.-Z., Zhang, W., ... Chen, Y.-H. (2011). Novel NKX2.5 mutations in patients with familial atrial septal defects. Pediatric Cardiology, 32(2), 193–201.

Liu, X.-Y., Yang, Y.-Q., Yang, Y., Lin, X.-P., & Chen, Y.-H. (2009a). Mutation of NKX2.5 gene in patients with atrial septal defect. Zhonghua er ke za zhi, 47(9), 696–700.

Liu, X.-Y., Yang, Y.-Q., Yang, Y., Lin, X.-P., & Chen, Y.-H. (2009b). Novel NKX2.5 mutations identified in patients with congenital ventricular septal defects. Zhonghua Yi Xue Za Zhi, 89(34), 2395–2399.
stimulates cardiac growth by opposing class II histone deacetylases. Cell, 125(3), 453–466.

Stallmeyer, B., Fenge, H., Nowak-Götti, U., & Schulze-Bahr, E. (2010). Mutational spectrum in the cardiac transcription factor gene NKX2.5 (Csx) associated with congenital heart disease. Clinical Genetics, 78(6), 533–540.

Su, W., Zhu, P., Wang, R., Wu, Q., Wang, M., Zhang, X., ... Dong, N. (2017). Congenital heart diseases and their association with the variant distribution features on susceptibility genes. Clinical Genetics, 91(3), 349–354.

Tanaka, M., Berul, C. I., Ishii, M., Jay, P. Y., Wakimoto, H., Douglas, P., ... Izumo, S. (2002). A mouse model of congenital heart disease: Cardiac arrhythmias and atrial septal defect caused by haploinsufficiency of the cardiac transcription factor Csx/Nkx2.5. Cold Spring Harbor Symposium on Quantitative Biology, 67, 317–325.

Tang, X., Ma, H., Han, L., Zheng, W., Lu, Y.-B., Chen, X.-F., ... Liu, D.-P. (2016). SIRT1 deacetylates the cardiac transcription factor Nkx2.5 and inhibits its transcriptional activity. Scientific Reports, 6, 36576.

Terada, R., Warren, S., Lu, J. T., Chien, K. R., Wessels, A., & Kasahara, H. (2011). Ablation of Nkx2.5 at mid-embryonic stage results in premature lethality and cardiac malformation. Cardiovascular Research, 91(2), 289–299.

Tian, L., Zhu, J.-F., Yang, J.-G., Zhu, Q.-H., Du, R., Li, J., & Li, W. (2008). Gene mutation in secundum atrial septal defect: Analysis of a Chinese family with 3 patients. Zhonghua Yi Xue Za Zhi, 88(4), 250–253.

Tong, Y.-F. (2016). Mutations of NKX2.5 and GATA4 genes in the development of congenital heart disease. Gene, 588(1), 86–94.

Töpf, A., Griffin, H. R., Glen, E., Soemedi, R., Brown, D. L., Hall, D., ... Goodship, J. A. (2014). Functionally significant, rare transcription factor variants in tetralogy of Fallot. PLOS One, 9(8), e95453.

van der Bom, T., Zomer, A. C., Zwinderman, A. H., Meijboom, F. J., Bourma, B. J., & Mulder, B. J. M. (2011). The changing epidemiology of congenital heart disease. Nature Reviews Cardiology, 8(1), 50–60.

van der Linde, D., Konings, E. E. M., Witsenburg, M., Helbing, W. A., Takkenberg, J. J. M., & Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. Journal of the American College of Cardiology, 58(21), 2241–2247.

Wakimoto, H., Kasahara, H., Maguire, C. T., Moskowitz, I. P. G., Izumo, S., & Berul, C. I. (2003). Cardiac electrophysiological phenotypes in postnatal expression of Nkx2.5 transgenic mice. Genesis, 37(3), 144–150.

Walsh, R., Thomson, K. L., Ware, J. S., Funke, B. H., Woodley, J., McGuire, K. J., ... Watkins, H. (2017). Reassessment of Mendelian gene pathogenicity using 7,855 cardiomypathy cases and 60,706 reference samples. Genetics in Medicine, 19(2), 192–203.

Wang, J., Chen, Q., Wang, L., Zhou, S., Cheng, L., Xie, X., ... Ma, X. (2011a). Identifying novel mutations of NKX2-5 congenital heart disease patients of Chinese minority groups. International Journal of Cardiology, 148(1), 102–104.

Wang, F., Liu, C., Jia, X., Liu, X., Xu, Y., Yan, S., ... Gu, M. (2017). Next-generation sequencing of NKX2.1, FOXE1, PAX8, NKX2.5, and TSHR in 100 Chinese patients with congenital hypothyroidism and athyreosis. Clinica Chimica Acta, 470, 36–41.

Wang, J., Liu, X. Y., & Yang, Y. Q. (2011b). Novel NKX2-5 mutations responsible for congenital heart disease. Genetics and Molecular Research, 10(4), 2905–2915.

Wang, J., Xin, Y.-F., Liu, X.-Y., Liu, Z.-M., Wang, X.-Z., & Yang, Y.-Q. (2011c). A novel NKX2-5 mutation in familial ventricular septal defect. International Journal of Molecular Medicine, 27(3), 369–375.

Wang, J., Zhang, H., Iyder, D., Feng, X.-H., & Schwartz, R. J. (2008). Regulation of cardiac specific Nkx2.5 gene activity by small ubiquitin-like modifier. The Journal of Biological Chemistry, 283(34), 23235–23243.

Watada, H., Mirmira, R. G., Kalamaras, J., & German, M. S. (2000). Intramolecular control of transcriptional activity by the NK2-specific domain in NK-2 homeodomain proteins. Proceedings of the National Academy of Sciences of the United States of America, 97(17), 9443–9448.

Watanabe, Y., Benson, D. W., Yano, S., Akagi, T., Yoshino, M., & Murray, J. C. (2002). Two novel frameshift mutations in NKKX2.5 result in novel features including visceral inversus and sinus venosus type ASD. Journal of Medical Genetics, 39(11), 807–811.

Winston, J. B., Erlich, J. M., Green, C. A., Aluko, A., Kaiser, K. A., Takematsu, M., ... Jay, P. Y. (2010). Heterogeneity of genetic modifiers ensures normal cardiac development. Circulation, 121(11), 1313–1321.

Xie, W.-H., Chang, C., Xu, Y.-J., Li, R.-G., Qu, X.-K., Fang, W.-Y., ... Yang, Y.-Q. (2011). Prevalence and spectrum of Nkx2.5 mutations associated with idiopathic atrial fibrillation. Clinics, 68(6), 777–784.

Xu, J.-H., Gu, J.-Y., Guo, Y.-H., Zhang, H., Qiu, X.-B., Li, R.-G., ... Yang, Y.-Q. (2017a). Prevalence and spectrum of NKX2-5 mutations associated with sporadic adult-onset dilated cardiomyopathy. International Heart Journal, 58(4), 521–529.

Xu, Y.-J., Qiu, X.-B., Yuan, F., Shi, H.-Y., Xu, L., Hou, X.-M., ... Li, R.-G. (2017b). Prevalence and spectrum of NKX2.5 mutations in patients with congenital atrial septal defect and atrioventricular block. Molecular Medicine Reports, 15(4), 2247–2254.

Yuan, F., Qiu, X.-B., Li, R.-G., Qu, X.-K., Wang, J., Xu, Y.-J., ... Liao, D.-N. (2015). A novel Nkx2.5 loss-of-function mutation predisposes to familial dilated cardiomyopathy and arrhythmias. International Journal of Molecular Medicine, 35(2), 478–486.

Yu, H., Xu, J.-H., Song, H.-M., Zhao, L., Xu, W.-J., Wang, J., ... Yang, Y.-Q. (2014). Mutational spectrum of the NKX2-5 gene in patients with lone atrial fibrillation. International Journal of Medical Sciences, 11(6), 554–563.

Zaidi, S., & Brueckner, M. (2017). Genetics and genomics of congenital heart disease. Circulation Research, 120(6), 923–940.

Zakariyah, A. F., Rajgara, R. F., Horner, E., Cattin, M.-E., Blais, A., Skerjanec, I. S., & Burgon, P. G. (2018). In vitro modeling of congenital heart defects associated with an Nkx2.5 mutation revealed a dysregulation in BMP/notch-mediated signaling. Stem Cells, 36(4), 514–526.

Zakariyah, A. F., Rajgara, R. F., Veinot, J. P., Skerjanec, I. S., & Burgon, P. G. (2017). Congenital heart defect causing mutation in Nkx2.5 displays in vivo functional deficit. Journal of Molecular and Cellular Cardiology, 105, 89–98.

Zhu, W., Shiojima, I., Hiroi, Y., Zou, Y., Akazawa, H., Mizukami, M., ... Komuro, I. (2000). Functional analyses of three Csx/Nkx2.5 mutations that cause human congenital heart disease. The Journal of Biological Chemistry, 275(45), 35291–35296.

**Supporting Information**

Additional supporting information may be found online in the Supporting Information section.