Carlos David Bruque ORCID iD: 0000-0003-0726-5418

Liliana Dain ORCID iD: 0000-0002-2155-4999

Update of genetic variants in the *NKX2-5* gene

Jorge E. Kolomenski[1,2], Marisol Delea[3], Leandro Simonetti[4]; Mónica C. Fabbro[5]; Lucía D. Espeche[3], Melisa Taboas[3], Alejandro D. Nadra[1,2], Carlos D. Bruque[3,6,*] and Liliana Dain[2,3,6*].

[1] Departamento de Química Biológica Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, IQUIBICEN-CONICET, Buenos Aires, Argentina.

[2] Departamento de Fisiología, Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Instituto de Biociencias, Biotecnología y Biomedicina, Universidad de Buenos Aires, Buenos Aires, Argentina.

[3] Centro Nacional de Genética Médica, ANLIS, Buenos Aires, Argentina.

[4] Department of Chemistry - Biomedical Centre, Uppsala University, Uppsala, Sweden.

[5] Laboratorio Novagen, Buenos Aires, Argentina.

[6] Instituto de Biología y Medicina Experimental, (IBYME-CONICET), Buenos Aires, Argentina.

* Corresponding authors. e-mails:

ldain@fbmc.fcen.uba.ar; bruquecarlos@gmail.com

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/humu.24030.

This work was supported by Health Scholarships Research “Dr. Abraam Sonis” 2017 from the National Secretary of Health (CDB), Grants from Agencia Nacional de Promoción Científica y Tecnológica PID 2012-0060 (LD) and Universidad Nacional de Buenos Aires: UBACyT 20020170100592BA (ADN and LD).

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 in order to build a unified, curated and updated compilation of all available GVs.

We retrieved a total of 1380 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 alpha-helix, where 11 missense variants may affect protein-DNA interaction or the hydrophobic core. We also pinpointed the likely location of pathogenic GVs in 4 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: NKX2-5; genetic variant evaluation; curated database; associated phenotypes; linear motif.

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; H. Kasahara et al. 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 et al. 2010). The HD is a conserved helix-loop-helix domain with three alpha helices (Pradhan, Gopal, and Nam 2014) that recognizes and binds to DNA and can homo or heterodimerize (Elliott et al., 2010). The NK2-SD was suggested to function as an intramolecular transactivation regulator (Watada et al. 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; Y.N. Liu et al., 2015), two putative nuclear localization signals (NLS) (H. Kasahara & Izumo, 1999; Ouyang et al., 2016),

a SUMOylation motif (E. Y. Kim et al., 2011; Jun Wang et al., 2008), a phosphorylation site and an acetylation site, both located in the HD (p.S164 and p.K183, respectively) (H. Kasahara & Izumo, 1999; T. 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 GIRAW 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 (C. Y. Chen & Schwartz, 1995; Dupays et al., 2015), as shown in its interaction with the *NPPA* gene promoter (Pradhan et al., 2012, 2016).

There are 3 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 and co-workers (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 et al., 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 (J. I. Hoffman, 2013; J. I. E. Hoffman & Kaplan, 2002; van der Linde et al., 2011). It is the leading non-infectious cause of death in the first year of life (J.-B. Huang et al., 2010; Zaidi & Brueckner, 2017) and the most frequent type of congenital disease (J.-B. 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), Tetralogy of Fallot (ToF), hypoplastic left heart syndrome, double outlet right ventricle, arrhythmia and sudden death, among others (Ellesøe et al., 2016; Stella Marie Reamon-Buettner & Borlak, 2010; Zakariyah et al., 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 (Stella Marie 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, in order to build a unified, curated and updated compilation of the GVs of this gene.

2 | *NKX2-5* database of genetic variants

We compiled the information of GVs in *NKX2-5* from 6 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 et al., 2010), ExAC (https://exac.broadinstitute.org/) (Lek et al., 2016), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar) (Landrum et al., 2018) and LOVD (http://www.lovd.nl/3.0/home/) (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, https://www.cardiodb.org/acgv/) (Walsh et al., 2017), but genetic variants were not found for *NKX2-5*.

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 in order to make them easily comparable.

The 381 published articles were obtained from PubMed (https://www.ncbi.nlm.nih.gov/pubmed) on August 2018 by searching for the different names for the gene in GeneCards (https://www.genecards.org/cgi-bin/carddisp.pl?gene=NKX2-5) and either of the terms “mutation” and “genetic variant”. The compiled GVs were restricted to variations spanning equal or under 50 base pairs.

The 64 CHD patients from the CNGM were analyzed by direct DNA sequencing (Supp. Text S1).

In order to integrate the information from different sources into a single database, we first had to establish a unique identifier. The two most frequently used identifiers are the refSNP identification number (rs#) and the HGVS identification standard (den Dunnen et al., 2016), the latter one being the best option for unambiguous identification of GVs. For this reason, our database uses an identifier following the HGVS identification standard.

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 didn't follow HGVS standards, as was the case for some deletions, insertions and duplications and some of the data from the scientific literature (Supp. Text S1).

In parallel, the cDNA 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 (Supp. 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 (Supp. 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 (Supp. Text S1).

In silico analysis were performed using PDBsum (Laskowski et al., 1997) in order 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 (post-translational modification sites, nuclear localization signals, etc.). This information was matched to predicted linear motifs from the eukaryotic linear motif resource (http://elm.eu.org/) (Gouw et al., 2018) in order 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 1380 unique GVs (Supp. Table S1); 1314 of these variants come from the different databases, which accumulated 1706 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 (Supp. 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 less than 1/10,000 individuals.

Types of GV in NKX2-5

From the total of 1380 unique variants, 626 were found in the translated region (TR), and 754 in the non-translated region (NTR, including the 5' and 3' UTRs 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: 7 deletions, 6 duplications and 1 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.

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 2.A and 2.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), 6 as conflicting evidence and 14 as likely benign in the compiled database (Supp. 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 (Table 2 and Supp. Table S1). Allele origin was ascertained for 108 GVs, from which 65 were familial and 10 de novo (Supp. 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 (Supp. 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 2 different *NKX2-5* GVs) and 6 were predicted to be benign/likely benign (for 4 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 less than 1/10,000 and 7 between 1/10,000-1/100, 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.

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, 1 stop loss. The two remaining GVs occuring 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).

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 4 variants, the HD (residues 138-197) has 49 GVs and the NK2-SD (residues 212-234) has 5 GVs. These account for 58 pathogenic GVs from the total of 141 in the TR. In addition, 16 pathogenic GVs are located in the YRR (residues 237-274), 3 in the NKX2-5 box (residues 291-304) and 3 in the GIRAW motif (residues 320-324), leaving 61 GVs in the rest of the protein (Table 2 and Figure 2B).

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: 3 in the SUMOylation motif and 1 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.

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 alpha 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 8 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, 4 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 alpha-helix, 7 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 (Table 2 and Supp. Table S1). A total of 22 GVs were located in the HD, 3 in the YRR and 3 in the TN domain.

From the 22 GVs in the homeodomain, 15 were missense, 5 nonsense, 1 frameshift and 1 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).

There were 11 GVs from the HD in which the effects on dimerization were tested and 6 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 (6 in the third alpha helix) included studies of DNA binding, all of which showed a reduction in comparison with wild type NKX2-5. From these, 6 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 to the wild type. Of note, p.Cys270Tyr is the variant classified as conflicting evidence that was defined 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 to wild type transactivation. Of note, there were two variants (p.Arg25Cys and p.Gln198*) which showed increased transactivation in some conditions and reduced in others.

7 | Discussion and future prospects

NKX2-5 is a homeobox protein that plays an important role in the formation of the early heart (Ellesøe et al., 2016; Shiojima et al., 1995).

Starting with Schott and co-workers (Schott et al., 1998), several pathological variants have been reported for this gene over the years, mostly in patients with CHD (Ellesøe et al., 2016; Stella Marie Reamon-Buettner & Borlak, 2010; Su et al., 2017).

The present update compiles 1380 GVs in *NKX2-5* retrieved from different sources, including 6 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 general databases, 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 (Stella Marie Reamon-Buettner et al., 2013). Additional studies searching for GVs in *NKX2-5* gene in Latin American countries would also be of interest in order 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 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* resuls, among others. It is important to note, however, that the finding of healthy carriers of *NKX2-5* GVs in some families and GDBs could skew predictions of pathogenicity. Incomplete penetrance has been repeatedly observed in cardiopathies (D. Woodrow Benson, 2002), which could explain the reported healthy carriers of pathogenic variants in the *NKX2-5* gene (D. W. Benson et al., 1999; De Luca et al., 2010; Hideko Kasahara & Benson, 2004; X.-Y. Liu et al., 2011; Stella Marie Reamon-Buettner & Borlak, 2010; Stallmeyer et al., 2010; J. Wang et al., 2011). In addition, some GVs in GDBs may represent individuals that are actually affected but were not registered as patients.

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 *NKX2-5* 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., 2016; Stella Marie Reamon-Buettner & Borlak, 2010; Su et al., 2017). We also found other non-cardiac phenotypes besides the already known thyroid-related ones: two GVs in patients with heterotaxy and asplenia or polysplenia (Izumi et al., 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 *NKX2-5* 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 alpha-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 alpha-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 genetic variants 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 (H. Kasahara et al., 2000). Haploinsufficiency has been demonstrated to be related to cardiac defects in animal models (Azhari et al., 2004; Biben et al., 2000; Tanaka et al., 2002; Winston et al., 2010). Moreover,

deletions encompassing *NKX2-5* have been described in patients presenting ASD, VSD and ventricular myocardial noncompaction (Joseph et al., 1990; Kleczkowska et al., 1993; Pauli et al., 1999). It should be remembered that, since 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 *NKX2-5* 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 NKX2-5 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, GIRAW 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 above mentioned regions and further add information on 5 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 genetic variants could explain the majority of CHD that lack a detectable monogenic basis (D'Alessandro et al., 2016; De Luca et al., 2010; Y. 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 genetic variant (Winston et al., 2010). Although physical and functional interactions between NKX2-5 and GATA4 and NKX2-5 and TBX5 are well documented (Durocher et al., 1997; Hiroi et al., 2001), it has been reported that NKX2-5 may interact with other proteins like KDM6 (Lee et al., 2012), JARID2 (T.G. Kim et al., 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; X. 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 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

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported in the manuscript.

We want to thank Isabel Lüthy, PhD. for her collaboration and support of this work at the IBYME-CONICET and Gustavo Ontiveros, M.D. for his clinical support in the cardiogenic field. JEK is a PhD fellow from the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). MD, LDE, MIT, CDB and LD are professional staff from the Administración Nacional de Laboratorios e Institutos de Salud (ANLIS). ADN and LD are staff researchers from the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). LS is a junior researcher at the Department of Chemistry - BMC, Uppsala University. MCF is a professional staff from Novagen Lab.

In memory of Pablo Kolomenski, father, husband and teacher.

Data Availability Statement

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

Author Contributions

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

References

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

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. (2002). 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., Smalls, O., Johnson, M. C., Watson, M. S., Seidman, J. G., Seidman, C. E., Plowden, J., & 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 Espanola de Cardiologia*, *70*(1), 51.

Biben, C., Weber, R., Kesteven, S., Stanley, E., McDonald, L., Elliott, D. A., Barnett, L., Köentgen, F., Robb, L., Feneley, M., & 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., Ramialison, M., Doan, T., de Jong, D., Bondue, A., Kaur, G., Mohamed, S., Fonoudi, H., Chen, C.-M., Wouters, M. A., Bhattacharya, S., Plachta, N., Dunwoodie, S. L., Chapman, G., Blanpain, C., & Harvey, R. P. (2015). NKX2-5 mutations causative for congenital heart disease retain functionality and are directed to hundreds of targets. *eLife*, *4*. https://doi.org/10.7554/eLife.06942

Brendolan, A., Ferretti, E., Salsi, V., Moses, K., Quaggin, S., Blasi, F., Cleary, M. L., & 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., Seki, T., Oh, S. P., Lu, J. T., Sumners, C., Raizada, M. K., Horikoshi, N., Weinberg, E. O., Yasui, K., Ikeda, Y., Chien, K. R., & 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., de Angelis, C., Doohan, R., Arques, C. G., Torres, M., & Hill, R. E. (2008). The dynamics of spleen morphogenesis. *Developmental Biology*, *318*(2), 303–311.

Chen, C. Y., & Schwartz, R. J. (1995). Identification of novel DNA binding targets and regulatory domains of a murine tinman homeodomain factor, nkx-2.5. *The Journal of Biological Chemistry*, *270*(26), 15628–15633.

Chen, Y., Mao, J., Sun, Y., Zhang, Q., Cheng, H.-B., Yan, W.-H., Choy, K. W., & Li, H. (2010). A novel mutation of GATA4 in a familial atrial septal defect. *Clinica Chimica Acta; International Journal of Clinical Chemistry*, *411*(21-22), 1741–1745.

Choquet, C., Nguyen, T. H. M., Sicard, P., Buttigieg, E., Tran, T. T., Kober, F., Varlet, I., Sturny, R., Costa, M. W., Harvey, R. P., Nguyen, C., Rihet, P., Richard, S., Bernard, M., Kelly, R. G., Lalevée, N., & Miquerol, L. (2018). Deletion of Nkx2-5 in trabecular myocardium reveals the developmental origins of pathological heterogeneity associated with ventricular non-compaction cardiomyopathy. *PLoS Genetics*, *14*(7), e1007502.

Chung, I.-M., & Rajakumar, G. (2016). Genetics of Congenital Heart Defects: The NKX2-5 Gene, a Key Player. *Genes*, *7*(2). https://doi.org/10.3390/genes7020006

Costa, M. W., Guo, G., Wolstein, O., Vale, M., Castro, M. L., Wang, L., Otway, R., Riek, P., Cochrane, N., Furtado, M., Semsarian, C., Weintraub, R. G., Yeoh, T., Hayward, C., Keogh, A., Macdonald, P., Feneley, M., Graham, R. M., Seidman, J. G., … Harvey, R. P. (2013). Functional characterization of a novel mutation in NKX2-5 associated with congenital heart disease and adult-onset cardiomyopathy. *Circulation. Cardiovascular Genetics*, *6*(3), 238–247.

Crawshaw, P. (1995). The new BPA classification. *Archives of Disease in Childhood*, *73*(6), 563–567.

D'Alessandro, L. C. A., Al Turki, S., Manickaraj, A. K., Manase, D., Mulder, B. J. M., Bergin, L., Rosenberg, H. C., Mondal, T., Gordon, E., Lougheed, J., Smythe, J., Devriendt, K., Bhattacharya, S., Watkins, H., Bentham, J., Bowdin, S., Hurles, M. E., & Mital, S. (2016). Exome sequencing identifies rare variants in multiple genes in atrioventricular septal defect. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *18*(2), 189–198.

De Luca, A., Sarkozy, A., Consoli, F., Ferese, R., Guida, V., Dentici, M. L., Mingarelli, R., Bellacchio, E., Tuo, G., Limongelli, G., Digilio, M. C., Marino, B., & Dallapiccola, B. (2010). Familial transposition of the great arteries caused by multiple mutations in laterality genes. *Heart*, *96*(9), 673–677.

den Dunnen, J. T., Dalgleish, R., Maglott, D. R., Hart, R. K., Greenblatt, M. S., McGowan-Jordan, J., Roux, A.-F., Smith, T., Antonarakis, S. E., & Taschner, P. E. M. (2016). HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Human Mutation*, *37*(6), 564–569.

Dentice, M., Cordeddu, V., Rosica, A., Ferrara, A. M., Santarpia, L., Salvatore, D., Chiovato, L., Perri, A., Moschini, L., Fazzini, C., Olivieri, A., Costa, P., Stoppioni, V., Baserga, M., De Felice, M., Sorcini, M., Fenzi, G., Di Lauro, R., Tartaglia, M., & Macchia, P. E. (2006). Missense mutation in the transcription factor NKX2-5: a novel molecular event in the pathogenesis of thyroid dysgenesis. *The Journal of Clinical Endocrinology and Metabolism*, *91*(4), 1428–1433.

Desvignes, J.-P., Bartoli, M., Delague, V., Krahn, M., Miltgen, M., Béroud, C., & Salgado, D. (2018). VarAFT: a variant annotation and filtration system

for human next generation sequencing data. *Nucleic Acids Research*, *46*(W1), W545–W553.

Draus, J. M., Jr, Hauck, M. A., Goetsch, M., Austin, E. H., 3rd, Tomita-Mitchell, A., & Mitchell, M. E. (2009). Investigation of somatic NKX2-5 mutations in congenital heart disease. *Journal of Medical Genetics*, *46*(2), 115–122.

Dupays, L., Shang, C., Wilson, R., Kotecha, S., Wood, S., Towers, N., & Mohun, T. (2015). Sequential Binding of MEIS1 and NKX2-5 on the Popdc2 Gene: A Mechanism for Spatiotemporal Regulation of Enhancers during Cardiogenesis. *Cell Reports*, *13*(1), 183–195.

Durocher, D., Charron, F., Warren, R., Schwartz, R. J., & Nemer, M. (1997). The cardiac transcription factors Nkx2-5 and GATA-4 are mutual cofactors. *The EMBO Journal*, *16*(18), 5687–5696.

El-Bouchikhi, I., Belhassan, K., Moufid, F. Z., Houssaini, M. I., Ouldim, K., & Atmani, S. (2017). Novel NKX2-5 germline mutation in a Moroccan child with transitional atrio-ventricular septal defect (tAVSD). *The Turkish Journal of Pediatrics*, *59*(5), 610–613.

Ellesøe, S. G., Johansen, M. M., Bjerre, J. V., Hjortdal, V. E., Brunak, S., & Larsen, L. A. (2016). Familial Atrial Septal Defect and Sudden Cardiac Death: Identification of a Novel NKX2-5 Mutation and a Review of the Literature. *Congenital Heart Disease*, *11*(3), 283–290.

Elliott, D. A., Kirk, E. P., Schaft, D., & Harvey, R. P. (2010). Chapter 9.1 - NK-2 Class Homeodomain Proteins: Conserved Regulators of Cardiogenesis. In N. Rosenthal & R. P. Harvey (Eds.), *Heart Development and Regeneration* (pp. 569–597). Academic Press.

Elliott, D. A., Kirk, E. P., Yeoh, T., Chandar, S., McKenzie, F., Taylor, P., Grossfeld, P., Fatkin, D., Jones, O., Hayes, P., Feneley, M., & Harvey, R. P. (2003). Cardiac homeobox gene NKX2-5 mutations and congenital heart disease: associations with atrial septal defect and hypoplastic left heart syndrome. *Journal of the American College of Cardiology*, *41*(11), 2072–2076.

Elliott, D. A., Solloway, M. J., Wise, N., Biben, C., Costa, M. W., Furtado, M. B., Lange, M., Dunwoodie, S., & Harvey, R. P. (2006). A tyrosine-rich domain within homeodomain transcription factor Nkx2-5 is an essential element in the early cardiac transcriptional regulatory machinery. *Development*, *133*(7), 1311–1322.

El Malti, R., Liu, H., Doray, B., Thauvin, C., Maltret, A., Dauphin, C., Gonçalves-Rocha, M., Teboul, M., Blanchet, P., Roume, J., Gronier, C., Ducreux, C., Veyrier, M., Marçon, F., Acar, P., Lusson, J.-R., Levy, M., Beyler, C., Vigneron, J., … Bouvagnet, P. (2016). A systematic variant screening in familial cases of congenital heart defects demonstrates the usefulness of molecular genetics in this field. *European Journal of Human Genetics: EJHG*, *24*(2), 228–236.

Esposito, G., Grutter, G., Drago, F., Costa, M. W., De Santis, A., Bosco, G., Marino, B., Bellacchio, E., Lepri, F., Harvey, R. P., Sarkozy, A., & Dallapiccola, B. (2009). Molecular analysis of PRKAG2, LAMP2, and NKX2-5 genes in a cohort of 125 patients with accessory atrioventricular connection. *American Journal of Medical Genetics. Part A*, *149A*(7), 1574–1577.

Evans, S. M. (1999). Vertebrate tinman homologues and cardiac differentiation. *Seminars in Cell & Developmental Biology*, *10*(1), 73–83.

Fagman, H., & Nilsson, M. (2011). Morphogenetics of early thyroid development. *Journal of Molecular Endocrinology*, *46*(1), R33–R42.

Fahed, A. C., Roberts, A. E., Mital, S., & Lakdawala, N. K. (2014). Heart failure in congenital heart disease: a confluence of acquired and congenital. *Heart Failure Clinics*, *10*(1), 219–227.

Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J., & den Dunnen, J. T. (2011). LOVD v.2.0: the next generation in gene variant databases. *Human Mutation*, *32*(5), 557–563.

Gioli-Pereira, L., Pereira, A. C., Mesquita, S. M., Xavier-Neto, J., Lopes, A. A., & Krieger, J. E. (2010). NKX2.5 mutations in patients with non-syndromic congenital heart disease. *International Journal of Cardiology*, *138*(3), 261–265.

Goldmuntz, E., Geiger, E., & Benson, D. W. (2001). NKX2.5 mutations in patients with tetralogy of fallot. *Circulation*, 104(21), 2565–2568.

Gouw, M., Michael, S., Sámano-Sánchez, H., Kumar, M., Zeke, A., Lang, B., Bely, B., Chemes, L. B., Davey, N. E., Deng, Z., Diella, F., Gürth, C.-M., Huber, A.-K., Kleinsorg, S., Schlegel, L. S., Palopoli, N., Roey, K. V., Altenberg, B., Reményi, A., … Gibson, T. J. (2018). The eukaryotic linear motif resource - 2018 update. *Nucleic Acids Research*, *46*(D1), D428–D434.

Granados-Riveron, J. T., Pope, M., Bu'lock, F. A., Thornborough, C., Eason, J., Setchfield, K., Ketley, A., Kirk, E. P., Fatkin, D., Feneley, M. P., Harvey, R. P., & Brook, J. D. (2012). Combined mutation screening of

NKX2-5, GATA4, and TBX5 in congenital heart disease: multiple heterozygosity and novel mutations. *Congenital Heart Disease*, *7*(2), 151–159.

Guntheroth, W., Chun, L., Patton, K. K., Matsushita, M. M., Page, R. L., & Raskind, W. H. (2012). Wenckebach periodicity at rest that normalizes with tachycardia in a family with a NKX2.5 mutation. *The American Journal of Cardiology*, *110*(11), 1646–1650.

Gutierrez-Roelens, I., De Roy, L., Ovaert, C., Sluysmans, T., Devriendt, K., Brunner, H. G., & Vikkula, M. (2006). A novel CSX/NKX2-5 mutation causes autosomal-dominant AV block: are atrial fibrillation and syncopes part of the phenotype? *European Journal of Human Genetics: EJHG*, *14*(12), 1313–1316.

Gutierrez-Roelens, I., Sluysmans, T., Gewillig, M., Devriendt, K., & Vikkula, M. (2002). Progressive AV-block and anomalous venous return among cardiac anomalies associated with two novel missense mutations in the CSX/NKX2-5 gene. *Human Mutation*, *20*(1), 75–76.

Hanley, A., Walsh, K. A., Joyce, C., McLellan, M. A., Clauss, S., Hagen, A., Shea, M. A., Tucker, N. R., Lin, H., Fahy, G. J., & Ellinor, P. T. (2016). Mutation of a common amino acid in NKX2.5 results in dilated cardiomyopathy in two large families. *BMC Medical Genetics*, *17*(1), 83.

Harvey, R. P. (1996). NK-2 homeobox genes and heart development. *Developmental Biology*, *178*(2), 203–216.

Hatemi, A. C., Güleç, C., Cine, N., Vural, B., Hatırnaz, O., Sayitoğlu, M., Oztunç, F., Saltık, L., Kansız, E., & Erginel Ünaltuna, N. (2011). Sequence variations of NKX2-5 and HAND1 genes in patients with atrial

isomerism. *Anadolu Kardiyoloji Dergisi: AKD = the Anatolian Journal of Cardiology*, *11*(4), 319–328.

Hermanns, P., Grasberger, H., Refetoff, S., & Pohlenz, J. (2011). Mutations in the NKX2.5 gene and the PAX8 promoter in a girl with thyroid dysgenesis. *The Journal of Clinical Endocrinology and Metabolism*, *96*(6), E977–E981.

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. In *Nature Genetics* (Vol. 28, Issue 3, pp. 276–280). https://doi.org/10.1038/90123

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., DePalma, S. R., McKean, D., Wakimoto, H., Gorham, J., Jin, S. C., Deanfield, J., Giardini, A., Porter, G. A., Jr, Kim, R., Bilguvar, K., López-Giráldez, F., Tikhonova, I., Mane, S., … 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., Shiojima, I., Hiroi, Y., Harada, M., Murakawa, Y., Hirata, 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., Lv, X.-D., Du, M., & Fan, X.-M. (2010). Molecular mechanisms of congenital heart disease. *Cardiovascular Pathology: The Official Journal of the Society for Cardiovascular Pathology*, *19*(5), e183–e193.

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., Colby, G., Gebreab, F., Gygi, M. P., Parzen, H., Szpyt, J., Tam, S., Zarraga, G., Pontano-Vaites, L., Swarup, S., White, A. E., Schweppe, D. K., Rad, R., Erickson, B. K., … Harper, J. W. (2017). Architecture of the human interactome defines protein communities and disease networks. *Nature*, *545*(7655), 505–509.

Ikeda, Y., Hiroi, Y., Hosoda, T., Utsunomiya, T., Matsuo, S., Ito, T., Inoue, J.-I., Sumiyoshi, T., Takano, H., Nagai, R., & Komuro, I. (2002). Novel point mutation in the cardiac transcription factor CSX/NKX2.5 associated with congenital heart disease. *Circulation Journal: Official Journal ofthe Japanese Circulation Society*, *66*(6), 561–563.

Izumi, K., Noon, S., Wilkens, A., & Krantz, I. D. (2014). NKX2.5 mutation identification on exome sequencing in a patient with heterotaxy. *European Journal of Medical Genetics*, *57*(10), 558–561.

Jarrell, D. K., Lennon, M. L., & Jacot, J. G. (2019). Epigenetics and Mechanobiology in Heart Development and Congenital Heart Disease.

Diseases (Basel, Switzerland), *7*(3). https://doi.org/10.3390/diseases7030052

Jeong, H.-S., Jung, E.-S., Sim, Y.-J., Kim, S.-J., Jang, J.-W., Hong, K.-S., Lee, W.-Y., Chung, H.-M., Park, K.-T., Jung, Y.-S., Kim, C.-H., & Kim, K.-S. (2015). Fbxo25 controls Tbx5 and Nkx2-5 transcriptional activity to regulate cardiomyocyte development. *Biochimica et Biophysica Acta*, *1849*(6), 709–721.

Joseph, P., Kimm, J., Kalyan-Raman, U. P., Nixon, J. P., & Hiller, J. (1990). Terminal deletion of the long arm of chromosome 5. *American Journal of Human Genetics*, *47*(A31).

Kasahara, H., Bartunkova, S., Schinke, M., Tanaka, M., & Izumo, S. (1998). Cardiac and extracardiac expression of Csx/Nkx2.5 homeodomain protein. *Circulation Research*, *82*(9), 936–946.

Kasahara, H., & Benson, D. W. (2004). Biochemical analyses of eight NKX2.5 homeodomain missense mutations causing atrioventricular block and cardiac anomalies. *Cardiovascular Research*, *64*(1), 40–51.

Kasahara, H., & Izumo, S. (1999). Identification of the in vivo casein kinase II phosphorylation site within the homeodomain of the cardiac tisue-specifying homeobox gene product Csx/Nkx2.5. *Molecular and Cellular Biology*, *19*(1), 526–536.

Kasahara, H., Lee, B., Schott, J. J., Benson, D. W., Seidman, J. G., Seidman, C. E., & Izumo, S. (2000). Loss of function and inhibitory effects of human CSX/NKX2.5 homeoprotein mutations associated with congenital heart disease. *The Journal of Clinical Investigation*, *106*(2), 299–308.

Khatami, M., Mazidi, M., Taher, S., Heidari, M. M., & Hadadzadeh, M. (2018). Novel Point Mutations in the NKX2.5 Gene in Pediatric Patients with Non-Familial Congenital Heart Disease. *Medicina*, *54*(3). https://doi.org/10.3390/medicina54030046

Kim, E. Y., Chen, L., Ma, Y., Yu, W., Chang, J., Moskowitz, I. P., & Wang, J. (2011). Expression of sumoylation deficient Nkx2.5 mutant in Nkx2.5 haploinsufficient mice leads to congenital heart defects. *PloS One*, *6*(6), e20803.

Kim, T.-G., Chen, J., Sadoshima, J., & Lee, Y. (2004). Jumonji represses atrial natriuretic factor gene expression by inhibiting transcriptional activities of cardiac transcription factors. *Molecular and Cellular Biology*, *24*(23), 10151–10160.

Kleczkowska, A., Fryns, J. P., & van den Berghe, H. (1993). A distinct multiple congenital anomalies syndrome associated with distal 5q deletion (q35.1qter). *Annales de Genetique*, *36*(2), 126–128.

Kodo, K., Nishizawa, T., Furutani, M., Arai, S., Ishihara, K., Oda, M., Makino, S., Fukuda, K., Takahashi, T., Matsuoka, R., Nakanishi, T., & Yamagishi, H. (2012). Genetic analysis of essential cardiac transcription factors in 256 patients with non-syndromic congenital heart defects. *Circulation Journal: Official Journal of the Japanese Circulation Society*, *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 homeobox transcription factor gene NKX2.5: identification of a

novel mutation. *Clinical Research in Cardiology: Official Journal of the German Cardiac Society*, *95*(9), 499–503.

Koss, M., Bolze, A., Brendolan, A., Saggese, M., Capellini, T. D., Bojilova, E., Boisson, B., Prall, O. W. J., Elliott, D. A., Solloway, M., Lenti, E., Hidaka, C., Chang, C.-P., Mahlaoui, N., Harvey, R. P., Casanova, J.-L., & 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., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., … 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., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., … 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.

Liu, X.-Y., Wang, J., Yang, Y.-Q., Zhang, Y.-Y., Chen, X.-Z., Zhang, W., Wang, X.-Z., Zheng, J.-H., & 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. Chinese journal of pediatrics*, *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.

Liu, Y.-N., Cheng, S.-S., Wang, C., Xing, D.-X., Liu, Y., & Tan, X.-J. (2015). Experimental and theoretical studies of the products of addition-elimination reactions between benzil dihydrazone and three isomeric chlorobenzaldehydes. *Acta Crystallographica. Section B: Structural Crystallography and Crystal Chemistry*, *71*(Pt 7), 554–563.

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

Li, Y., Yagi, H., Onuoha, E. O., Damerla, R. R., Francis, R., Furutani, Y., Tariq, M., King, S. M., Hendricks, G., Cui, C., Saydmohammed, M., Lee, D. M., Zahid, M., Sami, I., Leatherbury, L., Pazour, G. J., Ware, S. M., Nakanishi, T., Goldmuntz, E., … Lo, C. W. (2016). DNAH6 and Its Interactions with PCD Genes in Heterotaxy and Primary Ciliary Dyskinesia. *PLoS Genetics*, *12*(2), e1005821.

Luna-Zurita, L., Stirnimann, C. U., Glatt, S., Kaynak, B. L., Thomas, S., Baudin, F., Samee, M. A. H., He, D., Small, E. M., Mileikovsky, M., Nagy, A., Holloway, A. K., Pollard, K. S., Müller, C. W., & Bruneau, B. G. (2016). Complex Interdependence Regulates Heterotypic Transcription Factor Distribution and Coordinates Cardiogenesis. *Cell*, *164*(5), 999–1014.

Lyons, I., Parsons, L. M., Hartley, L., Li, R., Andrews, J. E., Robb, L., & Harvey, R. P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes & Development*, *9*(13), 1654–1666.

Matsuoka, R. (2005). Mutations of transcription factors in human with heart disease for understanding the development and mechanisms of congenital cardiovascular heart disease. *Advances in Experimental Medicine and Biology*, *565*, 349–357; discussion 405–415.

Mattapally, S., Singh, M., Murthy, K. S., Asthana, S., & Banerjee, S. K. (2018). Computational modeling suggests impaired interactions between NKX2.5 and GATA4 in individuals carrying a novel pathogenic D16N NKX2.5 mutation. *Oncotarget*, *9*(17), 13713–13732.

McElhinney, D. B., Geiger, E., Blinder, J., Benson, D. W., & Goldmuntz, E. (2003). NKX2.5 mutations in patients with congenital heart disease. *Journal of the American College of Cardiology*, *42*(9), 1650–1655.

Mottaz, A., David, F. P. A., Veuthey, A.-L., & Yip, Y. L. (2010). Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*, *26*(6), 851–852.

Muntean, I., Togănel, R., & Benedek, T. (2017). Genetics of Congenital Heart Disease: Past and Present. *Biochemical Genetics*, *55*(2), 105–123.

Ouyang, P., Saarel, E., Bai, Y., Luo, C., Lv, Q., Xu, Y., Wang, F., Fan, C., Younoszai, A., Chen, Q., Tu, X., & Wang, Q. K. (2011). A de novo mutation in NKX2.5 associated with atrial septal defects, ventricular noncompaction, syncope and sudden death. *Clinica Chimica Acta; International Journal of Clinical Chemistry*, *412*(1-2), 170–175.

Ouyang, P., Zhang, H., Fan, Z., Wei, P., Huang, Z., Wang, S., & Li, T. (2016). A R/K-rich motif in the C-terminal of the homeodomain is required for complete translocating of NKX2.5 protein into nucleus. *Gene*, *592*(2), 276–280.

Pabst, S., Wollnik, B., Rohmann, E., Hintz, Y., Glänzer, K., Vetter, H., Nickenig, G., & Grohé, C. (2008). A novel stop mutation truncating critical regions of the cardiac transcription factor NKX2-5 in a large family with autosomal-dominant inherited congenital heart disease. *Clinical Research in Cardiology: Official Journal of the German Cardiac Society*, *97*(1), 39–42.

Pauli, R. M., Scheib-Wixted, S., Cripe, L., Izumo, S., & Sekhon, G. S. (1999). Ventricular Noncompaction and Distal Chromosome 5q Deletion

American Journal of Medical Genetics. *American Journal of Medical Genetics*, *85*, 419–423.

Peng, T., Wang, L., Zhou, S.-F., & Li, X. (2010). Mutations of the GATA4 and NKX2.5 genes in Chinese pediatric patients with non-familial congenital heart disease. *Genetica*, *138*(11-12), 1231–1240.

Perera, J. L., Johnson, N. M., Judge, D. P., & Crosson, J. E. (2014). Novel and highly lethal NKX2.5 missense mutation in a family with sudden death and ventricular arrhythmia. *Pediatric Cardiology*, *35*(7), 1206–1212.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612.

Pradhan, L., Genis, C., Scone, P., Weinberg, E. O., Kasahara, H., & Nam, H.-J. (2012). Crystal structure of the human NKX2.5 homeodomain in complex with DNA target. *Biochemistry*, *51*(32), 6312–6319.

Pradhan, L., Gopal, S., Li, S., Ashur, S., Suryanarayanan, S., Kasahara, H., & Nam, H.-J. (2016). Intermolecular Interactions of Cardiac Transcription Factors NKX2.5 and TBX5. *Biochemistry*, *55*(12), 1702–1710.

Pradhan, L., Gopal, S., & Nam, H. J. (2014). Crystallization and preliminary X-ray analysis of the cardiac transcription factor complex of NKX2.5 and TBX5 with DNA. *Acta Crystallographica. Section F, Structural Biology and Crystallization Communications*, *70*(Pt 5), 592–595.(5), 405–424.

Rifai, L., Maazouzi, W., & Sefiani, A. (2007). Novel point mutation in the NKX2-5 gene in a Moroccan family with atrioventricular conduction

disturbance and an atrial septal defect in the oval fossa. *Cardiology in the Young*, *17*(1), 107–109.

Ritchie, M. D., Rowan, S., Kucera, G., Stubblefield, T., Blair, M., Carter, S., Roden, D. M., & Darbar, D. (2012). Chromosome 4q25 variants are genetic modifiers of rare ion channel mutations associated with familial atrial fibrillation. *Journal of the American College of Cardiology*, *60*(13), 1173–1181.

Sarkozy, A., Conti, E., Neri, C., D'Agostino, R., Digilio, M. C., Esposito, G., Toscano, A., Marino, B., Pizzuti, A., & Dallapiccola, B. (2005). Spectrum of atrial septal defects associated with mutations of NKX2.5 and GATA4 transcription factors. *Journal of Medical Genetics*, *42*(2), e16.

Schott, J. J., Benson, D. W., Basson, C. T., Pease, W., Silberbach, G. M., Moak, J. P., Maron, B. J., Seidman, C. E., & Seidman, J. G. (1998). Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science*, *281*(5373), 108–111.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., & Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, *29*(1), 308–311.

Shiojima, I., Komuro, I., Inazawa, J., Nakahori, Y., Matsushita, I., Abe, T., Nagai, R., & Yazaki, Y. (1995). Assignment of cardiac homeobox gene CSX to human chromosome 5q34. *Genomics*, *27*(1), 204–206.

Shiojima, I., Komuro, I., Mizuno, T., Aikawa, R., Akazawa, H., Oka, T., Yamazaki, T., & Yazaki, Y. (1996). Molecular cloning and characterization of human cardiac homeobox gene CSX1. *Circulation Research*, *79*(5), 920–929.

Song, K., Backs, J., McAnally, J., Qi, X., Gerard, R. D., Richardson, J. A., Hill, J. A., Bassel-Duby, R., & Olson, E. N. (2006). The transcriptional coactivator CAMTA2 stimulates cardiac growth by opposing class II histone deacetylases. *Cell*, *125*(3), 453–466.

Stallmeyer, B., Fenge, H., Nowak-Göttl, 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., Mei, L., Tang, J., Kumar, M., Wang, X., Su, L., & 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., Yamasaki, N., Kawamoto, T., Gehrmann, J., Maguire, C. T., Schinke, M., Seidman, C. E., Seidman, J. G., Kurachi, Y., & 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 Symposia on Quantitative Biology*, *67*, 317–325.

Tang, X., Ma, H., Han, L., Zheng, W., Lu, Y.-B., Chen, X.-F., Liang, S.-T., Wei, G.-H., Zhang, Z.-Q., Chen, H.-Z., & 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., Rahman, T. J., Eloranta, J. J., Jüngst, C., Stuart, A. G., O'Sullivan, J., Keavney, B. D., & Goodship, J. A. (2014). Funcionally 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., Bouma, 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., Slager, M. A., 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., Mazzarotto, F., Blair, E., Seller, A., Taylor, J. C., Minikel, E. V., Exome Aggregation Consortium, MacArthur, D. G., Farrall, M., Cook, S.

A., & Watkins, H. (2017). Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genetics in Medicine: Official Journal of the American College of Medical Genetics*, *19*(2), 192–203.

Wang, F., Liu, C., Jia, X., Liu, X., Xu, Y., Yan, S., Jia, X., Huang, Z., Liu, 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; International Journal of Clinical Chemistry*, *470*, 36–41.

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

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

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

Wang, J., Zhang, H., Iyer, 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 NKX2.5 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., Barlow, R. S., Sureka, A. O., LaPage, M. J., Janss, L. L., & 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., Liu, X., & Yang, Y.-Q. (2013). 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., Shi, H.-Y., Liu, H., Yang, X.-X., Xu, Y.-J., Qu, X.-K., & Yang, Y.-Q. (2017). 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., Qu, X.-K., Liu, X., Huang, R.-T., Xue, S., Yang, Y.-Q., & Li, R.-G. (2017). 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., Liu, X., Fang, W.-Y., Yang, Y.-Q., & 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., Li, R.-G., Xu, L., Jiang, W.-F., Qiu, X.-B., Jiang, J.-Q., Qu, X.-K., Liu, X., Fang, W.-Y., Jiang, J.-F., & 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., Skerjanc, 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., Skerjanc, 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., Toko, H., Yazaki, Y., Nagai, R., & Komuro, I. (2000). Functional analyses of three Csx/Nkx-2.5 mutations that cause human congenital heart disease. *The Journal of Biological Chemistry*, *275*(45), 35291–35296.

FIGURES

Figure 1. Compilation of the genetic variants. Pipeline used to compile the different genetic variants is shown as a workflow. Databases in orange are general databases from which frequency information was obtained. Databases in green are clinical databases from which phenotype information was obtained. Publications and the cohort of CHD patients also supplied phenotype information. The databases listed are: NCBI's ClinVar (https://www.ncbi.nlm.nih.gov/clinvar) (Landrum et al., 2018); 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 et al., 2010); ExAC (https://exac.broadinstitute.org/) (Lek et al., 2016); LOVD (http://www.lovd.nl/3.0/home/) (Fokkema et al., 2011). CHD: Congenital heart disease; GV: Genetic variant.

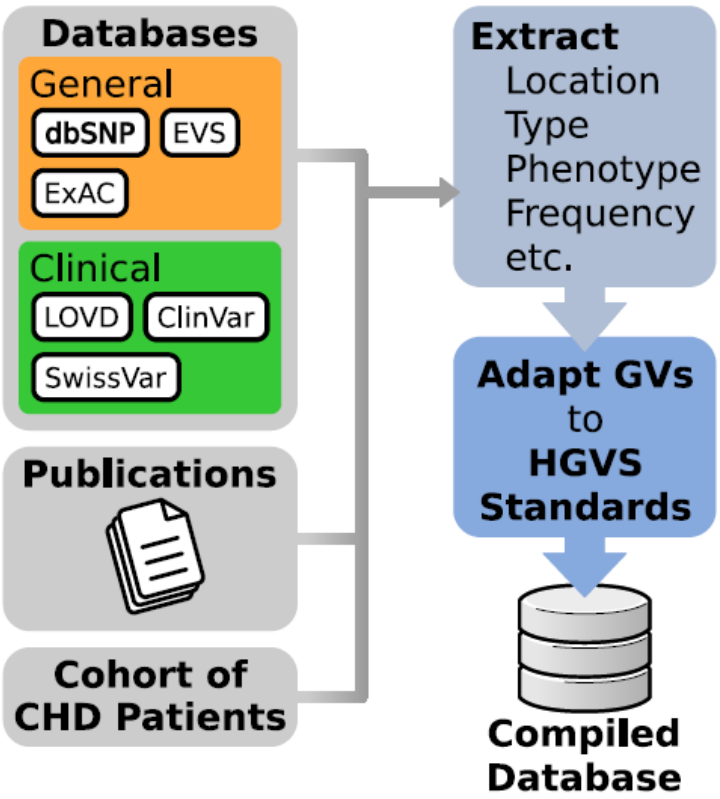

Figure 2. Distribution of genetic variants over the NKX2-5 gene and in the protein. (A) Genetic variants in the compiled database are depicted along with its location in the NKX2-5 gene structure. Boxes represent the two exons of the gene, while the thin line represents the intron. The analysis was limited to the region encompassing the whole canonical transcript of NKX2-5 (chr5: 173,232,109-173,235,311, RefSeq: NC_000005.10). Every line above the NKX2-5 gene represents the position of each of the genetic variants: in blue, all the 1380 variants retrieved; in red, variants classified as pathogenic; in green, variants retrieved from general databases with a frequency above 1/10,000. The two pathogenic splice site variants are detailed below, represented by their "c." descriptor (RefSeq: NM_004378.1). 3'UTR and 5'UTR: 3' and 5' untranslated region; GV: Genetic variant; TR: translated region. (B) The positions of pathogenic genetic variants from the compiled database are shown. In the middle, a representation of the NKX2-5 protein highlighting the known regions/domains of the protein. On the upper part of the figure, every line in red represents the position of each genetic variant classified as pathogenic in the translated region, while on the bottom, the same genetic variants are noted by their "p." descriptor (RefSeq: NP_004378.1). If two different variants cause the same effect on the protein sequence, the "p." identifier is noted only once. The line graphs were generated with the PyGame library in the Python 3 programming language (http://www.python.org). HD: Homeodomain; NK2-SD: NK2-Specific domain; TN: Tinman domain; YRR: Tyrosine-rich region.

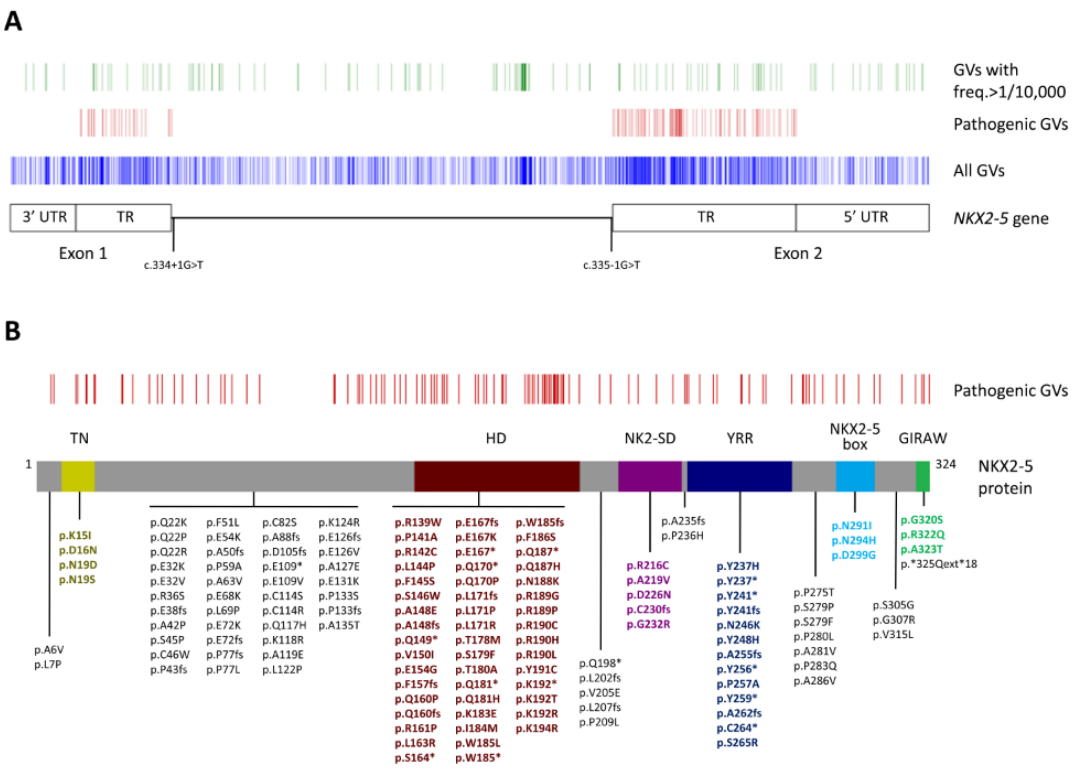

Figure 3. Analysis of distribution of genetic variants in the NKX2-5 protein. Number of missense variants along the canonical NKX2-5 protein (residues 1-324), in windows of three residues. In green, variants from the general databases with a frequency over 1/10,000 individuals. In red, variants classified as pathogenic. Thick lines below represent different regions/domains of the protein: in yellow, the Tinman domain; in bordeaux, the homeodomain; in red the 3 alpha-helices of the homeodomain; in purple, the NK2-specific domain; in blue, the Tyrosine-rich region (YRR); in cyan, the NKX2-5 box motif; in green, the GIRAW motif. The figure was generated with the PyPlot library in the Python 3 programming language (http://www.python.org). GVs: Genetic variants.

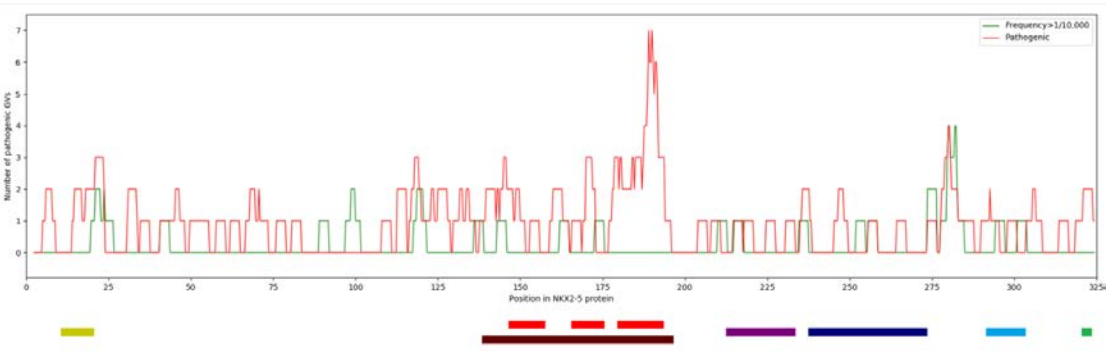

Figure 4. Tridimensional structure of the NKX2-5 Homeodomain and its interaction with DNA. The image was generated with help from UCSF Chimera (http://www.rbvi.ucsf.edu/chimera). A. Schematic 3D representation of the NKX2-5 homeodomain along with a segment of the ANF promoter (PDB ID 3RKQ, B chain). In red, residues that contain genetic variants classified as pathogenic. H1-3: Helixes 1 to 3. B. Pipe and planks representation of the NKX2-5 homeodomain. Helices 1 to 3 are represented as yellow pipes (H1-3). Loops between helices are shown in blue. The hydrophobic residues that point towards the inner face or "hydrophobic core"

of the alpha helices are shown in beige. C. Schematic representation of the ANF sequence of DNA 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.

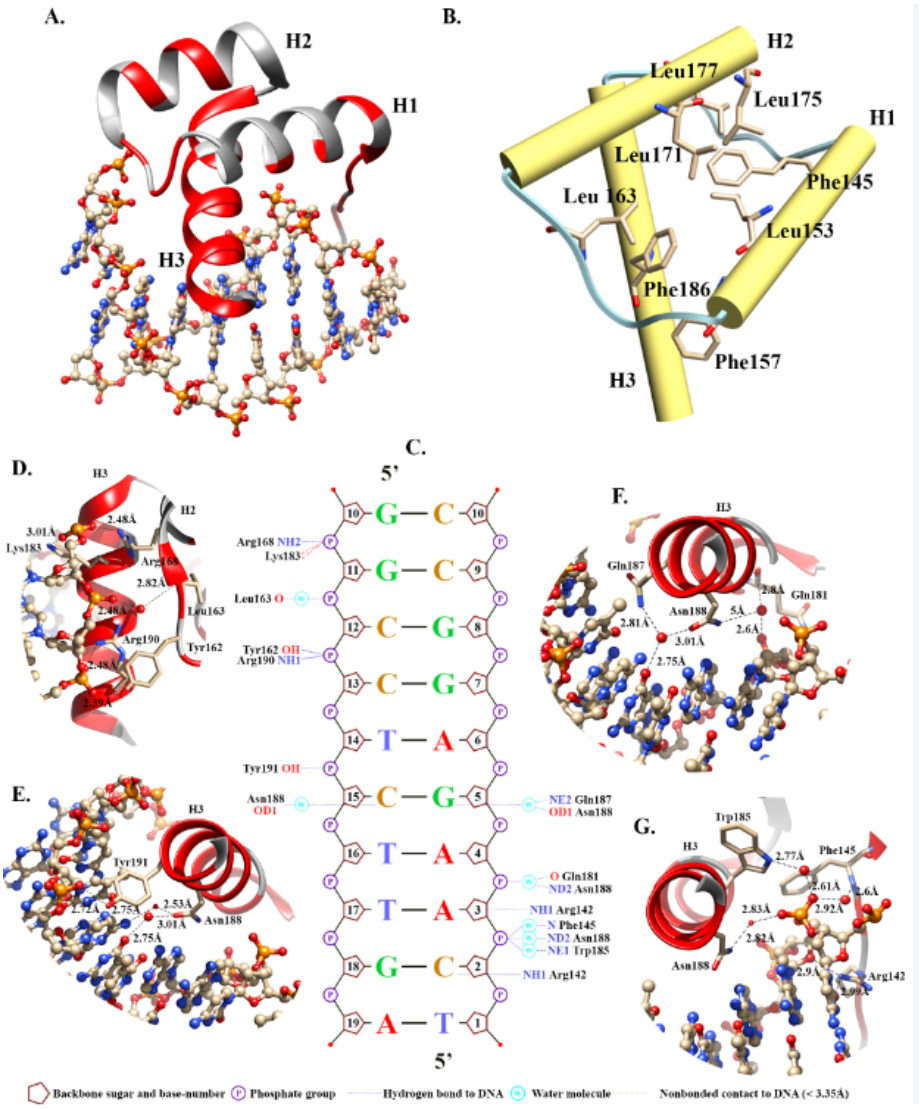

Table 1. Types of genetic variants in the different regions/domains of the *NKX2-5* gene.

			Missense	Synonymous	Nonsense	Inframe	Frameshift	Stop loss	Total
	Tinman		10 (4)	6 (0)	0 (0)	0 (0)	1 (0)	0 (0)	17 (4)
Translated region	**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 (47)	92 (0)	3 (2)	8 (0)	18 (12)	1 (1)	320 (62)

	Total	**389 (101)**	**169 (0)**	**20 (17)**	**14 (0)**	**33 (23)**	**1 (1)**	**626 (142)**
Untranslated region								**755 (2)**
							Total:	**1381 (144)**

Number and type of unique genetic variants in the compiled database, grouped by protein region/domain. In parenthesis, number of pathogenic variants. HD: Homeodomain; NK2-SD: NK2-Specific Domain; YRR: Tyrosine-Rich Region.

Table 2. Genetic variants classified as pathogenic in the compiled database.

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
c.973T>C	**p.(*325Glnext*18)**	**GIRAW**		**ASD, VSD, AVSD**	**VUS**		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,

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							2004; Reamon-Buettner et al., 2004
c.943G>T	p.Val315Leu		0.035	ToF	VUS	No effect	Rauch et al., 2010
c.919G>A	**p.(Gly307Arg)**			**VSD, PS**	**VUS**		Pulignani et al., 2018
c.913A>G	p.(Ser305Gly)			VSD	VUS		Reamon-Buettner and Borlak, 2004; Reamon-Buettner et al., 2004
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,

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							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.835T>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-**		**TD**	**P**	**TA**	Hermanns et al.,

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
		Rich					2011
c.792C>A	**p.(Cys264*)**	**Tyr-Rich**		**ASD, AVB**	**P**		Ikeda et al., 2002
c.783del	p.(Ala262Argfs*32)	Tyr-Rich		AVB	P		Gunthero th et al., 2012
c.777C>G	p.(Tyr259*)	Tyr-Rich			P		Predicted by ClinVar
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		VSD	LP		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., 2006
c.762del	**p.(Ala255Profs*39)**	**Tyr-Rich**		**ASD, AVB**	**P**		Stallmey er et al., 2010
c.742T>C	p.(Tyr248His)	Tyr-Rich		VSD	VUS		Reamon-Buettner and Borlak, 2004; Reamon-Buettner et al., 2004

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
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.(Tyr241Trpfs*51)	Tyr-Rich		ASD, AVB, VT, DCM, SD	P		El Malti et al., 2016
c.723C>G	p.(Tyr241*)	Tyr-Rich			P		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	P	TA	Koss et al., 2012
c.701_702insTCCCT	**p.(Ala235Profs*61)**			**ASD, AVB**	**P**		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			P		Predicted by ClinVar

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
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 et al., 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)**			**ASD**	**LP**		Wang et al., 2011[a]
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

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
c.605_606del	p.(Leu202Argfs*49)			ASD, VSD, AVB	P		Sarkozy et al., 2005
c.592C>T	p.Gln198*			ASD, AVB, SD	P	TA[1], DB	Schott et al., 1998; Hosoda et al., 1999; Kasahara et al., 2000; 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

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
c.572A>G	p.Tyr191Cys	HD (H3)		ASD, VSD, AVB	P	TA, DB	Benson et al., 1999; Kasahara et al., 2000; Kasahara and Benson, 2004; Costa et al., 2013
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 et al., 2000; Kasahara and Benson, 2004
c.564C>A	p.Asn188Lys	HD (H3)		ASD, EA, TVA, AVB	P	TA, D**B**	Benson et al., 1999; Kasahara et al.,

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							2000; Kasahara and Benson, 2004
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
c.554_555insC	p.(Trp185Cysfs*67)	HD (H3)			P		Predicted by ClinVar
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**	**TA, DB stability[2]**	Costa et al., 2013; Hanley et al., 2016

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
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		Perera et al., 2014
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**	Yu et al., 2014
c.536C>T	**p.(Ser179Phe)**	**HD (H3)**		**ASD, VSD, AVB**	**P**		Liu et al., 2009a, 2009b, 2011
c.533C>T	p.Thr178Met	HD		ASD, VSD, HLHS, AVB, SSS	P	TA, DB	Schott et al., 1998; Kasahara et al., 2000; Zhu et al., 2000; Elliott et al., 2003; Kasahara and Benson, 2004; Reamon-Buettner and Borlak, 2004; Reamon-Buettner et al., 2004;

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							Hirayama-Yamada et al., 2005; 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***	Kasahara and Benson, 2004
c.510_511dup	**p.Leu171Argfs*6**	**HD (H2)**		**ASD, AVB, SD**	**P**	**TA, localization[3]**	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**	Schott et al., 1998; Kasahara et al., 2000; Zhu et al., 2000; Hatemi et al., 2011
c.499G>T	**p.Glu167***	**HD (H2)**		**ASD, AVB, DCM**	**P**	**TA**	Xu et al., 2017a
c.499G>A	**p.(Glu167Lys)**	**HD (H2)**		**ASD, VSD, PA, VNC**	**LP**		Bermúdez-Jiménez et al., 2017

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	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.478_480delinsGTACCGTT	p.(Gln160Valfs*18)	HD		CHD	P		Predicted by ClinVar
c.479A>C	**p.(Gln160Pro)**	**HD**		**ASD, AVB**	**P**		Rifai et al., 2007
c.471_472del	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)**		**ASD, VSD, ToF, AVB**	**P**	**TA, DB, D**	Benson et al., 1999; Kasahara et al., 2000; Bjørnstad and Leren,

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							2009
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	Ritchie et al., 2012; Huang et al., 2013
c.431T>C	p.(Leu144Pro)	HD	4.65	ASD, AVSD	VUS		Reamon-Buettner and Borlak, 2004; Reamon-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,**	**P**	**TA**	Xu et al., 2017[a]

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
				DCM			
c.403G>A	p.(Ala135Thr)			ASD, AVSD	VUS		Reamon-Buettner and Borlak, 2004; Reamon-Buettner et al., 2004
c.397_400del	**p.(Pro133Glyfs *42)**			**AVSD, DORV, AVR, HT, AP**	**LP**		Izumi et al., 2014
c.397C>T	p.(Pro133Ser)			VSD	VUS		Reamon-Buettner and Borlak, 2004; Reamon-Buettner et al., 2004
c.391G>A	p.(Glu131Lys)			ASD	VUS		Predicted by ClinVar
c.380C>A	p.(Ala127Glu)			ASD	LP		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

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
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;

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							Reamon-Buettner et al., 2004
c.335-1G>T	p.?	splice variant		ASD, VSD	P		Predicted by ClinVar
c.334+1G>T	p.?	splice variant		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, 2005
c.244T>A	p.Cys82Ser		0.017	IAVC	VUS	TA[4]	Esposito et al., 2009
c.230C>T	p.(Pro77Leu)			VSD	VUS		Reamon-Buettner and Borlak, 2004; Reamon-Buettner et al.,

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
							2004
c.228_229del	p.(Pro77Phefs*30)			ASD, AVB	P		Watanabe et al., 2002
c.215_221del	p.(Glu72Alafs*102)			ASD, AVB, AF, HT, PP	P		Watanabe et al., 2002
c.214G>A	**p.(Glu72Lys)**			**ASD**	**VUS**		Tian et al., 2008; Mattapally et al., 2018
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_163delinsGCCTCCT	p.(Ala50Profs*123)			ASD	P		Predicted by ClinVar
c.160G>A	**p.(Glu54Lys)**			**ToF**	**LP**		Wang et al., 2011b

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
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., 2011b
c.95A>T	**p.(Glu32Val)**			**VSD, ToF**	**P**		Khatami et al., 2018

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
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
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

cDNA variant	Protein variant	Affected region	Frequency (per 1000)	Associated phenotype	ACMG Classification	In vitro effect impaired	References
c.17C>T	**p.Ala6Val**			**ToF**	**P**	**TA**	Kodo et al., 2012

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. 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: Atrioventricular 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: Alpha-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 región/domain; VF: Ventricular fibrillation; VNC: Ventricular noncompaction; VSD: Ventricular septal defect; VT: Ventricular tachycardia; VUS: Variant of uncertain significance. [1]: In some experiments, diminished/weaker, in some others increased/stronger. [2]: Increased protein stability. [3]: Protein in nucleus and cytoplasm. [4]: Small effect. *: showed decreased interaction also with GATA4 and TBX5. ** showed decreased interaction with GATA4.

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; Wang et al., 2008; Kim et al., 2011
Acetylation site	183	1	p.(K183E)	Li et al., 2007; Tang et al., 2016
CKII Phosphorylation	161-167	6	p.R161P, p.(L163R), p.(S164*), p.(E167fs),	ELM; Kasahara and Izumo, 1999

motif			p.(E167K), p.E167*	
Nuclear localization signal	136-143	4	**p.(A135T)**, p.R139W, p.(P141A), p.R142C	ELM; Kasahara and Izumo, 1999
Nuclear localization signal	192-199	5	p.K192*, p.(K192T), p.(K192R), p.(K194R), **p.Q198***	ELM; Ouyang et al., 2016

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). CKII: Casein kinase II; ELM: Predicted by the Eukaryotic Linear Motif resource (http://elm.eu.org/) (Gouw et al., 2018); GV: Genetic variant.