

# SCIENTIFIC REPORTS



OPEN

## Partial microduplication in the histone acetyltransferase complex member *KANSL1* is associated with congenital heart defects in 22q11.2 microdeletion syndrome patients

Luis E. León<sup>1</sup>, Felipe Benavides<sup>1</sup>, Karen Espinoza<sup>1</sup>, Cecilia Vial<sup>1</sup>, Patricia Alvarez<sup>2</sup>, Mirta Palomares<sup>3,4</sup>, Guillermo Lay-Son<sup>1,5</sup>, Macarena Miranda<sup>4</sup> & Gabriela M. Repetto<sup>1,5</sup>

22q11.2 microdeletion syndrome (22q11.2DS) is the most common microdeletion disorder in humans, with an incidence of 1/4000 live births. It is caused by a heterozygous deletion of 1.5–3 Mb on chromosome region 22q11.2. Patients with the deletion present features that include neuropsychiatric problems, craniofacial abnormalities and cardiovascular malformations. However, the phenotype is highly variable and the factors related to the clinical heterogeneity are not fully understood. About 65% of patients with 22q11.2DS have congenital heart defects (CHD). The main goal of this study was to identify common CNVs in 22q11.2DS patients that could be associated with the incomplete penetrance of CHD. Analysis of genomic DNA from 253 patients with 22q11.2DS using array technology showed an association between a microduplication located in region 17q21.31 and CHD (p-value = 0.023, OR = 2.75, 95% CI = 1.17–7.03). This region includes the first three exons of *KANSL1* gene. Bioinformatic analysis showed that *KANSL1* and *CRKL*, a gene in the commonly deleted region of 22q11.2DS, are part of the same regulatory module in a miRNA-mRNA network. These results show that a *KANSL1* microduplication, in combination with the 22q11.2 deletion, is associated with increased risk of CHD in these patients, suggesting that *KANSL1* plays a role as a modifier gene in 22q11.2DS patients.

22q11.2 deletion syndrome (22q11.2DS, OMIM #192430, #188400) is the most common microdeletion syndrome in humans and is associated with a wide variety of phenotypic features, including congenital heart defects (CHD), craniofacial anomalies, intellectual disability, increased risk of psychosis and schizophrenia, among others<sup>1,2</sup>. Ninety percent of the patients have a ~3 Mb hemizygous deletion and less than 10% of patients have a smaller 1.5 or 2 Mb deletion<sup>3,4</sup>. Interestingly, different deletion sizes have no apparent effects in disease severity<sup>5</sup>. In addition, the phenotype is highly variable even within the patients who share the same 3 Mb deletion. This heterogeneity may involve a genetic component such as single nucleotide variants (SNVs) or copy number variations (CNVs), which could be affecting modifier genes<sup>6</sup>. CNVs are known contributors to complex diseases including CHD<sup>7</sup>, being potential modifiers for several different diseases<sup>8–10</sup>.

Little is known about genetic modifiers that may influence the clinical variability observed in patients with 22q11.2DS. Putative modifier genes have been identified in mouse models and include *FGF8*, *FGF10*, *GBX2*, *PITX2*, *CRKL*, *VEGF*, *TGFβ*, *CHRD* and *SHH* (reviewed in Aggarwal *et al.*)<sup>6</sup>, but their role in humans with 22q11.2DS has not been established. To date, only three studies have performed genome-wide search strategies for modifiers of CHD in 22q11.2DS<sup>11–13</sup>. In these studies, 22q11.2DS patients with and without structural heart defects were compared.

In the first study, Mlynarski *et al.* performed a genome-wide association study using microarray technology. The study used probe-intensity information from microarray plus bioinformatics processing to identify CNVs,

<sup>1</sup>Centro de Genética y Genómica, Facultad de Medicina, Clínica Alemana Universidad del Desarrollo, Santiago, Chile.

<sup>2</sup>Hospital Dr. Roberto del Río, Santiago, Chile. <sup>3</sup>Fundación Gantz, Santiago, Chile. <sup>4</sup>Hospital Dr. Luis Calvo Mackenna, Santiago, Chile. <sup>5</sup>Hospital Padre Hurtado, Santiago, Chile. Correspondence and requests for materials should be addressed to G.M.R. (email: [gretto@udd.cl](mailto:gretto@udd.cl))

finding a statistically significant association between a duplication encompassing the *SLC2A3* gene and CHD phenotype in 22q11.2DS patients ( $p$ -value =  $2.68 \times 10^{-4}$ , OR = 5.08 [95% confidence interval 2.0–17.51]). The authors show that the *SLC2A3* duplication by itself is not pathogenic unless it is inherited in combination with 22q11.2 deletion<sup>11</sup>. Moreover, functional studies in mice show that *SLC2A3* gene is a genetic modifier of the cardiac phenotype. The *SLC2A3* gene, also known as *GLUT3*, encodes a glucose transporter, which is important in tissues with high energy demands and metabolic rates, as cardiac myocytes and brain, among others.

The second study performed whole exome sequencing in 184 patients with 22q11.2DS, finding enrichment for rare deleterious SNVs in *MINA*, *RREB1*, *KDM7A* and *JMJD1* genes in the CHD group compared to the non-CHD group. These genes code for proteins related to epigenetic regulation that activates transcription by demethylation of histones, mainly H3K9 and H3K27. The results of this study suggest that variants in histone modification genes increase the risk of CHD in the presence of deletion<sup>13</sup>.

The third study analyzed rare CNVs using Affymetrix SNP 6.0 technology. The authors demonstrated no significant difference in overall burden of rare CNVs in patients with CHD compared to controls. However, an enrichment of CNVs overlapping with protein-coding cardiac-related genes was found in 22q11.2DS individuals with heart defects ( $n = 607$ ) compared to normal hearts ( $n = 339$ ). Furthermore, network analysis revealed that CNVs in specific cardiac networks, such as Wnt signaling, were overrepresented in 22q11.2DS CHD cases but not in 22q11DS, controls suggesting that specific CNVs located outside of the 22q11.2 region may increase the risk for CHDs in some 22q11DS patients<sup>12</sup>.

Although these genes were identified as genetic modifiers of CHD phenotype, they explain only a small proportion of the incomplete penetrance of CHDs in 22q11.2DS patients, suggesting that other modifiers of this condition exist. To search for new genetic modifiers and to address the effect of CNVs in a Chilean cohort of patients with 22q11.2DS, a genome-wide search was performed using Affymetrix SNP 6.0 arrays (Santa Clara, CA). First, the overall burden of CNVs in 22q11.2DS patients with and without CHD was compared, including absolute CNV counts and genomic length, and then proceeded to perform association analysis between CHD phenotype and CNVs. To minimize false positives, a combination of PennCNV for CNV detection and ParseCNV for association study, followed by a validation of the CNVs with a second independent technique was chosen<sup>14,15</sup>. Results indicated that a microduplication in the first three exons of KAT8 Regulatory NSL Complex Subunit 1 gene (*KANSL1*) is associated with CHD. Finally, a network integrative analysis approach was performed with computational biology tools to identify a biological interaction between *KANSL1* gene and genes within 22q11.2 deletion.

## Results

A genome-wide analysis using Affymetrix SNP 6.0 arrays in cohort one was performed, corresponding to 253 patients with 22q11.2DS. After stringent filter criteria detailed in Methods, 39 samples were excluded from the analysis (28 with CHD, 11 without): 29 for Log R Ratio (LRR) standard deviation greater than 0.45, four for quality control in wave adjustment and six for having over 100 CNV counts. In summary, 214 samples from unrelated subjects with 22q11.2DS (105 CHD and 109 without CHD) passed quality controls. The association between 22q11.2 deletion size and CHD phenotype was first assessed. There were 195 patients (91%) with a 3 Mb deletion (97 CHD and 98 No-CHD), 6 patients (2.8%) with 2 Mb (3 CHD and 3 No-CHD), and 13 (6.1%) with 1.5 Mb deletion (5 CHD and 8 No-CHD) ( $\chi^2 = 0.63$ ,  $p$ -value = 0.73), showing no association between deletion size and CHD phenotype, consistent with previous publications<sup>11,16</sup>.

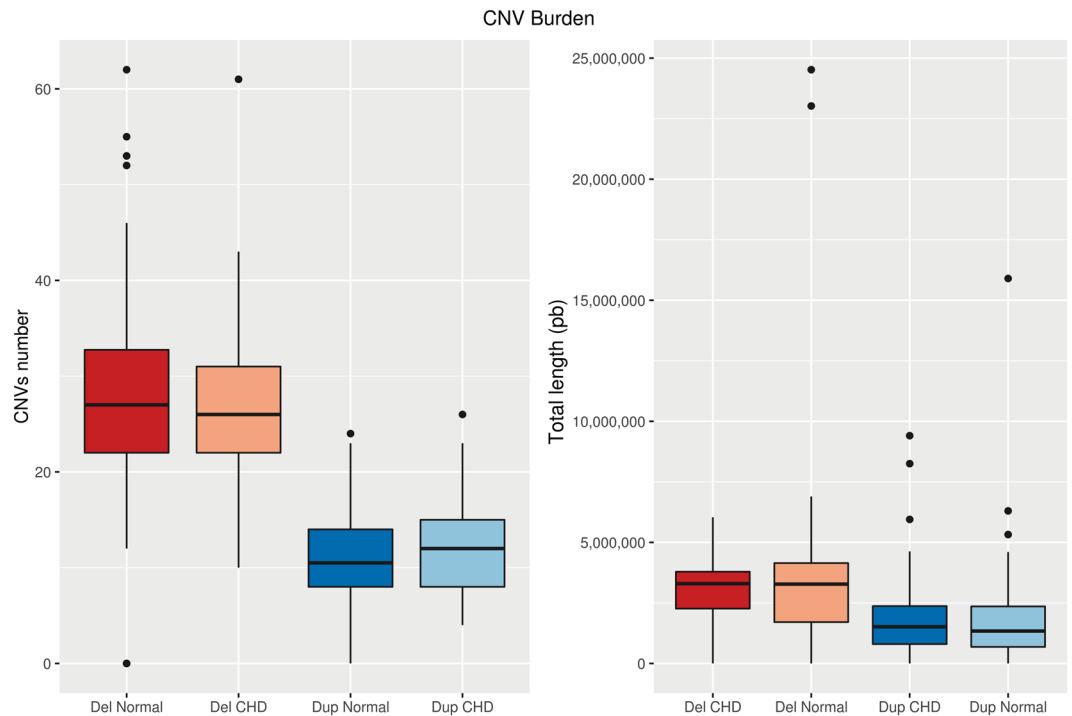
**CNV analysis.** There were no statistically significant differences in global CNV burden between CHD and no-CHD groups. Global CNV burden measures included average CNV length (2.7 Mb for CHD group vs. 2.6 Mb for no-CHD group) and average total number of CNVs per individual (18 for CHD group vs. 19 for no-CHD group). The same analysis was performed comparing only deletions between CHD and no-CHD groups and then comparing only duplications. No statistically significant differences were observed in this analysis (Fig. 1).

Then, the statistical association of CNVs with the cardiac condition was assessed by using the overlapping regions method with the ParseCNV pipeline<sup>15</sup>. None of the CNVs reached the uncorrected  $p$ -value that met conservative multiple-testing significance of association ( $p < 5 \times 10^{-4}$ )<sup>15</sup>. Therefore, the top-ranked CNVs obtained by this method with an uncorrected cutoff of  $p < 0.05$  were selected, in search of biologically plausible associations. Eleven losses and three gains were found with differences in frequencies between CHD and no-CHD groups, ranging in size from 0.122 to 117.4 kb. These CNVs are listed in Table 1 by ascending  $p$ -value.

**Gene prioritization.** To prioritize a candidate CNV region, a network analysis of genes located in the top-ranked 11 deletions and three duplications with phenotype information was performed using the PhenogramViz<sup>17</sup> app available as a plugin for CytoScape<sup>18</sup>. To establish a relationship with the phenotype of interest, we used Human Phenotype Ontology (HPO) terms associated with structural heart defects.

Briefly, PhenogramViz integrates data from different sources (including OMIM, MGI, ZFIN and Orphanet) with the selected HPO terms and then ranks CNVs according to pathogenicity. For this study, the CNV located in chr17:44176984-44272000 ranked as number one because it was the only pathogenic CNV in association with CHD-related HPO terms. This genomic region partially overlaps with the first three exons of *KANSL1*. Observations indicate that *KANSL1* is highly interconnected with the selected HPO terms in the phenogram created with genes overlapping CNV associated regions (Fig. 2).

**Frequency of *KANSL1* duplication.** The frequency of *KANSL1* microduplication was calculated in three Chilean cohorts. The first is the 22q11.2DS cohort; the frequency of CHD was 49% ( $n = 105$ ), and the frequency of *KANSL1* duplication was 12.61% ( $n = 27$ ). In patients with the partial *KANSL1* microduplication ( $n = 27$ , 19 patients with CHD and 8 without CHD), CHD frequency was 70%, showing an enrichment of CHD in these



**Figure 1.** Total CNV burden in 22q11.2DS patients. Standard box-and-whiskers plot for the distribution of the total length and number of CNV (deletions and duplications) in CHD vs. no- CHD. Boxes represent the 1st and 3rd quartiles of each distribution, thick horizontal lines represent the median value, and circles represent outliers.

subjects. This increase in frequency suggests an association between *KANSL1* microduplication and CHD or that *KANSL1* microduplication could be more frequent in 22q11.2DS patients.

To evaluate whether *KANSL1* duplication could be more frequent in 22q11.2DS patients with respect to control samples, CNV enrichment analysis was performed in the 22q11.2DS cohort and compared with respect to two independent cohorts of subjects without 22q11.2DS.

The second cohort corresponded to Chilean adult population controls subjects<sup>19</sup> and showed that *KANSL1* duplication was present in 12.68% of samples, similar to the frequency in the 22q11.2DS cohort. A third cohort corresponding to 40 individuals with congenital anomalies and/or cognitive disabilities who had chromosome microarray (CMA) for clinical reasons<sup>20</sup> was also studied. Eight of these 40 patients had the *KANSL1* duplication (20%) and five (12.5%) had CHD. Only one patient had both traits, the duplication of the *KANSL1* gene and CHD ( $p$  value = 1, OR = 0.85, CI = 0.015–10.37). These results indicate that *KANSL1* duplication and CHD are not statistically associated in patients without 22q11.2DS.

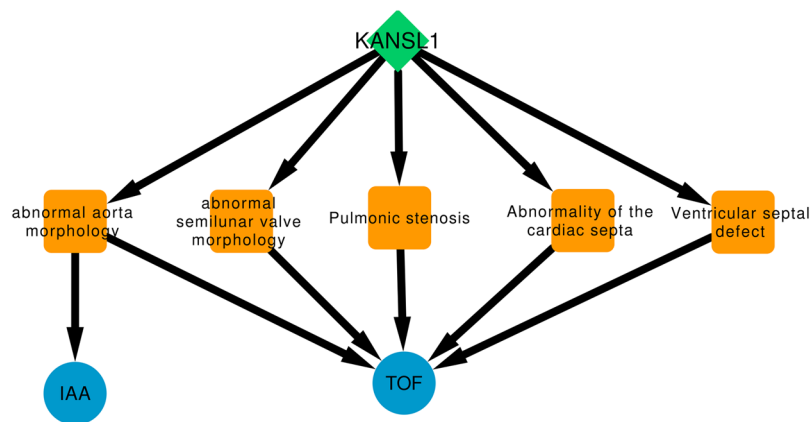
**CNV validation by MLPA.** To validate the microduplication in *KANSL1* by an independent technique, MLPA analysis in all samples with 22q11.2DS using SALSA kit P443 (MRC Holland) was performed. This kit contains probes for all 15 exons of the *KANSL1* gene. The analysis of the samples showed that the duplication includes at least exons 1, 2 and 3 of the *KANSL1* gene. Nevertheless, the design of MLPA probes does not allow to define the exact breakpoints of the duplication.

**Biological interactions networks.** If *KANSL1* is a modifier of CHD in 22q11.2DS patients, there must be a biological relationship with genes located in the 22q11.2 deletion region. To examine this hypothesis, a series of bioinformatic tools related to biological networks were used. Importantly, to be more stringent, only data with experimentally-validated biological interactions were considered.

As a first approach, an analysis was conducted to determine existing protein-protein interaction between *KANSL1* and protein-coding genes located in 22q11.2 deletion using GeneMANIA<sup>21</sup>. This tool showed no direct interactions between *KANSL1* and protein-coding genes located in 22q11.2 deletion. Therefore, a second approach included the search for regulatory interactions, and a miRNA-mRNA network was constructed. *KANSL1* and the 46 protein coding genes in the 22q11.2 deletion region were used as input in miWalk database to identify miRNAs that regulate these genes<sup>22</sup>. A miRNA-mRNA network was constructed using mRNA as the target node and the corresponding miRNA/s pairs found in the miWalk database as the source node. This analysis showed that *KANSL1* is regulated by miR-106b-5p, miR-148a-3p, miR-23b-3p, miR-17-5p, miR-149-5p and miR-130b-3p. These miRNAs also regulate *DGCR14*, *DGCR2*, *TXNRD2*, *MRPL40* and *CRKL* genes, which are included in the 22q11.2 deletion. Moreover, in a second miRNA-mRNA network analysis, previously reported modifier genes (*MINA*, *RREB1*, *KDM7A* and *JMJD1* and *SLC2A3*)<sup>10,12</sup> were also included. Interestingly, *SLC2A3* and *RREB1* genes, which are not included in 22q11.2 deletion, are regulated by the same miRNAs that regulate *KANSL1* (Fig. 3). To assess whether these results could occur by chance, the analyses was repeated using 50

CNV (hg19)	CNV size (bp)	Two tailed p value	OR (95% CI)	gain/loss	CNV Cases (n = 105)	CNV Controls (n = 109)	Distance to closest gene (bp)	Putative gene/s affected by CNV
chr13:84161129–84161251	122	0.013	14.3 (0.80–257.22)	loss	6	0	290092	<i>SLITRK1</i>
chr19:43646784–43691318	44534	0.013	14.30 (0.80–257.22)	loss	6	0	0	<i>PSG5</i>
chr16:55796389–55796594	205	0.014	2.92 (1.26–7.45)	loss	20	8	0	<i>CES</i>
chr1:16869363–16986851	117488	0.017	5.28 (1.32–38.34)	gain	10	2	0	<i>DQ585677</i> , <i>FLJ00313</i> , <i>MST1</i> , <i>NBPF1</i> , <i>NBPF10</i>
chr6:32570046–32574381	4335	0.017	7.88 (1.38–200.89)	loss	8	1	12484	<i>HLA-DRB1</i> , <i>HLA-DRB5</i>
chr3:148963395–148968121	4726	0.017	5.28 (1.32–38.34)	loss	10	2	23563	<i>CP</i>
chr17:44176984–44272000	95016	0.023	2.75 (1.17–7.03)	gain	19	8	0	<i>KANSL1</i>
chr15:22548025–22571758	23733	0.024	3.02 (1.18–8.88)	loss	16	6	74672	<i>IGH</i>
chr17:1619932–1680318	60386	0.028	11.99 (0.65–219.50)	loss	5	0	0	<i>SERPINF1</i> , <i>SERPINF2</i> , <i>WDR81</i>
chr8:7826922–7866162	39204	0.028	11.99 (0.65–219.50)	loss	5	0	0	<i>DEFB109</i>
chr11:5789601–5794708	5107	0.035	2.22 (1.09–4.66)	loss	26	14	0	<i>TRIM22</i> , <i>TRIM5</i>
chr13:57758270–57778380	20110	0.037	2.34 (1.09–5.31)	loss	22	11	13918	<i>PRR20</i>
chr3:89394592–89417171	22579	0.038	3.29 (1.09–12.48)	loss	12	4	0	<i>EPHA3</i>
chr11:18959245–18960115	870	0.042	2.42 (1.06–5.95)	gain	19	9	2696	<i>MRGPRX1</i>

**Table 1.** List of top-ranked CNVs associated with CHD in 22q11.2DS patients. A summary of top-ranked CNVs associated with CHD is shown.



**Figure 2.** Phenogram of associated CNVs. Human genes located or overlapped with identified CNVs are shown as green diamonds, phenotypes found in the 22q11.2DS patients are displayed as blue circles and the common ancestors (link between gene's phenotype annotation and patient's phenotypes) are displayed as orange squares. IAA = Interrupted Aortic Arch; TOF = Tetralogy of Fallot.

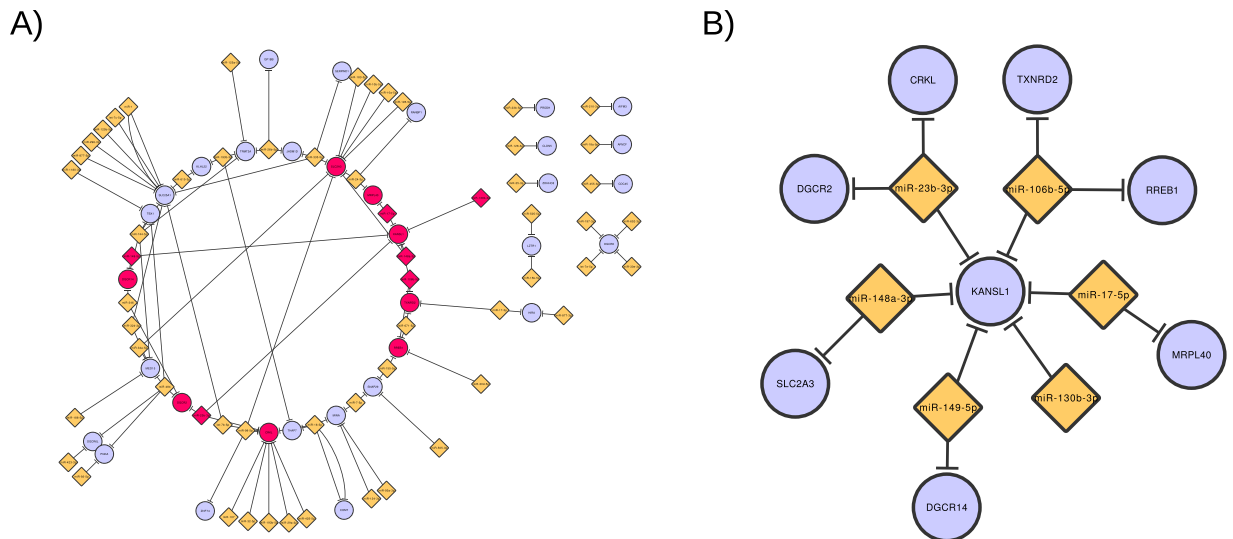
random genes in network constructions. No shared miRNA regulation between random genes and those in the deletion region was observed.

## Discussion

We report the results of a genome-wide analysis in search for cardiac modifier CNVs in regions outside the 22q11.2 deletion using high-density arrays in a case-control study design. Results showed that a microduplication that contains at least the first 3 exons of the *KANSL1* gene is statistically associated with the presence of CHD in 22q11.2DS patients and biologically related with modifier genes and genes located in 22q11.2 deletion region. Even though statistical association did not reach the multiple hypothesis testing correction, the integration of the top ranked p-values with interaction data gives a robust combination of the results in a biological context. Moreover, other studies have shown that even when a non-statistically significant p-value is obtained, the use of biological networks gives important information related to the condition<sup>12, 23–25</sup>.

Findings show that *KANSL1* duplication is a common CNV, seen in approximately 12% of a large group of Chilean individuals, indicative that by itself it is frequent and benign, but that in combination with 22q11.2 deletion, it increases the risk of CHD.

*KANSL1* is a protein-coding gene that belongs to a histone acetyltransferase (HAT) complex. This gene is encoded by *KAT8*, modifies the acetylation state of histone H4, mainly H4K16, and plays a role in the control of gene expression<sup>26, 27</sup>. Haploinsufficiency of this gene is implicated in Koolen-De Vries syndrome (KDVS, OMIM #610443)<sup>28</sup>. Interestingly, 39% of KDVS patients have CHD, mainly atrial septal defects or ventricular septal defects<sup>29</sup>. Analysis of the 17q21.31 region and characterization of KDVS patients without deletion of 17q21.31



**Figure 3.** miRNA-mRNA interactions networks. The yellow diamonds represent miRNAs and the light purple, mRNA targets. The edge connecting two nodes indicate regulation. **(A)** The red filled nodes correspond to miRNAs, *KANSL1* and genes of deletion co-regulated by the same miRNAs in the context of the whole network. **(B)** Sub-network showing only *KANSL1* interacting module. The graphs were generated in Cytoscape 3.2 and drawing using the Circle Layout.

show that SNVs located in *KANSL1* genes can be the cause of this syndrome<sup>28</sup>. Moreover, one of every four patients with point mutations in *KANSL1* has CHD, suggesting that this is a CHD-related gene, although further functional studies are needed to confirm this link<sup>28</sup>.

CNVs in *KANSL1* may be associated with CHD, with deletions causing KDVS and duplications acting as modifiers of the cardiovascular phenotype in 22q11.2DS. The literature reports show the effect of duplications in *KANSL1* gene in individuals with development delay, microcephaly and mild dysmorphic features<sup>30–32</sup>. These duplications span at least the complete *KANSL1* sequence and the microtubule-associated protein Tau gene (*MAPT*), which differs from the partial duplications found in 22q11.2 patients in this research. Mutations in *MAPT* gene are described in neurodegenerative disorders such as Alzheimer's disease; therefore, it is highly likely that the duplicated region of *MAPT* could be involved in the phenotype of 17q31.21 duplication patients. However, patients with 17q21.31 microduplication present milder phenotypes than patients with the microdeletion, suggesting that microduplications cases are probably unascertained<sup>30,31</sup>.

In the group of 40 patients with CMA, the duplication was frequent but not associated with CHD. However, in the group of patients with 22q11.2DS, the duplication was associated with CHD. This suggests that *KANSL1* microduplication, in combination with 22q11.2 deletion, increases the risk of developing CHD. The effects of the partial duplication in gene function and mechanisms of the interaction are to be elucidated; however, it is known that *KANSL1* is an important factor involved in the epigenetic modifications of histones. Variants in other genes with similar function affect conotruncal heart phenotypes in patients with 22q11.2DS<sup>13</sup>.

The relationship between histone modifier genes and heart disease has been previously described. It is widely known that epigenetic regulation of gene expression is mainly controlled by dynamic modification of chromatin structure<sup>33</sup>. This regulation mechanism is important during heart development as has been demonstrated in model organisms<sup>33,34</sup>. Genes that encode proteins involved in binding or modification of histones proteins have been implicated as disease genes causing heart defects<sup>35</sup>. These data support the idea of a link at a functional level between normal heart development, heart defects, and epigenetic regulation. Moreover, Kabuki (OMIM #147920) and CHARGE syndrome (OMIM #214800) are two examples of diseases that have CHD among other traits in which responsible genes are chromatin modifiers<sup>36,37</sup>.

In addition, the miRNA-mRNA network analysis revealed a common regulation between some genes in the 22q11.2 deletion region and *KANSL1*. Interestingly, other genes previously reported as modifiers of the CHD phenotype are regulated in the same module, showing a biological link between modifiers and genes located in 22q11.2 deletion region. This suggests that genes belonging to a common miRNA regulatory network could act as genetic modifiers, and variation in any one of these genes increases the risk of CHD. In this context, miRNA regulatory networks are an emergent field in 22q11.2DS research. The *DGCR8* gene, located in the 22q11.2 deletion, is involved in miRNA biogenesis since it is the cofactor for nuclease DROSHA, which processes pri-miRNAs fold into hairpin structures containing imperfectly base-paired stems. DROSHA is an important regulator of miRNA biogenesis, suggesting that miRNA networks could be deregulated during development. Some miRNAs, including miR-150, miR-194 and miR-185, have been reported as differentially expressed in individuals with 22q11.2DS compared to controls<sup>38</sup>. It was proposed that this is related to *DGCR8* haploinsufficiency. Importantly, the expression level of some miRNAs correlated with a set of phenotypes, including heart defects, brain measures, and thyroid abnormalities, suggesting that the deregulated miRNAs may contribute to these characteristics<sup>38</sup>. Results



in this study suggest that specific miRNAs regulate genes within the deletion as well as modifiers elsewhere in the genome, increasing the risk of CHD.

In summary, this study presents robust data showing that a duplication in the 17q31.21 region spanning the exons 1–3 of *KANSL1* gene is statistically associated with CHD in Chilean patients with 22q11.2DS. Moreover, *KANSL1* and genes located in 22q11.2 deletion share regulatory mechanisms mediated by miRNA. Future functional studies are needed in order to identify the exact mechanism of CHD mediated by *KANSL1*.

## Materials and Methods

The study and subject recruitment process were approved by the Ethics Committee of all the Institutional Review Boards at participating institutions (Facultad de Medicina, Clínica Alemana at the Universidad del Desarrollo, Hospital Padre Hurtado, Hospital Dr. Luis Calvo Mackenna, Hospital Dr. Roberto del Río, and Fundación Gantz). The study was carried out in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all patients and/or parents.

**Subject Cohorts.** Blood samples were obtained from 253 patients diagnosed with 22q11.2 microdeletion demonstrated previously by fluorescence *in situ* hybridization (FISH) in clinical cytogenetic laboratories (Cohort 1). Clinical information including echocardiogram reports was collected from medical records.

The CEL files of cohorts 2 and 3 were obtained from previous studies: cohort 2 corresponds to Chilean population controls without 22q11.2DS genotyped with Affymetrix SNP 6.0 arrays<sup>19</sup> and cohort 3 corresponds to Chilean patients without 22q11.2DS who had CMA with Affymetrix CytoScanHD arrays due to congenital anomalies or development disabilities<sup>20</sup> according to ASHG guidelines<sup>39</sup>.

**Genotyping with Affymetrix SNP 6.0 array.** Genomic DNA was purified from whole blood samples using AxyPrep blood genomic DNA kit (Axygen, Corning, NY, USA). Processing of genomic DNA and array hybridization was performed in accordance with manufacturer's recommendations. Affymetrix Human SNP 6.0 array (Affymetrix, Santa Clara, CA, USA) contains 1.8 million genetic markers and more than 946,000 non-polymorphic CNV probes. Fluorescence intensities were quantified by Affymetrix Array Scanner 3000 7G. Affymetrix Gene Chip Console (AGCC) software was used for data management and quality controls checks.

**CNV calling and filters.** Every sample that passed standard SNP QC procedures according to Affymetrix Genotyping Console was entered into the CNV analysis pipeline. These samples were divided into two groups according to cardiac phenotype: patients with CHD (CHD,  $n = 133$ ) and patients with normal cardiac anatomy (No CHD,  $n = 120$ ). Affymetrix Power Tools software was used to compute LRR and B allele frequency measures. CNV calls were done with PennCNV software standard procedure<sup>14</sup>, using the standard hg18 "all" PennCNV hidden Markov model, population frequency of B allele files and GC model correction for the Affymetrix SNP 6.0 arrays available from PennCNV web-page.

Samples that failed PennCNV quality control were excluded, beginning with samples with LRR standard deviation  $>0.45$ . Samples that failed QC after the wave adjustment procedure were also removed. PennCNV frequently generates a great number of CNV calls for samples with very low quality even if LRR standard deviation appears normal, so the samples that generated over 100 CNV calls were also removed. Given that some centromeric and telomeric regions are not well mapped, potentially resulting in CNV-calling errors in these regions, they were removed from the analysis by delimiting them to a 500 kb window. Finally, genomic regions encoding immunoglobulin genes have previously shown to be potential sites of false-positive PennCNV calls and were removed using genomics coordinates as suggested in the PennCNV software website<sup>14</sup>.

**Hierarchical sorting of CNV with bioinformatics approach.** In order to identify gene or genes located in CNV regions that could potentially act as genetic modifiers of heart anatomy, the PhenogramViz plugin<sup>17</sup> of CytoScape<sup>18</sup> was used. This plugin facilitates a phenotype guided interpretation of CNVs by using the integrated cross-species phenotype ontology Uberpheno. The results were visualized as gene-to-phenotype association in a network of nodes and edges called phenograms. The UCSC Genome Browser LiftOver tool was used to convert CNV coordinates to the GRCh37/hg19 build<sup>40</sup>. A phenogram was created using the coordinates of CNVs of lowest  $p$ -value ( $p < 0.05$ ) obtained with ParseCNV<sup>15</sup> and the HPO terms "Tetralogy of Fallot" (HPO:0001636) and "Interrupted aortic arch" (HPO:0011611). The resulting phenogram was visualized in CytoScape 3.2.1.

**CNV validation of *KANSL1* microduplication by MLPA.** Validation analysis was performed in samples from patients with 22q11.2DS using the SALSA MLPA kit P443 (MRC-Holland, Amsterdam, Netherlands), ABI-Prism 310 (Applied Biosystems, Foster City, CA) for capillary electrophoresis, and GeneMaker 2.0 software (SoftGenetics, State College, PA) to ascertain CNVs. Data was population-normalized, and probe ratios below 0.75 were considered as the threshold for deletion, while probe ratios above 1.25 were regarded as an indication of duplication. Validation was performed in 253 samples, with 7 showing results in cutoff limits; these were excluded from subsequent MLPA statistical analysis.

**Assessment of *KANSL1* microduplication frequency in non-22q11.2DS controls.** To determine the frequency of *KANSL1* microduplication in patients without 22q11.2DS, a second set of data from Affymetrix SNP 6.0 microarray originating from population controls (cohort two) was employed<sup>19</sup>. The cardiac phenotype information in these individuals is unavailable. The CNV filter and calling pipeline for this cohort of patients were the same as for 22q11.2DS cohort. The CNVs were identified, and the frequency of *KANSL1* microduplication was calculated. In addition, data from a third cohort<sup>20</sup> of patients with available cardiac phenotype information was processed using ChAS 2.0 software (Affymetrix, Santa Clara, CA). The association between CHD phenotype and the presence of *KANSL1* microduplication was calculated by Fisher's exact test in the corresponding cohorts.

**Biological interaction of *KANSL1* and genes located in 22q11.2 deletion.** To identify a link between *KANSL1* gene and genes located in the deleted region in 22q11.2DS, integrative bioinformatics approaches were employed. First, protein-protein interactions between *KANSL1* and proteins coded by genes located in the deletion region (3 Mb) were analyzed using the GeneMANIA tool, and only validated protein-protein interactions were used as edge connections<sup>21</sup>. A second approach assessed whether *KANSL1* and genes located in 22q11.2 deletion could be involved in common regulatory mechanisms, using miRWalk database<sup>22</sup>, which has information about miRNA-mRNA pairs that have been experimentally validated. Briefly, a search for all mRNA-miRNA pairs for all genes located in 22q11.2 deletion and *KANSL1* was done in the miRWalk webpage (<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/>). Genes previously described as modifiers, including *SLC2A3*, *MINA*, *RREB1* and *JMJD1* were used in this analysis<sup>11,13</sup>. To avoid finding regulatory interactions due to chance, the same procedure was repeated using 50 random genes and genes located in 22q11.2 deletion. The list of random genes was extracted from the HUGO database using an in-house script. These lists of miRNA-mRNA pairs were used to construct miRNA-target regulatory networks, which were visualized and analyzed with Cytoscape 3.2 as published<sup>41</sup>.

**Statistical Analysis.** A case/control association analysis was performed using ParseCNV, in which the main parameters were set by default using 1000 permutations. For CNV burden analysis, the number and size of deletions and duplications in CHD and no-CHD samples were counted. A Shapiro-Wilk test for normality of the data was performed, followed by a Wilcoxon rank sum test to compare CNV burden data between groups. The OR and CI of CNVs were calculated using the *epitools* package in R. All these analyses were done in R 3.2. software for statistical computing (Vienna, Austria).

## References

- McDonald-McGinn, D. M. *et al.* 22q11.2 deletion syndrome. *Nat. Rev. Dis. Prim.* **1**, 15071, doi:10.1038/nrdp.2015.71 (2015).
- Hacıhamdioğlu, B., Hacıhamdioğlu, D. & Delil, K. 22q11 deletion syndrome: current perspective. *Appl. Clin. Genet.* **8**, 123–32, doi:10.2147/TACG.S82105 (2015).
- Shaikh, T. H. *et al.* Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. *Hum. Mol. Genet.* **9**, 489–501, doi:10.1093/hmg/9.4.489 (2000).
- Emanuel, B. S. Molecular mechanisms and diagnosis of chromosome 22q11.2 rearrangements. *Dev. Disabil. Res. Rev.* **14**, 11–18, doi:10.1002/ddrr.3 (2008).
- Sandrin-Garcia, P. *et al.* Typical phenotypic spectrum of velocardiofacial syndrome occurs independently of deletion size in chromosome 22q11.2. *Mol. Cell. Biochem.* **303**, 9–17, doi:10.1007/s11010-007-9450-5 (2007).
- Aggarwal, V. S. & Morrow, B. E. Genetic modifiers of the physical malformations in velo-cardio-facial syndrome/DiGeorge syndrome. *Developmental Disabilities Research Reviews* **14**, 19–25, doi:10.1002/ddrr.4 (2008).
- Glessner, J. T. *et al.* Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ. Res.* **115**, 884–96, doi:10.1161/CIRCRESAHA.115.304458 (2014).
- McCarroll, S. A. & Altshuler, D. M. Copy-number variation and association studies of human disease. *Nat. Genet.* **39**, S37–42, doi:10.1038/ng2080 (2007).
- Jiang, Q., Ho, Y.-Y., Hao, L., Nichols Berrios, C. & Chakravarti, A. Copy number variants in candidate genes are genetic modifiers of Hirschsprung disease. *PLoS One* **6**, e21219, doi:10.1371/journal.pone.0021219 (2011).
- Shen, Y. *et al.* A significant effect of the TSPY1 copy number on spermatogenesis efficiency and the phenotypic expression of the *gr/gr* deletion. *Hum. Mol. Genet.* **22**, 1679–95, doi:10.1093/hmg/ddt004 (2013).
- Mlynarski, E. E. *et al.* Copy-Number Variation of the Glucose Transporter Gene SLC2A3 and Congenital Heart Defects in the 22q11.2 Deletion Syndrome. *Am. J. Hum. Genet.* **96**, 753–64, doi:10.1016/j.ajhg.2015.03.007 (2015).
- Mlynarski, E. E. *et al.* Rare copy number variants and congenital heart defects in the 22q11.2 deletion syndrome. *Hum. Genet.* **135**, 273–85, doi:10.1007/s00439-015-1623-9 (2016).
- Guo, T. *et al.* Histone Modifier Genes Alter Conotruncal Heart Phenotypes in 22q11.2 Deletion Syndrome. *Am. J. Hum. Genet.* **97**, 869–77, doi:10.1016/j.ajhg.2015.10.013 (2015).
- Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–74, doi:10.1101/gr.6861907 (2007).
- Glessner, J. T., Li, J. & Hakonarson, H. ParseCNV integrative copy number variation association software with quality tracking. *Nucleic Acids Res.* **41**, e64–e64, doi:10.1093/nar/gks1346 (2013).
- Michaelovsky, E. *et al.* Genotype-phenotype correlation in 22q11.2 deletion syndrome. *BMC Med. Genet.* **13**, 122, doi:10.1186/1471-2350-13-122 (2012).
- Köhler, S. *et al.* Clinical interpretation of CNVs with cross-species phenotype data. *J. Med. Genet.* **51**, 1–7, doi:10.1136/jmedgenet-2014-102633 (2014).
- Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–504, doi:10.1101/gr.1239303 (2003).
- Eyheramendy, S., Martinez, F. I., Manevy, F., Vial, C. & Repetto, G. M. Genetic structure characterization of Chileans reflects historical immigration patterns. *Nat. Commun.* **6**, 6472, doi:10.1038/ncomms7472 (2015).
- Lay-Son, G. *et al.* Chromosomal microarrays testing in children with developmental disabilities and congenital anomalies. *J. Pediatr. (Rio. J.)* **91**, 189–95, doi:10.1016/j.jpeds.2014.07.003 (2015).
- Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* **38**, W214–20, doi:10.1093/nar/gkq537 (2010).
- Dweep, H., Sticht, C., Pandey, P. & Gretz, N. miRWalk – Database: Prediction of possible miRNA binding sites by ‘walking’ the genes of three genomes. *J. Biomed. Inform.* **44**, 839–847, doi:10.1016/j.jbi.2011.05.002 (2011).
- Jia, P., Zheng, S., Long, J., Zheng, W. & Zhao, Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* **27**, 95–102, doi:10.1093/bioinformatics/btq615 (2011).
- Pan, W. Network-based model weighting to detect multiple loci influencing complex diseases. *Hum. Genet.* **124**, 225–34, doi:10.1007/s00439-008-0545-1 (2008).
- Lage, K. *et al.* A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat. Biotechnol.* **25**, 309–16, doi:10.1038/nbt1295 (2007).
- Dias, J. *et al.* Structural analysis of the *KANSL1*/WDR5/*KANSL2* complex reveals that WDR5 is required for efficient assembly and chromatin targeting of the NSL complex. *Genes Dev.* **28**, 929–942, doi:10.1101/gad.240200.114 (2014).
- Cai, Y. *et al.* Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J. Biol. Chem.* **285**, 4268–72, doi:10.1074/jbc.C109.087981 (2010).

28. Koolen, D. A. *et al.* Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat. Genet.* **44**, 639–641, doi:10.1038/ng.2262 (2012).
29. Koolen, D. A. *et al.* The Koolen-de Vries syndrome: a phenotypic comparison of patients with a 17q21.31 microdeletion versus a KANSL1 sequence variant. *Eur. J. Hum. Genet.* 1–8, doi:10.1038/ejhg.2015.178 (2015).
30. Mc Cormack, A., Taylor, J., Te Weehi, L., Love, D. R. & George, A. M. A case of 17q21.31 microduplication and 7q31.33 microdeletion, associated with developmental delay, microcephaly, and mild dysmorphic features. *Case Rep. Genet.* **2014**, 658570–6, doi:10.1155/2014/658570 (2014).
31. Natacci, F. *et al.* Chromosome 17q21.31 duplication syndrome: Description of a new familiar case and further delineation of the clinical spectrum. *Eur. J. Paediatr. Neurol.* **20**, 183–7, doi:10.1016/j.ejpn.2015.09.010 (2015).
32. Kitsiou-Tzeli, S. *et al.* Microdeletion and microduplication 17q21.31 plus an additional CNV, in patients with intellectual disability, identified by array-CGH. *Gene* **492**, 319–324, doi:10.1016/j.gene.2011.10.023 (2012).
33. Mathiyalagan, P., Keating, S. T., Du, X.-J. & El-Osta, A. Chromatin modifications remodel cardiac gene expression. *Cardiovasc. Res.* **103**, 7–16, doi:10.1093/cvr/cvu122 (2014).
34. Martinez, S. R., Gay, M. S. & Zhang, L. Epigenetic mechanisms in heart development and disease. *Drug Discov. Today* **20**, 799–811, doi:10.1016/j.drudis.2014.12.018 (2015).
35. Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–3, doi:10.1038/nature12141 (2013).
36. Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat. Genet.* **42**, 790–3, doi:10.1038/ng.646 (2010).
37. Vissers, L. E. L. M. *et al.* Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat. Genet.* **36**, 955–7, doi:10.1038/ng1407 (2004).
38. Sellier, C. *et al.* Decreased DGCR8 expression and miRNA dysregulation in individuals with 22q11.2 deletion syndrome. *PLoS One* **9**, e103884, doi:10.1371/journal.pone.0103884 (2014).
39. Miller, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am. J. Hum. Genet.* **86**, 749–64, doi:10.1016/j.ajhg.2010.04.006 (2010).
40. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598, doi:10.1093/nar/gkj144 (2006).
41. León, L. E. & Calligaris, S. D. Visualization and Analysis of miRNA-targets interactions networks in. *Methods in molecular biology.* **1509**, 209–220, doi:10.1007/978-1-4939-6524-3\_19 (2017).

## Acknowledgements

This work was supported by Fondecyt grants 1100131 and 1130392, Comisión Nacional de Ciencia y Tecnología (CONICYT-Chile). The authors would like to thank Dr. Anne Bassett (University of Toronto) and Dr. Anne Bliss (University of Colorado-Boulder) for critical reading of the manuscript.

## Author Contributions

Conceived and designed the analysis: L.E.L., K.E., C.V. and G.M.R. Sample and clinical data collection: P.A., M.P., G.L.S., M.M. and G.M.R. Performed the analysis: L.E.L., F.B., K.E. and C.V. Analyzed and interpreted the data: L.E.L., F.B., G.M.R. Wrote the manuscript: L.E.L. and G.M.R.

## Additional Information

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017