Title

Complete Sequence of the 22q11.2 Allele in 1,053 Subjects with 22q11.2 Deletion Syndrome Reveals Modifiers of Conotruncal Heart Defects

Journal

American Journal of Human Genetics, 106(1)

ISSN

0002-9297

Authors

Zhao, Yingjie
Diacou, Alexander
Johnston, H Richard
et al.

Publication Date

2020

DOI

10.1016/j.ajhg.2019.11.010

Peer reviewed

ARTICLE

Complete Sequence of the 22q11.2 Allele in 1,053 Subjects with 22q11.2 Deletion Syndrome Reveals Modifiers of Conotruncal Heart Defects

Yingjie Zhao,[1] Alexander Diacou,[1] H. Richard Johnston,[2] Fadi I. Musfee,[3] Donna M. McDonald-McGinn,[4,5] Daniel McGinn,[4,5] T. Blaine Crowley,[4,5] Gabriela M. Repetto,[6] Ann Swillen,[7] Jeroen Breckpot,[7] Joris R. Vermeesch,[7] Wendy R. Kates,[8,9] M. Cristina Digilio,[10] Marta Unolt,[10,11] Bruno Marino,[11] Maria Pontillo,[12] Marco Armando,[12,13] Fabio Di Fabio,[11] Stefano Vicari,[12,14] Marianne van den Bree,[15] Hayley Moss,[15] Michael J. Owen,[15] Kieran C. Murphy,[16] Clodagh M. Murphy,[17,18] Declan Murphy,[17,18] Kelly Schoch,[19] Vandana Shashi,[19] Flora Tassone,[20]

(Author list continued on next page)

The 22q11.2 deletion syndrome (22q11.2DS) results from non-allelic homologous recombination between low-copy repeats termed LCR22. About 60%–70% of individuals with the typical 3 megabase (Mb) deletion from LCR22A-D have congenital heart disease, mostly of the conotruncal type (CTD), whereas others have normal cardiac anatomy. In this study, we tested whether variants in the hemizygous LCR22A-D region are associated with risk for CTDs on the basis of the sequence of the 22q11.2 region from 1,053 22q11.2DS individuals. We found a significant association (FDR $p < 0.05$) of the CTD subset with 62 common variants in a single linkage disequilibrium (LD) block in a 350 kb interval harboring *CRKL*. A total of 45 of the 62 variants were associated with increased risk for CTDs (odds ratio [OR] ranges: 1.64–4.75). Associations of four variants were replicated in a meta-analysis of three genome-wide association studies of CTDs in affected individuals without 22q11.2DS. One of the replicated variants, rs178252, is located in an open chromatin region and resides in the double-elite enhancer, GH22J020947, that is predicted to regulate *CRKL* (CRK-like proto-oncogene, cytoplasmic adaptor) expression. Approximately 23% of patients with nested LCR22C-D deletions have CTDs, and inactivation of *Crkl* in mice causes CTDs, thus implicating this gene as a modifier. Rs178252 and rs6004160 are expression quantitative trait loci (eQTLs) of *CRKL*. Furthermore, set-based tests identified an enhancer that is predicted to target *CRKL* and is significantly associated with CTD risk (GH22J020946, sequence kernal association test (SKAT) $p = 7.21 \times 10^{-5}$) in the 22q11.2DS cohort. These findings suggest that variance in CTD penetrance in the 22q11.2DS population can be explained in part by variants affecting *CRKL* expression.

Introduction

The vast majority of individuals with 22q11.2 deletion syndrome (22q11.2DS [MIM: 192430]) have a 3 megabase (Mb) hemizygous deletion of chromosome 22q11.2.[1] Occuring in 1/4,000 live births[2,3] and 1/1,000 fetuses,[4,5] this syndrome is the most frequent chromosomal microdeletion disorder The 22q11.2DS results from *de novo* non-allelic homologous recombination events between four low-copy repeats (LCRs) termed LCR22A, B, C, and D.[6,7]

[1]Department of Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, USA; [2]Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA; [3]Department of Epidemiology, Human Genetics and Environmental Sciences, UTHealth School of Public Health, Houston, Texas 77225, USA; [4]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia 19104, USA; [5]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia 19104, USA; [6]Center for Genetics and Genomics, Facultad de Medicina Clinica Alemana-Universidad del Desarrollo, Santiago 7710162, Chile; [7]Center for Human Genetics, University of Leuven (KU Leuven), Leuven 3000, Belgium; [8]Department of Psychiatry and Behavioral Sciences, SUNY Upstate Medical University, Syracuse, NY 13202, USA; [9]Program in Neuroscience, SUNY Upstate Medical University, Syracuse, NY 13202, USA; [10]Department of Medical Genetics, Bambino Gesù Hospital, Rome 00165, Italy; [11]Department of Pediatrics, Gynecology, and Obstetrics, La Sapienza University of Rome, Rome 00185, Italy; [12]Department of Neuroscience, Bambino Gesù Hospital, Rome 00165, Italy; [13]Developmental Imaging and Psychopathology Lab, University of Geneva, Geneva 1211, Switzerland; [14]Department of Psychiatry, Catholic University, Rome 00153, Italy; [15]Medical Research Council Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Wales CF24 4HQ, UK; [16]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin 505095, Ireland; [17]Department of Forensic and Neurodevelopmental Sciences, King's College London, Institute of Psychiatry, Psychology, and Neuroscience, London SE5 8AF, UK; [18]Behavioural and Developmental Psychiatry Clinical Academic Group, Behavioural Genetics Clinic, National Adult Autism and ADHD Service, South London and Maudsley Foundation National Health Service Trust, London SE5 8AZ, UK; [19]Department of Pediatrics, Duke University, Durham, NC 27710, USA; [20]Department of Psychiatry and Behavioral Sciences, MIND Institute, University of California, Davis, CA 95817, USA; [21]The *Virtual Center* for Velo-Cardio-Facial Syndrome, Syracuse, NY 13206, USA; [22]School of Psychology, University of Newcastle, Newcastle 2258, Australia; [23]Department of Medical Genetics, Aix-Marseille University, Marseille 13284, France; [24]Genomics of Health and Unit of Molecular Diagnosis and Clinical Genetics, Son Espases University Hospital, Balearic Islands Health Research Institute, Palma de Mallorca 07120, Spain; [25]Institute of Medical and Molecular Genetics, University Hospital La Paz, Madrid 28046, Spain; [26]See Table S1; [27]Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles, Los Angeles, CA 90095, USA; [28]Department of Psychiatry and Psychology, Maastricht University, Maastricht, 6200 MD, the Netherlands; [29]Program in Genetics and Genome Biology, Research Institute, Toronto, Ontario, Canada; [30]Department of Psychiatry, The Hospital for Sick Children, Toronto, Ontario, Canada; [31]Department of Psychiatry, University of Toronto, Toronto, Ontario, M5S 1A1, Canada; [32]Department of Psychiatry, University Medical Center Utrecht Brain Center, Utrecht, 3584 CG, the Netherlands; [33]The Child Psychiatry Unit, Edmond and Lily Sapfra Children's Hospital, Sackler Faculty of Medicine, Tel Aviv University and Sheba Medical Center, Tel Aviv, 52621, Israel; [34]Department of

(Affiliations continued on next page)

Tony J. Simon,[20] Robert J. Shprintzen,[21] Linda Campbell,[22] Nicole Philip,[23] Damian Heine-Suñer,[24] Sixto García-Miñaúr,[25] Luis Fernández,[25] International 22q11.2 Brain and Behavior Consortium,[26] Carrie E. Bearden,[27] Claudia Vingerhoets,[28] Therese van Amelsvoort,[28] Stephan Eliez,[13] Maude Schneider,[13] Jacob A.S. Vorstman,[29,30,31,32] Doron Gothelf,[33] Elaine Zackai,[4,5] A.J. Agopian,[3] Raquel E. Gur,[34,35] Anne S. Bassett,[36,37,38] Beverly S. Emanuel,[4,5] Elizabeth Goldmuntz,[39,40] Laura E. Mitchell,[3] Tao Wang,[41] and Bernice E. Morrow,[1,*]

More than 85% of affected individuals carry a 3 Mb hemizygous deletion between LCR22A and LCR22D.[8] However, nested proximal (LCR22A– LCR22B, 1.5 Mb, 5%; LCR22A–LCR22C, 2 Mb, 2%;[6,7]) and distal (LCR22B–LCR22D, 1.5 Mb, 4%; LCR22C–LCR22D, 0.7 Mb, 1%) deletions are present in some individuals with 22q11.2DS.[1] In patients with the LCR22A–LCR22D and proximal nested LCR22A–LCR22B and LCR22A–LCR22C deletions the prevalence of congenital heart disease (CHD) is approximately 65%.[9,10] A somewhat lower prevalence (~32%) is observed in individuals with distal nested deletions.[11–14] Hence, both nested proximal and distal hemizygous deletions are associated with the occurrence of CHD.

Most 22q11.2DS patients with CHD have conotruncal heart defects (CTDs [MIM: 217095][15]), affecting the development of the cardiac outflow tract, including the aortic arch. Such defects that occur in 22q11.2DS patients include tetralogy of Fallot (TOF [MIM: 187500]), persistent truncus arteriosus (PTA), interrupted aortic arch type B (IAAB), right-sided aortic arch (RAA), and abnormal branching of the subclavian arteries. Some have isolated atrial septal defects (ASD), ventricular septal defects (VSD), and rarely, other cardiac malformations. Among the known coding genes in the LCR22A–LCR22B region, *TBX1* (*T-box 1* [MIM: 602054]), which encodes a T-box transcription factor,[16,17] is the strongest candidate gene for CTDs, as first suggested by gene inactivation studies in mouse models.[16–18] Inactivation of one allele of *Tbx1* resulted in mild aortic-arch anomalies, whereas inactivation of both alleles resulted in a PTA and perinatal lethality.[16–18] Furthermore, missense variants in *TBX1* have been found in individuals without a known genetic condition; these individuals partially phenocopied those with 22q11.2DS, implicating *TBX1* as a human CTD gene.[19–24] There is another gene, *CRKL* (CRK-like protooncogene adaptor protein [MIM: 602007]), that has been considered as a candidate. *CRKL*, mapping to the LCR22C-D region, is also of strong interest because inactivation of both alleles in mouse models results in CTDs with late gestational lethality.[25]Interestingly, a genetic interaction was observed between *Tbx1* and *Crkl* in mouse models, suggesting that they might participate in the same functional pathway during embryogenesis.[26]

In contrast to CHD's prevalence of about 0.5%–1% in the general population,[27] the dramatically elevated CHD risk in the 22q11.2DS population is attributed largely to the presence of the hemizygous deletion. Phenotypic variability, however, cannot be fully explained by the presence of the 22q11.2 deletion or deletion size and is most likely due to the existence of additional genetic and/or environmental modifiers. Identification of modifiers can provide insight into the biological mechanism of heart development and disease. Although some insights into the genetic architecture of CHD in 22q11.2DS have been gained through array genotyping and whole-exome sequencing efforts, whole-genome sequencing (WGS) methods in large cohorts are needed.[28–33] The remaining 22q11.2 allele is particularly vulnerable to second-hit variants because only one functional copy of genes is present. To test the hypothesis that common and/or rare single-nucleotide variants (SNVs; including small indels) on the remaining allele might be associated with CHD, we used WGS data from 1,053 22q11.2DS subjects, all with the same typical 3 Mb LCR22A–LCR22D deletion. We performed a case-control association study involving 22q11.2-affected individualswith CHD or subtypes within CHD, such as CTDs, and control individuals with 22q11.2DS but who also have a normal heart and/or aortic arch. Furthermore, to determine whether associations identified in the 22q11.2DS cohort were also observed in CTD cases from the general population, we analyzed existing genome-wide association data from a meta-analysis of CTDs in cases without a 22q11.2 deletion.[34]

Methods

Study Population

22q11.2DS Cohort

Recruitment of the study subjects has been previously described.[31,35] In brief, subjects with a known 22q11.2 deletion, existing DNA samples, and approval by institutional research ethics boards (Albert Einstein College of Medicine; Committee of

Psychiatry, Perelman School of Medicine of the University of Pennsylvania Philadelphia, PA 19104, USA; [35]Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; [36]Dalglish Family 22q Clinic, Clinical Genetics Research Program, Toronto M5T 1L8, Ontario Canada; [37]Toronto General Hospital, Centre for Addiction and Mental Health, Toronto M5T 1L8, Ontario, Canada; [38]Department of Psychiatry, University of Toronto, Toronto M5T 1L8, Ontario, Canada; [39]Division of Cardiology, Children's Hospital of Philadelphia Philadelphia, PA 19104, USA; [40]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [41]Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA
*Correspondence: bernice.morrow@einsteinmed.org
https://doi.org/10.1016/j.ajhg.2019.11.010.

Clinical Investigation; CCI#1999-201) were recruited in part from the International 22q11.2 Brain and Behavior Consortium[36] (Table S1). A total of 1,595 samples had a clinical diagnosis of 22q11.2DS and carried a laboratory-confirmed 22q11.2 deletion.

Congenital-Heart-Disease Phenotypes in the Study Population

We obtained cardiac phenotype information from cardiology records, including echocardiography reports, as previously described.[35] Individuals with missing cardiac records were excluded from these analyses. Individuals with the LCR22A-D deletion and any intracardiac or aortic arch defect were considered to be CHD-affected individuals. Individuals with no heart or aortic arch defect, except for those with only a patent foramen ovale or VSD and/or ASD that spontaneously closed in infancy and/or a bicuspid aortic valve, were considered to be controls. For the CTD subset, any of the following cardiac defects were considered to be CTD-affected individuals in the present study: TOF, PTA, IAAB, RAA, or abnormal origin of the right or left subclavian artery. The difference between CHD- and CTD-afected individuals was that CTD-affected individuals did not have isolated VSD or ASD, but both CHD-affected and CTD-affected individuals had LCR22A-D deletions. CTD can be separated into two different groups based upon differences in embryological origin. These groups are characterized by (1) cardiac outflow tract (OFT) defects that include TOF, PTA, PS (pulmonic stenosis), and/or PA (pulmonary atresia) and (2) aortic-arch defects or defects that affect arterial branching from the aortic arch; such defects include RAA, IAAB, or other aortic-arch defects, such as abnormal origin of the right subclavian artery. We also performed other sub-phenotype comparisons as described.

Complete Sequencing of the 22q11.2 Region and Quality-Control Measures

WGS with a median depth of 39-fold was performed on 1,595 subjects as part of the International 22q11.2 Brain and Behavior Consortium.[36] In brief, samples were sequenced with the Illumina HiSeq X Ten for the first 100 samples and the Illumina HiSeq 2500 platform for all other samples at Hudson Alpha. Sequence reads were mapped to genome build hg38 (December 2013; GRCh38/hg38) with PEMapper (90% stringency for 2 × 100 bp reads and 95% stringency for 2 × 150 bp reads;[37]). Deletion sizes were confirmed by the coverage at the 22q11.2 region. Variants on the remaining 22q11.2 allele (LCR22A–LCR22D region; chr22: 18115819–21432004, hg38) were called by PECaller in haploid mode. Variants were called if ≥90% of the variants at the site had a posterior probability ≥95%. Variants were removed if Hardy-Weinberg equilibrium (HWE) p value were $< 1.0 \times 10^{-5}$. Exclusion was based upon the quality-control (QC) results of the dataset from the genome-wide diploid variants. Samples from relatives, those in which sex did not match that revealed by genetic assays, and samples with poor-quality sequence were removed. For all samples with the LCR22A–LCR22D deletion, variants with a genotype call rate of <0.95 and monomorphic variants were removed. Variants within LCR22 regions were removed because of their repetitive nature.

Principal-Component Analysis

Principal-component analysis (PCA) was conducted with PLINK 1.9 beta on the basis of the dataset of genome-wide diploid variants as well as Hapmap 3 r3 (International HapMap project phase III release 3) data. First, shared variants in this dataset and the Hapmap 3 r3 dataset were extracted and combined into one dataset. Of note, we removed variants with A>T, T>A, G>C, and C>G allele types to avoid any potential strand-flip issues. Second, variants with minor-allele frequency <0.05 and variants in the sex chromosome were excluded. After this, we pruned autosomal common variants by using the –indep function to ensure only independent variants were used for PCA. Lastly, PCA was conducted with the –pca function. European Caucasian ancestry of the subjects was determined by the Multidimensional Outlier Detection method as implemented in SVS Golden Helix software. First, a median centroid vector was calculated as [median (column1), median (column2), median (column3)] on the basis of the top three principal components (PCs) for all the samples plus Hapmap CEU (Utah residents with ancestry from northern and western Europe) and TSI (Tuscans in Italy) samples (combined, referred to as Caucasian). A distance score was then calculated for each sample as follows:

$$threshold = \sqrt{\sum_{n=1}^{N} Q3_n^2} + M * \sqrt{\sum_{n=1}^{N} IQR_n^2}$$

The outlier threshold was calculated as follows:

$$dist_{sample} = \sqrt{\sum_{n=1}^{N} \left(value_{sample,n} - median_n\right)^2}$$

Where Q3 and IQR are the third quartile and inner quartile range of each sample (1 … N), respectively, and M is a user-specified multiplier; in this study, 2 was adopted. Outliers of Caucasian samples were examined in the scatterplot of PC1 versus PC2; samples that clustered with the HapMap Gujarati Indians in Houston, Texas (GIH) population or the Mexican ancestry in Los Angeles, California (MEX) population were grouped as Hispanics. Populations dispersed toward Yoruba in Ibadan, Nigeria (YRI) were grouped as African-admixed populations. We also conducted PCA in the three subpopulations to obtain the top several PCs that could serve as covariants so that we could adjust for possible population stratification in the stratified analyses.

SNV-Based Analyses

Logistic regression analyses for common variants were conducted in all 1,053 samples as well as in the Caucasian, African-admixed, and Hispanic subsets; the sex and corresponding number of PCs were adjusted for CHD, CTD, TOF, and TOF-PTA-IAAB risk. We employed the false-discovery rate (FDR) to correct for multiple testing issues. For rare variants, we conducted the Fisher's exact test in the Caucasian population. Because most genome-wide association studies (GWASs) adopted a suggestive significant threshold at 1.0×10^{-5} for 2.0×10^6 to 1.0×10^6 variants, we set the suggestive significant threshold for this study of a few thousand variants at $p = 1.0 \times 10^{-3}$ for both common and rare variants. PLINK 1.90 was used for SNV-based analyses.[38] For the top variants that were located in the LCR22C–LCR22D region and showed evidence of association with CTD risk in the 22q11.2DS cohort, we analyzed existing data from a meta-analysis of three published GWASs of CTDs in individuals without a 22q11.2 deletion.[34]

Functional Annotation Filtering of Rare Loss-of-Function, High-Confidence-Deleterious Missense, and Splicing Variants

Variants that passed QC were annotated for possible biological function via Bystro,[39] snpEff,[40] and dbNSFP.[41] We adopted the definition of loss-of-function (LoF) variants as previously described.[42] n brief, LoF variants included any predicted

indel-frameshift, stop-gain, splice-donor, splice-acceptor, stop-loss, start-loss, and low-frequency variants, typically with an alternate allele frequency (AAF) less than 0.001 in gnomAD (the Genome Aggregation Database). These would be highly damaging variants. Only consistent Bystro and snpEff predictions for LoF variants were included. As an ensemble annotation tool, Bystro also has additional annotations for possible function; these additional annotations include phastCons and phyloP, combined annotation -dependent depletion score (CADD),[43] and AAF of gnomAD.[39] High-confidence-damaging missense (D-Mis) variants are defined by having been predicted to be "damaging" or "deleterious" or by having passed the suggested threshold for scores by at least half of the 29 algorithms compiled in dbNSFP (database for functional prediction for non-synonymous variants) for SNVs (such algorithms include SIFT,[44] SIFT4G,[45] PolyPhen2-HDIV,[46] PolyPhen2-HVAR,[46] LRT,[47] MutationTaster2,[48] and others[41]). The probability that a particular SNV affected splicing was annotated by ada and random forest scores by dbscSNV (database for splicing consensus variants), which is a companion database in dbNSFP. High-confidence-damaging splicing variants (D-splicing variants) included in this study were those that passed the suggested threshold of both algorithms (0.6 of the 0–1 range). Variants in evolutionarily conserved regions were defined as those with both phastCons and phyloP values in the top quantile. Variants that mapped to multiple transcripts and therefore had multiple annotations required prioritization of the RefSeq site type, as follows: exon > UTR > intron > intergenic. These variants also required the following prioritization of the RefSeq exonic allele function: indel-frameshift = start loss = stop gain = stop loss > indel-non-Frameshift > nonsynonymous > synonymous.

LINSIGHT measures potential non-coding variants for which primate evolution suggests effects on fitness.[49] The LINSIGHT score of the top associated variants in the LCR22C–LCR22D region were downloaded from the UCSC Genome Browser. Expression quantitative trait loci (eQTLs) of *CRKL* were identified in the GTEx database portal.[50] Variants were evaluated in ATAC-seq data of open chromatin regions during differentiation from human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) to early cardiomyocytes (GSE85330[51]).

Enrichment Analyses of Rare Variants

On the basis of variant annotation results, rare variants (alternative allele frequency, AAF < 0.01 in gnomAD) that fell into 14 categories were extracted. Variant categories included all LoF variants, those with CADD score[52] > 30, those with CADD score > 15, missense variants with CADD > 15, all missense variants, synonymous variants with CADD > 15, all synonymous variants, all exonic variants, all variants in UTR regions, all intronic variants, all variants in promoter regions (2 kb both upstream and downstream of transcriptional start site, TSS), variants in double-elite enhancers,[53] variants in conserved regions, and variants in non-coding RNAs (ncRNAs). Double-elite enhancers are defined as those that have two or more sources of functional genomic evidence for being a regulatory element (e.g., chromatin conformation assays), have associations with a gene target, and are supported by two or more functional genomic methods (e.g., eQTLs).[53] For these 14 categories of rare variants, we compared the average number that mapped to each category in affected individuals to the average number in controls. The significance of enrichment of any of the categories in CTD-affected individuals versus controls was accessed by 10,000 label-swapping permutations.

Set-Based Analyses

Set-based analyses were conducted for common and rare variants. Sets in the LCR22A–LCR22D region included RefSeq genes between the TSS and transcriptional termination site (TTS) as well as genes both 2 kb upstream and 2 kb downstream, promoter regions (both 2 kb upstream and 2 kb downstream of the TSS), and the double-elite set of curated high-confidence enhancers in the GeneHancer database.[53] All detailed information about the sets were downloaded from the UCSC Genome Browser (assembly hg38). The 3 Mb region contains 72 genes, including known or predicted coding as well as non-coding genes, and therefore 72 promoters, as well as 96 double-elite enhancers. Therefore, the Bonferroni multiple-correction threshold for the set-based test was set at $p = 2.1 \times 10^{-4}$. The suggestive significant threshold was set at $p = 1.0 \times 10^{-3}$. We used the burden test[35] to evaluate whether there were any rare variants were enriched in any set, and we tested whether the effects are in the same direction. We used SKAT to test for common variants and considered the situation where a large fraction of the variants in a region are non-causal or the effects of causal variants were in different directions.[54] The set-based tests were implemented in the SKAT R package.[55]

Results

22q11.2DS Cohort and Study Design

The study cohort consisted of 1,053 individuals who had 22q11.2DS and WGS data and who had a 3 Mb LCR22A–LCR22D deletion (Figure 1A and Table S2). A total of 14,158 SNVs within the LCR22A–LCR22 region from WGS passed quality-control measures and were used in the downstream analyses (Table S3). Demographic characteristics of the study cohort and frequency of various cardiac defects among 1,053 subjects are listed in Table 1. The conotruncal region of the heart contains the aorta, pulmonary trunk, and arterial branches and is shown in Figure 1B. Among the 1,053 individuals with 22q11.2DS, 584 (55.5%) had a diagnosis of CHD, 424 (40.3%) had a CTD, 105 (10.0%) had an isolated VSD, and 55 (5.2%) had an isolated ASD (Table 1; Figure 1C). The most common phenotype among those with a CTD was TOF (n = 194; 18.4%; Table 1; Figure 1C). The remaining 469 (44.5%) individuals with the LCR22A–LCR22D deletion that had a normal heart and aortic arch and were designated as controls (Table 1; Figure 1C).

Ancestry of the 22q11.2DS cohort determined by PCA is shown in Figure 2A and listed in Table 1. A total of 790 of the 1,053 subjects were of European descent, and the rest were of African-admixed (n = 161) or Hispanic (n = 102) descent, as indicated (Figure 2A; Table 1). As shown in Figure S1, the top five (n = 1,053), four (n = 790 CEU), four (n = 104 Hispanics), and five (n = 161 African-admixed) PCs accounted for the majority of the population variance and, hence, were used as covariates

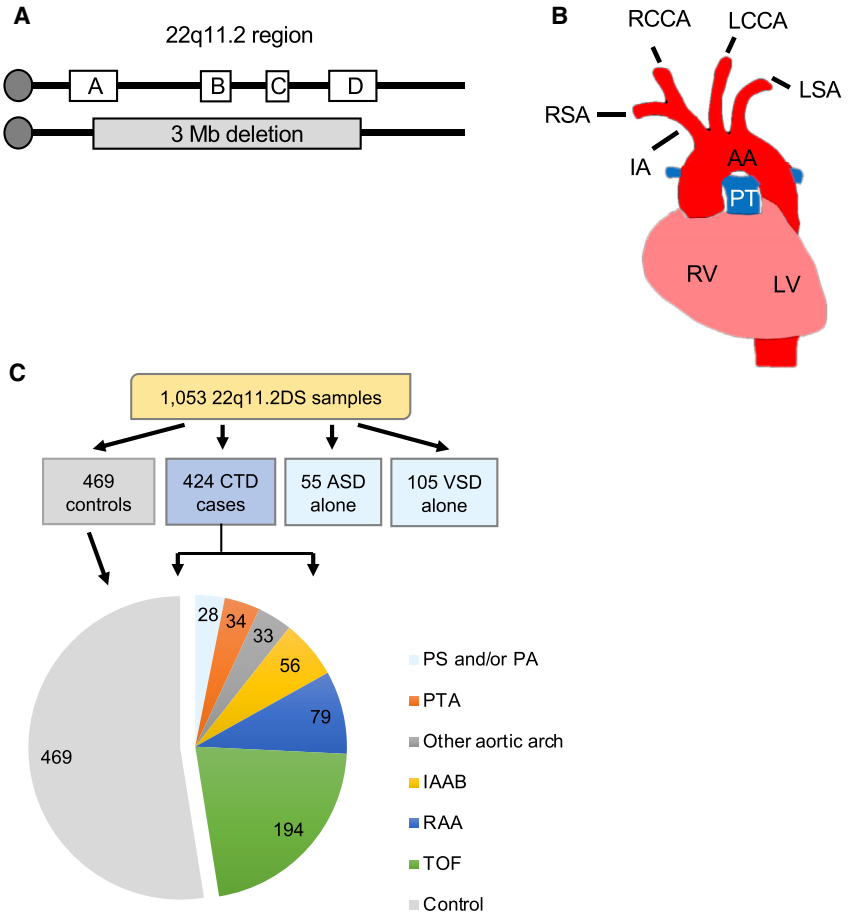

Figure 1. The 22q11.2 Deletion Syndrome Cohort

(A) An illustrative map of the 22q11.2 region depicts the locations of LCR22A, LCR22B, LCR22C, and LCR22D (box) and the 3 Mb deleted region (gray box) downstream of the centromere (dark gray circle).

(B) Heart in mammals indicating the aortic arch, pulmonary trunk (PT), and branches of the aortic arch. RV, right ventricle; LV, left ventricle. RCCA and LCCA, right and left common carotid artery; RSA and LSA, right and left subclavian artery; IA, innominate artery.

(C) Pie chart of cardiac and aortic-arch phenotypes, including 469 with a normal heart and aortic arch (gray), and phenotypic breakdown of 424 individuals comprising the cohort with CTDs. Among the 424, 194 had tetralogy of Fallot (TOF; green), 79 had a right-sided aortic arch (RAA; dark blue), 56 had interrupted aortic arch type B (IAAB; yellow), 34 had a persistent truncus arteriosus (PTA; dark gray), 28 had pulmonary stenosis or pulmonic atresia (PS/PA; light blue), and 33 had other aortic-arch defects such as an abnormal right or left subclavian artery in the absence of other cardiac or aortic arch anomalies.

for logistic-regression analyses of common variants among all subjects and stratified analyses for the three subpopulations.

The frequency distribution of the 14,158 variants that passed quality control is presented in Figure 2B. There were 9,821 rare variants with an AAF < 0.01 and 4,337 low-frequency and common variants. The low-frequency variants were grouped into common variants for all analyses, and for simplification, they are referred to herein as one group named common variants throughout this study.

The study design is presented in Figure 3. In brief, haploid variants in the 22q11.2 region were classified into common variants, rare variants, and the most damaging category of rare variants, including LoF and deleterious missense variants and splicing variants. Then different analytical strategies, which will be described in detail in the next sections, were applied to the three groups of variant categories.

SNV-Based Common-Variant Analyses

We performed an association study of common variants by comparing 584 CHD-affected individuals to 469 controls with 22q11.2DS and found suggestive evidence for association with several variants in the LCR22C–LCR22D region (Figure S2A). We repeated these analyses but restricted the affected individuals to only those with CTDs ($n = 424$) and identified significant associations in the same region. The association signal was in a cluster of 62 SNVs that are located in a 350 kb region largely within LCR22C–LCR22D (chr22: 20607741–2095814; hg38; Figures 4A and 4B) and that are in strong linkage disequilibrium (LD) (Figure 4C). The SNVs have an AAF > 0.05 (Figure 4D), with p values ranging from 1.00×10^{-3} to 2.59×10^{-5}, and 45 passed FDR correction from logistic regression analyses (Table S4). This region contains four functional genes that include *PI4KA* (phosphatidylinositol 4-kinase alpha [MIM: 600286]), *SERPIND1* (serpin family D member 1 [MIM: 142360]), *SNAP29* (synaptosome associated protein 29 [MIM: 604202]), and *CRKL* as shown in Figure 4E. Variants that passed FDR correction were highlighted in red for all CTD samples (Figure 4A) and for the Caucasian subset (Figure 4B). These variants were not significantly associated with CTDs in the African-admixed or Hispanic populations, possibly because the sample sizes were very small (Figures S2B and S2C). Of note, none of the SNVs in the coding or non-coding regions of *TBX1* showed association with risk for CTDs (all $p > 0.05$) in any subpopulation (Figures 4B and S2B and S2C).

We next performed phenotype subset analyses. The difference between the CHD and CTD categories was the inclusion of isolated VSD or ASD in the CHD category. Logistic-regression analyses for isolated VSD or ASD showed that they do not contribute to the association signals between LCR22C and LCR22D (Figures S2D and S2E). In fact, adding

Table 1. Demographic and Clinical Characteristics of the 22q11.2DS Cohort

Variables	No. (%) for Categorical Variables for All 1,053 Samples	No. (%) for Categorical Variables for 893 Samples in the Main Studied CTD Cohort
Gender		
Male	512 (48.6%)	430 (48.2%)
Female	541 (51.4%)	463 (51.8%)
Deletion Type		
AD	1,053 (100%)	893 (100%)
CHD Status[a]		
Normal heart	469 (44.5%)	469 (52.5%)
CTD	424 (40.3%)	424 (47.5%)
CHD	584 (65.0%)	NA
TOF-PTA-IAAB	284 (26.9%)	NA
TOF	194 (18.4%)	NA
ASD alone	55 (5.2%)	NA
VSD alone	105 (10.0%)	NA
Population Origin		
Caucasian	790 (75.2%)	669 (74.9%), 312:357[c]
AA[b] and admixed	161 (15.3%)	135 (15.1%), 68:67[c]
Hispanic	102 (9.7%)	89 (10.0%), 44:45[c]

[a]Definition of CHD phenotypes: patients are coded as cases for the specific CHD categories if they satisfied the corresponding definitions. Controls are those with no heart or aortic-arch anomalies. CTD, conotruncal heart defect; TOF, tetralogy of Fallot; PTA, persistent truncus arteriosus; and IAAB, interrupted aortic arch.
[b]AA, African-admixed.
[c]Ratio of CTD affected individuals to controls.

the isolated VSD or ASD categories to the analysis reduced the association signals (Figure S2A), suggesting that they have different risk factors in the 22q11.2DS population.

After this, we tested for evidence regarding whether association differs within subgroups of CTDs in the Caucasian subpopulation with the caveat that group sizes vary. We first compared the subgroup of TOF-affected individuals with the same controls and then compared the subgroup that included TOF, PTA, or IAAB to the same controls also. Variants with a $p < 1.0 \times 10^{-3}$ for all of the categories were compiled into a table that included variants occurring from two or more subgroups. If any variant was included more than twice, the smallest p value with its corresponding phenotype category was retained. In total, there were 69 variants in the LCR22C–LCR22D region (Table S4). We did not observe a difference in the association test for the TOF or the combined TOF-PTA-IAAB versus the CTD categories and the same controls, suggesting that these subsets share a similar genetic risk.

The cardiac OFT forms from the second heart field mesoderm and neural crest mesenchymal cells, whereas the aortic arch and arterial branches develop from the pharyngeal arch arteries that contain vascular endothelium that is derived from mesoderm as well as smooth muscle cells that are derived from neural crest.[56,57] The CTD category consists of subjects with defects in cardiac OFT defects and/or the aortic arch. Because of their different embryological origins, we compared the two. As shown in Figure S3, the group with defects in cardiac OFT (Figure S3A) contributes much more to the strength of the association signal in the 350 kb region—the top variants pass FDR for multiple-testing correction—than do those in the group with defects in the aortic arch (Figure S3B). This result supports the findings that these have different developmental and anatomical origins, and therefore, their genetic control may also be distinct.

Set-Based Analysis for Common Variants and Odds Ratios

Focusing on CTDs in the 22q11.2DS population, we performed a set-based SKAT test to determine the genetic risk conferred by genes, promoters, and enhancers. Results from the set-based test for common variants is presented in Figure 5A and Table S5. We found that the gene *SERPIND1* and one double-elite enhancer, GH22J020946, were significantly associated with CTD risk after Bonferroni correction (Figure 5A). Data on chromatin regulation and chromatin interactions are shown in Figure 5B. These analyses indicate significant interactions between the four coding genes mapping to the 350 kb interval (Figure 5C).

There is functional genomic evidence for regulatory activity of double-elite enhancers and for particular gene targets.[53] In the 350 kb region, there are eleven double-elite enhancers (GH22J020748, GH22J020775,

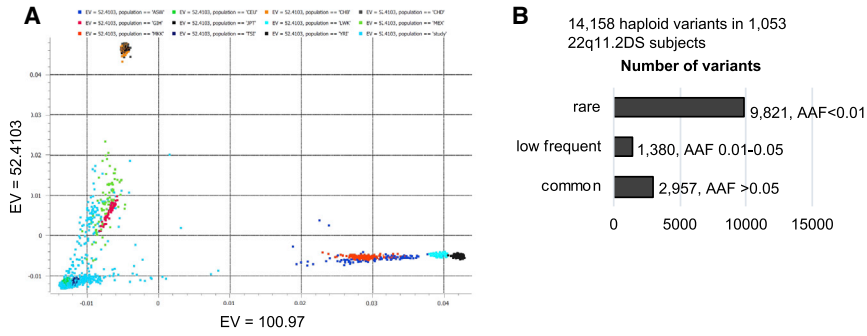

Figure 2. Ethnicity and Constituency of Variants in the 1,053 22q11.2DS Cohort
(A) Scatterplot of the first PC (EV = 100.97) versus the second PC (EV = 52.41) calculated from PCA based on common independent shared variants in the 22q11.2DS cohort (aqua color) and HapMap 3 r3 samples (International HapMap project phase III release 3, including 1,397 individuals from 11 populations across the globe; colored based upon the population). ASW, African ancestry in Southwest USA; CEU, Utah residents with ancestry from northern and western Europe; CHB, Han Chinese in Beijing, China; CHD, Chinese in Metropolitan Denver, Colorado; GIH, Gujarati Indians in Houston, Texas; JPT, Japanese in Tokyo, Japan; LWK, Luhya in Webuye, Kenya; MEX, Mexican ancestry in Los Angeles, California; MKK, Maasai in Kinyawa, Kenya; TSI (T): Tuscan in Italy; YRI (Y), Yoruba in Ibadan, Nigeria (West Africa).
(B) Frequency distribution of 14,158 variants that passed QC measures. In total, there are 9,821 rare variants with AAF < 0.01, 1,380 low-frequency variants (AAF between 0.01-0.05), and 2,957 common variants (AAF > 0.05).

GH22J020855, GH22J020866, GH22J020883, GH22J020916, GH22J020936, GH22J020939, GH22J020940, GH22J020946, and GH22J020947), as indicated in Table S6. These double-elite enhancers have not yet been assayed for function in human or mouse cardiac development but are of interest because they harbor non-coding common variants, such as the double-elite enhancer GH22J020946.

We examined the odds ratio (OR) of the SNVs in the 350 kb region to determine the individual and overall risk of the variants. The distribution of the OR as well as the corresponding 95% confidence intervals (CIs) were plotted in this 350 kb region (Figures 5D and 5E). Most of the variants were associated with increased odds of CTDs (median OR = 2.96;range = 1.64–4.75), whereas alternate alleles of three variants were associated with an OR less than 1 (OR range: 0.48–0.52) (Table S4).

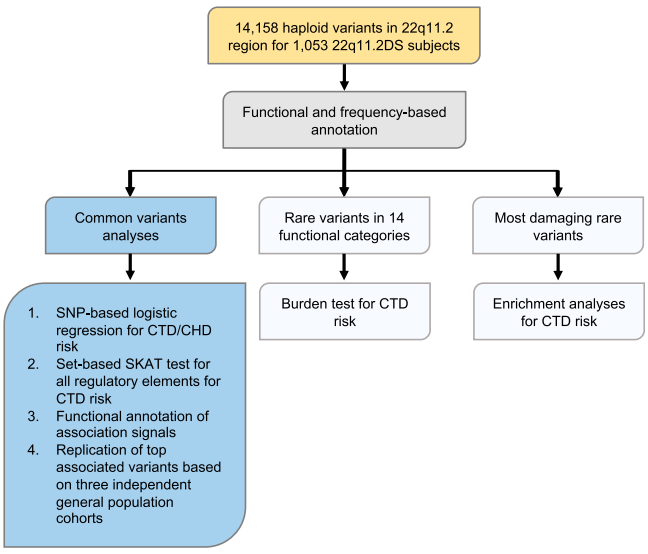

Figure 3. Data-Analyses Flowchart of this Study
Common variants and rare variants were determined by the alternate allele frequency (AAF) of individuals in the gnomAD database at the threshold of 0.01.

Replication of the Top Associated Variants Based in Three Different CTD Cohorts without a 22q11.2 Deletion

Of the 69 top associated variants, 49 were included in a meta-analysis of three GWAS of CTDs in no-deletion individuals.[34] Four of these variants had meta-analysis p values < 0.05, and the direction of association was the same as that observed in the 22q11.2-deleted cohort (Table 2). Three of these SNVs, rs165912, rs6004160, and rs738059, are in complete LD (i.e., $R^2 = 1$, based on Hapmap r3 Caucasian data) and were similarly associated with CTDs (OR = 1.10, 95% CI, 1.00–1.21, meta-p = 0.04) in the meta-analysis. Among the three, rs6004160 is an eQTL of *CRKL*. The most significantly replicated SNV, rs178252 (OR = 1.16 [1.04–1.30], meta-p = 0.006), is also an eQTL of *CRKL* (Table S7, Figure 5E). The SNV rs178252 resides in the double-elite enhancer GH22J020947 (10 kb in sequence), and one of the top predicted targets is *CRKL*, according to the GeneHancer database (Figure 5B). Moreover, among the 69 variants, the SNV rs178252 is one of two that maps to open chromatin regions,[51] as shown in Table 2, Figure S4, and Figure 5E.

Functional Annotation of Association Signals in the LCR22C–LCR22D Region

The presence of a large LD block poses a challenge to the discovery of possible causal variants in the associated 350 kb region. To identify possible functional variants, we generated CADD, LINSIGHT, and phastCons scores for the 69 SNVs. Detailed associations of the 69 variants are presented in Table S4. Of note, genotype-tissue expression (GTEx) data (muscle-skeletal) indicate that 12 of the SNVs are eQTLs of *CRKL*, suggesting that they might be functional. According to data from human stem cells, two SNVs reside in an open chromatin region[51] (Figure S4). All except one of the 69 SNVs are in non-coding sequences. The one coding variant, rs165854, has a low CADD score of 5.97, indicating it is less likely to be causal. The AAF of SNVs in both this 22q11.2DS cohort and that of the gnomAD database were plotted in Figure 4D. Variants have

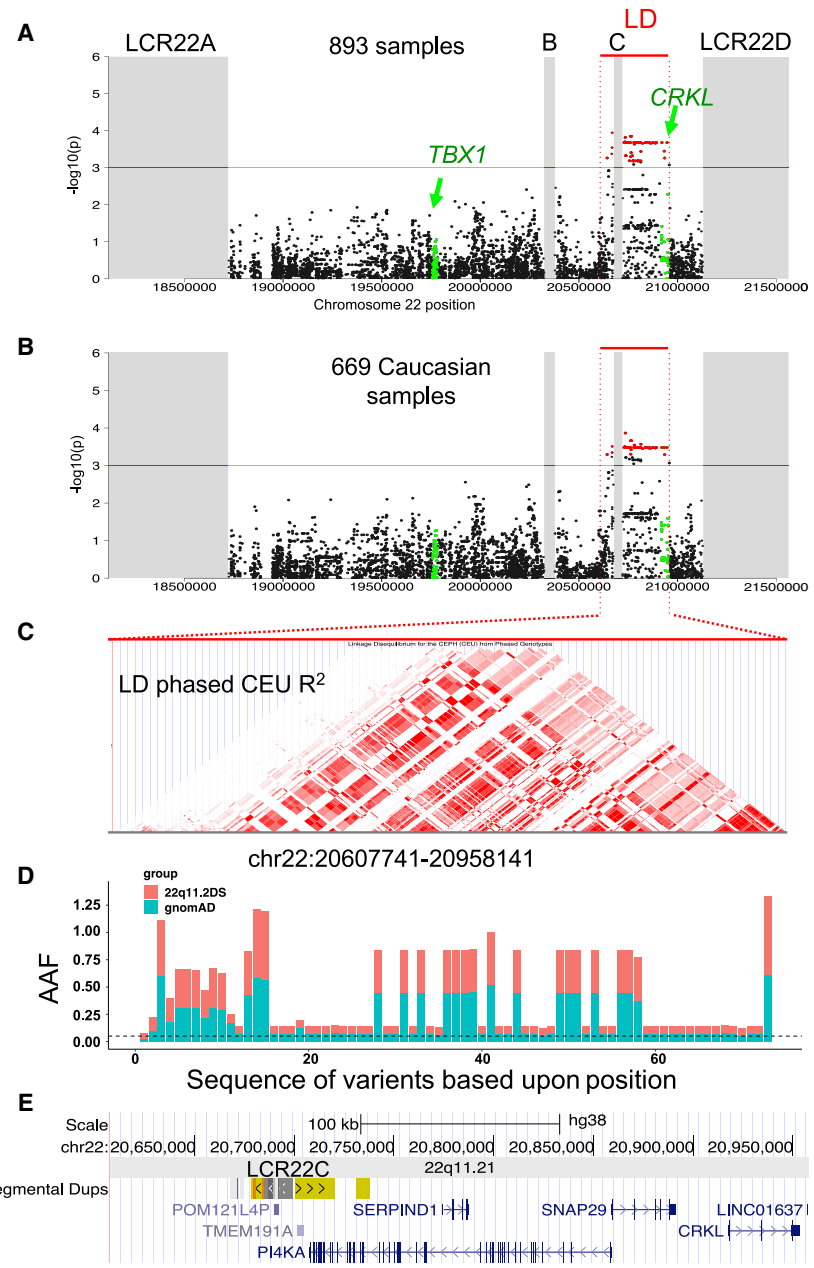

Figure 4. Regional Plot of Logistic Regression Analyses Identifies Significant Association to a 350 kb LCR22C- LCR22D Interval

(A and B) Regional plot for common variants versus CTD risk in the remaining allele of 22q11.2 with adjustment of sex and top five PCs in 893 samples (A) and top four PCs in 669 Caucasian samples (B). Variants in *TBX1* and *CRKL* loci are indicated (green). Gray blocks represent LCR22A, LCR22B, LCR22C, and LCR22D in the 22q11.2 region. Blue horizontal lines represent the threshold of suggestive significant association for multiple testing at 1.0×10^{-3}. Two vertical red dashed lines denote top associated variants (chr22: 20607741–20958141 in USCS Genome Browser Consortium assembly hg38). A red bar indicates LD block. Variants in red survived a false discovery rate (FDR) for multiple testing correction.

(C) LD structure in the Caucasian population as determined from HapMap data in the top associated region in the UCSC Genome Browser.

(D) Bar plot of the alternate allele frequency (AAF) of top variants in the 22q11.2DS population (red) as well as in the gnomAD database (aqua) in the top associated region.

(E) Snapshot of the UCSC Genome Browser indicating the genomic interval of 22q11.2, Segmental Duplication track indicating LCR22C, and the RefSeq gene alignment in the top associated region.

similar AAFs in both datasets indicating both that our data are of high quality and that subsets of the variants have similar AAFs in either dataset. The latter of these is evidence that those variants are in high to complete LD.

Enrichment Analyses of Rare Deleterious Variants in the 22q11.2DS Cohort

We also tested the hypothesis that rare variants might alter risk for CHD or CTDs in the cohort. Three rare LoF variants, six high-confidence D-Mis variants, and four high-confidence D-splicing variants were identified in the remaining allele of 22q11.2 (Table S8). Of note, none of these LoF, D-Mis, and D-splicing variants were located in *TBX1* or *CRKL*.[1] Those variants occurred in nine CHD-affected individuals, seven CTD-affected individuals, and ten controls, and there was no enrichment in CTD-affected individuals, indicating that it is unlikely that predicted damaging variants in the 22q11.2 region could account for CHD or CTD risk and explain the phenotypic heterogeneity in this population. No significant enrichment of rare variants in the Caucasian subpopulation in any of the 14 most functional plausible categories was identified (all empirical $p > 0.05$; Figure S5). When taken together, these resuts indicate that it is unlikely that rare coding variants in this region influence risk for CHD or CTDs in individuals with 22q11.2DS.

SNV-Based Fisher's Exact Test and Set-Based Burden Test for Analyses for Rare Variants in the Caucasian Population

We then tested the hypothesis that the association signal in the 350 kb region in the LCR22C–LCR22D interval for CTDs is driven by rare variants in the same region. Because of issues related to population stratification for rare-variant association analysis, we focused on the largest population, which was the Caucasian population. Fisher's exact test was run in samples from 669 Caucasian subjects, including those with CTDs and controls. The distribution of rare variants, individually, between CTD-affected individuals and

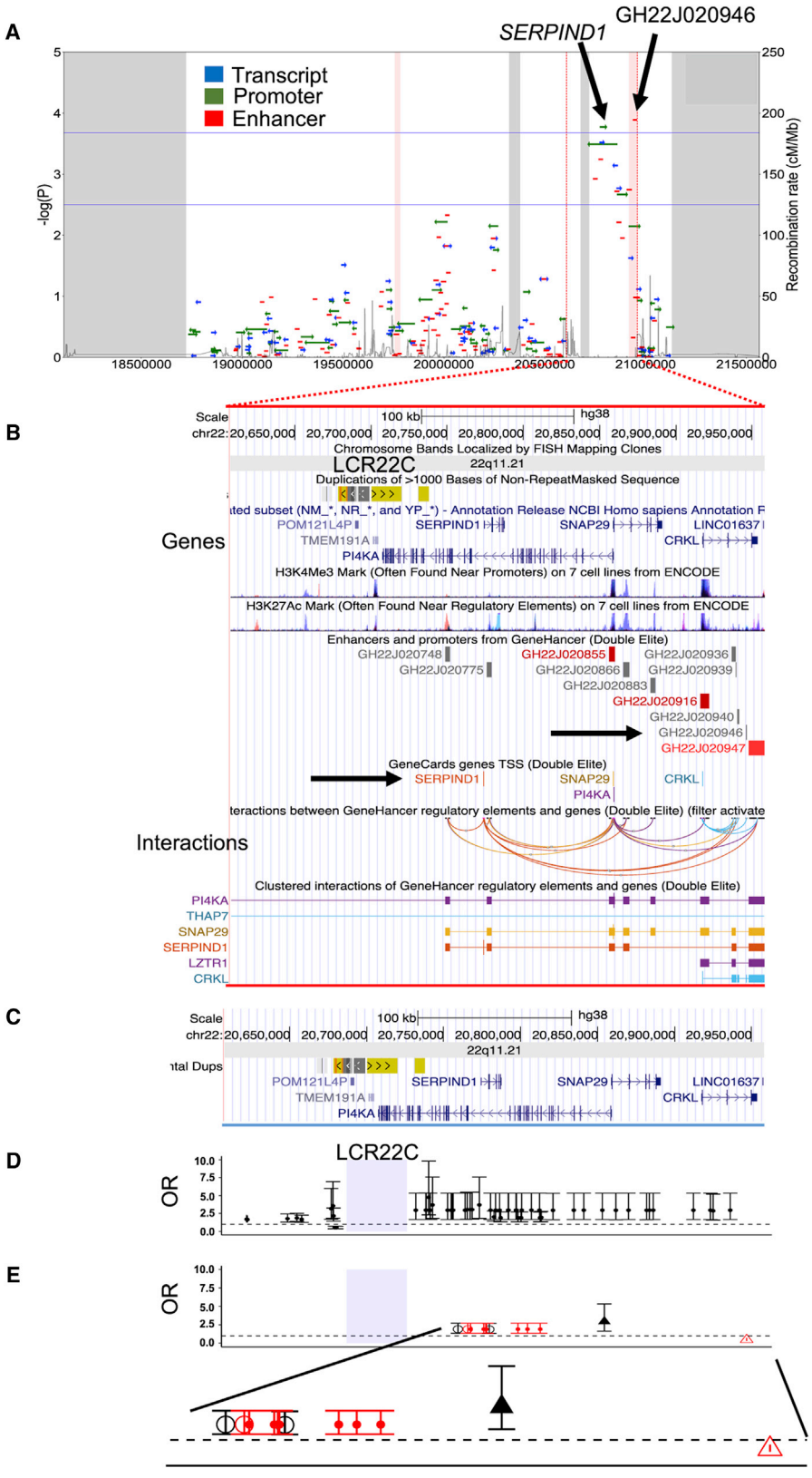

Figure 5. Set-Based SKAT Test for Common Variants Identifies Non-coding Variants in Gene-Regulatory Regions in the 350 kb LCR22C–LCR22D Region

(A) SKAT (sequence kernel association test) with adjustment of sex and top four PCs was applied to common variants in the cohort of 669 Caucasian samples. The set includes all 72 predicted coding and non-coding genes (blue; between transcriptional start site [TSS] and transcriptional termination site [TTS] plus 2 kb both upstream and downstream) from LCR22A–LCR22D, 72 putative promoters of these genes (green; both 2 kb upstream and downstream of TSS), and 96 curated double-elite sets of enhancers (red) in the 22q11.2 region.

(legend continued on next page)

controls was not significantly different (Figure S6). Moreover, there was no significant difference between CTD-affected individuals and controls for the joint effect of rare variants when they were aggregated into variants in genes, promoter regions, and enhancer regions on the basis of burden tests (Figure S7). Of note, the burden test had a $p > 0.05$ for both *TBX1* and *CRKL*. These results indicate that it is unlikely that rare variants account for the significant association found for common variants of CTD risk in LCR22C–LCR22D. Therefore, the association signals of common variants are unlikely to be driven by rare variants, pointing to the possibility that common variants might drive the signal that was observed.

Discussion

In this report, we found that common variants in a 350 kb region on the remaining allele, largely within the LCR22C–LCR22D interval, are associated with moderate increased risk (OR range: 1.6–4.8) for CTDs in individuals with the typical 3 Mb 22q11.2 deletion. Complete sequence data indicate that this association was not driven by rare variants individually or jointly in the same region, suggesting that associated common variants are among the top causal variants.

Different 22q11.2 Deletions and Similar Phenotypes

Individuals diagnosed with 22q11.2DS have highly variable clinical phenotypes. One of the early prevailing hypotheses was that this clinical variation was due to differences in deletion sizes and that these differences resulted in haploinsufficiency of different genes. Although early data suggested that the LCR22A–LCR22B region was the critical region for the syndrome,[6,7,58–60] patients were identified recently with non-overlapping LCR22B–LCR22D and LCR22C–LCR22D deletions.[11,13] Despite having non-overlapping deletions, these individuals had similar CTDs with half the frequency.[14] The existence of patients with LCR22B–LCR22D or LCR22C–LCR22D deletions and CTDs, coupled with the data presented in this report, supports the likelihood that the LCR22C–LCR22D region contains modifiers of CTD risk. Individuals with the LCR22A–LCR22C deletion have a CTD frequency similar to that in the LCR22A–LCR22A B or LCR22A–LCR22A D region. Thus, we suggest that *CRKL* acts as a risk factor but that the 22q11.2 region is quite complex and that our understanding of the biology of the region is still in its infancy.

Evidence that Non-coding Variants in *CRKL* Might Modify the Phenotype in 22q11.2DS

We identified a significant association between common variants in a 350 kb interval within the LCR22C–LCR22A D region. Because the variants in the 350 kb region are in LD, it is not possible to rule out individual genes without further functional annotation. There are four known protein coding genes that map to the interval: *PI4KA*, *SERPIND1*, *SNAP29*, and *CRKL*. The *PI4KA* gene product functions as a critical enzyme in the metabolism of plasma membrane phosphoinositides by catalyzing one of the early steps of membrane lipid formation.[61] Inactivation of *PI4KA* in the mouse results in early embryonic lethality, and it is therefore an essential gene for embryonic development.[62] Recessive mutations in *PI4KA* were discovered in one family in which three fetuses carrying compound heterozygous mutations had polymicrogyria and cerebellar hypoplasia, as well as other malformations, but cardiovascular anomalies were not present.[63] In studies in zebrafish, knock down of *pi4ka* by morpholino injection resulted in abnormal fin development, shortened body axis, and pericardial edema, which could implicate functions in the heart.[64] Thus, it is not known whether non-coding variants in the locus could alter *PI4KA* expression and, thus, could increase risk for CTDs. As mentioned above, *SERPIND1* encodes heparin co-factor II, which inhibits thrombin activity. Inactivation of *SERPIND1* in humans causes excess thrombin and deep-vein thrombosis.[64–66] In one report, *Serpind1*-null mutant mice were found to survive in normal Mendelian ratios,[67] and in another report of a similar null mutant, the mice were embryonic lethal and had vascular remodeling defects, but they did not have cardiac developmental defects.[68] The *SNAP29* gene product is located in the cytoplasm and is involved in intracellular membrane vesicle fusion and membrane trafficking.[69] Recessive mutations in *SNAP29*

The pink background highlights *TBX1* and *CRKL*. Two blue horizontal lines represent suggestive and Bonferroni-corrected significant thresholds for multiple testing at 2.5×10^{-3} and 2.1×10^{-4}, respectively. Two vertical dashed red lines denote where the top associated signals reside from logistic-regression analyses (chr22: 20607741–20958141, in hg38). Arrows depict a significant gene (*SERPIND1*) and enhancer (GH22J020946). Four gray blocks represent LCR22A, LCR22B, LCR22C, and LCR22D.

(B) Snapshot of the UCSC Genome Browser in hg38 assembly showing the genomic context in the 350 kb top associated region. Included are the segmental duplication track (LCR22C), RefSeq genes, H3K4Me3 (promoter), and H2K27Ac (enhancer) marks from cell lines, double-elite sets of enhancers (gray) and promoters (red), indication of TSSs, and the interactions between enhancers and genes in this region. Arrows point to the double-elite enhancer and gene found above (A). Note that five double-elite enhancers (GH22J020947, GH22J020946, GH22J020940, GH22J020939, and GH22J020936) regulate *CRKL*, as indicated below the regulatory region interactions.

(C) Gene alignment in the top associated region is based on the RefSeq gene track in the UCSC Genome Browser.

(D and E) Distribution of the ORs of the 69 top associated variants ($p < 1.0 \times 10^{-3}$), of which seven variants are eQTLs of *CRKL* (red dots in [E]); three empty circles and one empty triangle in (E) represent variants of which the association with CTD risk was also found in the CTD cohorts without 22q11.2 deletion; two closed black triangles in (E) denote variants located in an open chromatin region. Of note, rs178252 includes all of the three features. An enlarged image is below.

Table 2. Association Results of Four Replicated Variants in Three Independent CTD Cohorts without 22q11.2 Deletion and Meta-Analysis, as well as Primary Results in the 22q11.2DS Cohort

Datasets	No. of Individuals	rs178252 (G > A)			rs165912 (C > T)			rs6004160 (G > A)			rs738059 (G > A)		
		AAF[a]	OR (95% CI)[b]	p Value	AAF	OR (95% CI)	p Value	AAF	OR (95% CI)	p Value	AFF	OR (95% CI)	p Value
22q11.2DS cohort	469:424[d]	0.720	1.67 (1.24–2.25)	6.90×10^{-4}	0.393	1.62 (1.23–2.14)	6.03×10^{-4}	0.939	1.65 (1.25–2.18)	3.66×10^{-4}	0.392	1.64 (1.24–2.16)	4.80×10^{-4}
Meta-analysis[c]			1.16 (1.04–1.30)	0.006		1.10 (1.00–1.21)	0.04		1.10 (1.00–1.21)	0.04		1.10 (1.00–1.21)	0.04
eQTL of *CRKL*[e]		Y			N			Y			N		
Open chromatin region[f]		Y			N			N			N		

[a]Relative risk based on comparison of heterozygote to common homozygous genotypes.
[b]Likelihood ratio test comparing the full model with inherited genotype modeled as an additive effect and the unrestricted maternal genotype to the model with just the unrestricted maternal genotype.
[c]Meta-analyses of three CTD cohorts without 22q11.2 deletion.
[d]Ratio of affected individuals to controls.
[e]eQTL data was downloaded from GTEx.
[f]Determination of whether the variants reside in open chromatin region is based on ATAC-seq data from human induced pluripotent stem cells and human embryonic stem cells[51]

in humans cause cerebral dysgenesis, neuropathy, and skin conditions, termed CEDNIK syndrome [MIM: 609528].[70] Among these four genes, *CRKL* is the most likely candidate gene for which altered expression by non-coding variants on the remaining allele of 22q11.2 might influence risk for CTDs.

The top associated variants identified in this study lie in putative regulatory regions of *CRKL*, as described above. *Crkl* is a ubiquitously expressed gene, and mouse model data suggest that it functions in neural-crest cells within the pharyngeal apparatus for cardiovascular development.[25] Inactivation of both alleles of *Tbx1* or *Crkl* in the mouse results in a similar spectrum and range of CTDs,[16–18] and these alleles genetically interact during embryogenesis.[26] The connection between these genes is based upon the hypothesis that *Tbx1* acts upstream of the growth factor ligand FGF8 (fibroblast growth factor 8 [MIM: 600483]) and that CRKL acts downstream and thereby activates the MAPK (mitogen-activated protein kinase) signaling pathway, which is critically important for heart, aortic arch, and arterial-branch formation.[71] Therefore, there is precedence from mouse genetic studies that *CRKL* might act as a modifier, and variants that might reduce expression of *CRKL* would increase risk for CTDs. One question is whether *CRKL* is sensitive to altered gene dosage. An allelic series generated in the mouse showed that *Crkl* is partially sensitive to altered gene dosage in embryonic development.[14] Various cardiovascular anomalies that included CTDs were identified.[14]

Strengths and Limitations in Examination of Sequence in the 22q11.2 Allele

The availability of WGS is a strength of the study, and good coverage of the interval was obtained. Although the sample size was large for a rare condition such as 22q11.2DS, we evaluated several thousand variants and, given statistical correction for multiple comparisons, power was still quite low for variants of low effect size. Consequently, we might have missed true associations. Fine mapping to identify so-called causal variants is exceedingly difficult because the variants of interest were in high to complete LD. With the availability of data sources such as GTEx and GeneHancer, we found that top associated variants might affect expression levels of *CRKL*. Another strength lies in the availability of summary p value statistics from no-deletion CTD-affected individuals for whom supportive data were available. Variants identified in this report can be tested in the future by functional validation and *in vivo* studies. Examination of chromatin structure by high-throughput chromatin conformation capture (Hi-C) implies that this 350 kb region is within a single topological associated domain.[72] On the other hand, a recent study reported that local and global chromatin interactions were altered dynamically in a multilayered fashion in 22q11.2 lymphoblastoid cell lines as compared to control cell lines without deletion.[72] Therefore, the deletion might affect chromatin interactions outside this interval, and other genes could serve as modifiers. It is therefore possible that a regulatory element within the 350 kb region could alter *TBX1* expression or that of another gene elsewhere on 22q11.2. Nonetheless, the fact that patients with LCR22B–LCR22D and LCR22C–LCR22D deletions have CHD at 20%–30% suggests that this region is critically important as a potential modifier of cardiac development in 22q11.2DS.

Conclusions

In this report, we found that a cluster of common SNVs in the LCR22C–LCR22D region on the remaining allele of 22q11.2 is associated with the risk for CTDs in individuals with 22q11.2DS. Haploinsufficiency of this region alone is associated with CTDs, and when taken together with mouse genetic studies, the results presented here implicate, mostly plausibly, *CRKL* as a possible target of non-coding putative regulatory variants.

Supplemental Data

Supplemental Data can be found online at https://doi.org/10.1016/j.ajhg.2019.11.010.

Acknowledgments

We would like to thank the 22q11.2DS-affected families who provided DNA and clinical information for this study. We acknowledge the Genomics and Molecular Cytogenetics Cores at the Albert Einstein College of Medicine. We thank the Pediatric Cardiac Genomics Consortium for data collection and management and for the use of published data, without which the replication of the findings in our 22q11.2DS cohort in CTD cohorts without a 22q11.2 deletion would have never been possible. Dr. Morrow was supported by a Leducq Foundation grant and National Institutes of Health grants R01HL132577, R01HL084410, U01MH101720, U54HD090260, and P01HD070454. Other funding sources are detailed in the Supplemental Information.

Declaration of Interests

The authors declare no competing interests.

Received: August 28, 2019
Accepted: November 13, 2019
Published: December 20, 2019

Web Resources

Bystro, https://bystro.io/
dbNSFP, https://sites.google.com/site/jpopgen/dbNSFP
DbSUPER, http://asntech.org/dbsuper/
Ensembl Regulatory Build, https://useast.ensembl.org/info/genome/funcgen/regulatory_build.html
EPDnew, https://epd.epfl.ch//index.php
FANTOM5, http://fantom.gsc.riken.jp/5/
GENCODE, https://www.encodeproject.org/)
GeneHancer database, https://www.genecards.org/
gnomAD, https://gnomad.broadinstitute.org/

GTEx, https://gtexportal.org/home
Hapmap 3 r3 data, ftp://ftp.ncbi.nlm.nih.gov/hapmap/
OMIM, https://www.omim.org/
PLINK, https://www.cog-genomics.org/plink/1.9/
SnpEff, http://snpeff.sourceforge.net/
SVS Golden Helix software, https://doc.goldenhelix.com/SVS/latest/svsmanual/numeric_data_quality.html
UCSC genome browser, https://genome.ucsc.edu/

References

1. McDonald-McGinn, D.M., Sullivan, K.E., Marino, B., Philip, N., Swillen, A., Vorstman, J.A., Zackai, E.H., Emanuel, B.S., Vermeesch, J.R., Morrow, B.E., et al. (2015). 22q11.2 deletion syndrome. Nat. Rev. Dis. Primers *1*, 15071.
2. Botto, L.D., May, K., Fernhoff, P.M., Correa, A., Coleman, K., Rasmussen, S.A., Merritt, R.K., O'Leary, L.A., Wong, L.Y., Elixson, E.M., et al. (2003). A population-based study of the 22q11.2 deletion: phenotype, incidence, and contribution to major birth defects in the population. Pediatrics *112*, ^{10}1–107.
3. Oskarsdóttir, S., Vujic, M., and Fasth, A. (2004). Incidence and prevalence of the 22q11 deletion syndrome: a population-based study in Western Sweden. Arch. Dis. Child. *89*, 148–151.
4. Grati, F.R., Molina Gomes, D., Ferreira, J.C., Dupont, C., Alesi, V., Gouas, L., Horelli-Kuitunen, N., Choy, K.W., García-Herrero, S., de la Vega, A.G., et al. (2015). Prevalence of recurrent pathogenic microdeletions and microduplications in over 9500 pregnancies. Prenat. Diagn. *35*, 801–809.
5. Maisenbacher, M.K., Merrion, K., Pettersen, B., Young, M., Paik, K., Iyengar, S., Kareht, S., Sigurjonsson, S., Demko, Z.P., and Martin, K.A. (2017). Incidence of the 22q11.2 deletion in a large cohort of miscarriage samples. Mol. Cytogenet. *10*, 6.
6. Edelmann, L., Pandita, R.K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R.S., Magenis, E., Shprintzen, R.J., and Morrow, B.E. (1999). A common molecular basis for rearrangement disorders on chromosome 22q11. Hum. Mol. Genet. *8*, 1157–1167.
7. Shaikh, T.H., Kurahashi, H., Saitta, S.C., O'Hare, A.M., Hu, P., Roe, B.A., Driscoll, D.A., McDonald-McGinn, D.M., Zackai, E.H., Budarf, M.L., and Emanuel, B.S. (2000). Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. Hum. Mol. Genet. *9*, 489–501.
8. Edelmann, L., Pandita, R.K., and Morrow, B.E. (1999). Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. Am. J. Hum. Genet. *64*, 1076–1086.
9. Burn, J., and Goodship, J. (1996). Developmental genetics of the heart. Curr. Opin. Genet. Dev. *6*, 322–325.
10. Unolt, M., Versacci, P., Anaclerio, S., Lambiase, C., Calcagni, G., Trezzi, M., Carotti, A., Crowley, T.B., Zackai, E.H., Goldmuntz, E., et al. (2018). Congenital heart diseases and cardiovascular abnormalities in 22q11.2 deletion syndrome: From well-established knowledge to new frontiers. Am. J. Med. Genet. A. *176*, 2087–2098.
11. Burnside, R.D. (2015). 22q11.21 Deletion Syndromes: A Review of Proximal, Central, and Distal Deletions and Their Associated Features. Cytogenet. Genome Res. *146*, 89–99.
12. Verhagen, J.M., Diderich, K.E., Oudesluijs, G., Mancini, G.M., Eggink, A.J., Verkleij-Hagoort, A.C., Groenenberg, I.A., Willems, P.J., du Plessis, F.A., de Man, S.A., et al. (2012). Phenotypic variability of atypical 22q11.2 deletions not including TBX1. Am. J. Med. Genet. A. *158A*, 2412–2420.
13. Rump, P., de Leeuw, N., van Essen, A.J., Verschuuren-Bemelmans, C.C., Veenstra-Knol, H.E., Swinkels, M.E., Oostdijk, W., Ruivenkamp, C., Reardon, W., de Munnik, S., et al. (2014). Central 22q11.2 deletions. Am. J. Med. Genet. A. *164A*, 2707–2723.
14. Racedo, S.E., McDonald-McGinn, D.M., Chung, J.H., Goldmuntz, E., Zackai, E., Emanuel, B.S., Zhou, B., Funke, B., and Morrow, B.E. (2015). Mouse and human CRKL is dosage sensitive for cardiac outflow tract formation. Am. J. Hum. Genet. *96*, 235–244.
15. Peyvandi, S., Lupo, P.J., Garbarini, J., Woyciechowski, S., Edman, S., Emanuel, B.S., Mitchell, L.E., and Goldmuntz, E. (2013). 22q11.2 deletions in patients with conotruncal defects: data from 1,610 consecutive cases. Pediatr. Cardiol. *34*, 1687–1694.
16. Merscher, S., Funke, B., Epstein, J.A., Heyer, J., Puech, A., Lu, M.M., Xavier, R.J., Demay, M.B., Russell, R.G., Factor, S., et al. (2001). TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. Cell *104*, 619–629.
17. Lindsay, E.A., Vitelli, F., Su, H., Morishima, M., Huynh, T., Pramparo, T., Jurecic, V., Ogunrinu, G., Sutherland, H.F., Scambler, P.J., et al. (2001). Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. Nature *410*, 97–101.
18. Papaioannou, V.E. (2014). The T-box gene family: emerging roles in development, stem cells and cancer. Development *141*, 3819–3833.
19. Gong, W., Gottlieb, S., Collins, J., Blescia, A., Dietz, H., Goldmuntz, E., McDonald-McGinn, D.M., Zackai, E.H., Emanuel, B.S., Driscoll, D.A., and Budarf, M.L. (2001). Mutation analysis of TBX1 in non-deleted patients with features of DGS/VCFS or isolated cardiovascular defects. J. Med. Genet. *38*, E45.
20. Yagi, H., Furutani, Y., Hamada, H., Sasaki, T., Asakawa, S., Minoshima, S., Ichida, F., Joo, K., Kimura, M., Imamura, S., et al. (2003). Role of TBX1 in human del22q11.2 syndrome. Lancet *362*, 1366–1373.
21. Paylor, R., Glaser, B., Mupo, A., Ataliotis, P., Spencer, C., Sobotka, A., Sparks, C., Choi, C.H., Oghalai, J., Curran, S., et al. (2006). Tbx1 haploinsufficiency is linked to behavioral disorders in mice and humans: implications for 22q11 deletion syndrome. Proc. Natl. Acad. Sci. USA *103*, 7729–7734.
22. Torres-Juan, L., Rosell, J., Morla, M., Vidal-Pou, C., García-Algas, F., de la Fuente, M.A., Juan, M., Tubau, A., Bachiller, D., Bernues, M., et al. (2007). Mutations in TBX1 genocopy the 22q11.2 deletion and duplication syndromes: a new susceptibility factor for mental retardation. Eur. J. Hum. Genet. *15*, 658–663.
23. Rauch, R., Hofbeck, M., Zweier, C., Koch, A., Zink, S., Trautmann, U., Hoyer, J., Kaulitz, R., Singer, H., and Rauch, A. (2010). Comprehensive genotype-phenotype analysis in 230 patients with tetralogy of Fallot. J. Med. Genet. *47*, 321–331.
24. Ogata, T., Niihori, T., Tanaka, N., Kawai, M., Nagashima, T., Funayama, R., Nakayama, K., Nakashima, S., Kato, F., Fukami, M., et al. (2014). TBX1 mutation identified by exome sequencing in a Japanese family with 22q11.2 deletion syndrome-like craniofacial features and hypocalcemia. PLoS ONE *9*, e91598.
25. Guris, D.L., Fantes, J., Tara, D., Druker, B.J., and Imamoto, A. (2001). Mice lacking the homologue of the human 22q11.2

gene CRKL phenocopy neurocristopathies of DiGeorge syndrome. Nat. Genet. *27*, 293–298.
26. Guris, D.L., Duester, G., Papaioannou, V.E., and Imamoto, A. (2006). Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. Dev. Cell *10*, 81–92.
27. van der Linde, D., Konings, E.E., Slager, M.A., Witsenburg, M., Helbing, W.A., Takkenberg, J.J., and Roos-Hesselink, J.W. (2011). Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. J. Am. Coll. Cardiol. *58*, 2241–2247.
28. Guo, T., McDonald-McGinn, D., Blonska, A., Shanske, A., Bassett, A.S., Chow, E., Bowser, M., Sheridan, M., Beemer, F., Devriendt, K., et al.; International Chromosome 22q11.2 Consortium (2011). Genotype and cardiovascular phenotype correlations with TBX1 in 1,022 velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome patients. Hum. Mutat. *32*, 1278–1289.
29. Mlynarski, E.E., Sheridan, M.B., Xie, M., Guo, T., Racedo, S.E., McDonald-McGinn, D.M., Gai, X., Chow, E.W., Vorstman, J., Swillen, A., et al.; International Chromosome 22q11.2 Consortium (2015). Copy-Number Variation of the Glucose Transporter Gene SLC2A3 and Congenital Heart Defects in the 22q11.2 Deletion Syndrome. Am. J. Hum. Genet. *96*, 753–764.
30. Mlynarski, E.E., Xie, M., Taylor, D., Sheridan, M.B., Guo, T., Racedo, S.E., McDonald-McGinn, D.M., Chow, E.W., Vorstman, J., Swillen, A., et al.; International Chromosome 22q11.2 Consortium (2016). Rare copy number variants and congenital heart defects in the 22q11.2 deletion syndrome. Hum. Genet. *135*, 273–285.
31. Guo, T., Repetto, G.M., McDonald McGinn, D.M., Chung, J.H., Nomaru, H., Campbell, C.L., Blonska, A., Bassett, A.S., Chow, E.W.C., Mlynarski, E.E., et al.; International 22q11.2 Consortium/Brain and Behavior Consortium* (2017). Genome-wide association study to find modifiers for tetralogy of Fallot in the 22q11.2 deletion syndrome identifies variants in the *GPR98* locus on 5q14.3. Circ Cardiovasc Genet *10*.
32. Guo, T., Chung, J.H., Wang, T., McDonald-McGinn, D.M., Kates, W.R., Hawuła, W., Coleman, K., Zackai, E., Emanuel, B.S., and Morrow, B.E. (2015). Histone modifier genes alter conotruncal heart phenotypes in 22q11.2 deletion syndrome. Am. J. Hum. Genet. *97*, 869–877.
33. Lin, J.R., Zhang, Q., Cai, Y., Morrow, B.E., and Zhang, Z.D. (2017). Integrated rare variant-based risk gene prioritization in disease case-control sequencing studies. PLoS Genet. *13*, e1007142.
34. Agopian, A.J., Goldmuntz, E., Hakonarson, H., Sewda, A., Taylor, D., Mitchell, L.E.; and Pediatric Cardiac Genomics Consortium* (2017). Genome-wide association studies and meta-analyses for congenital heart defects. Circ Cardiovasc Genet *10*, e001449.
35. Guo, T., Diacou, A., Nomaru, H., McDonald-McGinn, D.M., Hestand, M., Demaerel, W., Zhang, L., Zhao, Y., Ujueta, F., Shan, J., et al.; International Chromosome 22q11.2, International 22q11.2 Brain and Behavior Consortia (2018). Deletion size analysis of 1680 22q11.2DS subjects identifies a new recombination hotspot on chromosome 22q11.2. Hum. Mol. Genet. *27*, 1150–1163.
36. Gur, R.E., Bassett, A.S., McDonald-McGinn, D.M., Bearden, C.E., Chow, E., Emanuel, B.S., Owen, M., Swillen, A., Van den Bree, M., Vermeesch, J., et al. (2017). A neurogenetic model for the study of schizophrenia spectrum disorders: the International 22q11.2 Deletion Syndrome Brain Behavior Consortium. Mol. Psychiatry *22*, 1664–1672.
37. Johnston, H.R., Chopra, P., Wingo, T.S., Patel, V., Epstein, M.P., Mulle, J.G., Warren, S.T., Zwick, M.E., and Cutler, D.J. (2017). Reply to Plüss et al.: The strength of PEMapper/PECaller lies in unbiased calling using large sample sizes. Proc. Natl. Acad. Sci. USA *114*, E8323.
38. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
39. Kotlar, A.V., Trevino, C.E., Zwick, M.E., Cutler, D.J., and Wingo, T.S. (2018). Bystro: Rapid online variant annotation and natural-language filtering at whole-genome scale. Genome Biol. *19*, 14.
40. Cingolani, P., Platts, A., Wang, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) *6*, 80–92.
41. Liu, X., Wu, C., Li, C., and Boerwinkle, E. (2016). dbNSFP v3.0: A one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. Hum. Mutat. *37*, 235–241.
42. MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., et al.; 1000 Genomes Project Consortium (2012). A systematic survey of loss-of-function variants in human protein-coding genes. Science *335*, 823–828.
43. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. *47* (D1), D886–D894.
44. Sim, N.L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P.C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res. *40*, W452–W457.
45. Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. Nat. Protoc. *11*, 1–9.
46. Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K., and Liu, X. (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. *24*, 2125–2137.
47. Chun, S., and Fay, J.C. (2009). Identification of deleterious mutations within three human genomes. Genome Res. *19*, 1553–1561.
48. Schwarz, J.M., Cooper, D.N., Schuelke, M., and Seelow, D. (2014). MutationTaster2: mutation prediction for the deep-sequencing age. Nat. Methods *11*, 361–362.
49. Huang, Y.F., Gulko, B., and Siepel, A. (2017). Fast, scalable prediction of deleterious noncoding variants from functional and population genomic data. Nat. Genet. *49*, 618–624.
50. Consortium, G.T.; and GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science *348*, 648–660.
51. Liu, Q., Jiang, C., Xu, J., Zhao, M.T., Van Bortle, K., Cheng, X., Wang, G., Chang, H.Y., Wu, J.C., and Snyder, M.P. (2017).

Genome-wide temporal profiling of transcriptome and open chromatin of early cardiomyocyte differentiation derived from hiPSCs and hESCs. Circ. Res. *121*, 376–391.
52. Nakagomi, H., Mochizuki, H., Inoue, M., Hirotsu, Y., Amemiya, K., Sakamoto, I., Nakagomi, S., Kubota, T., and Omata, M. (2018). Combined annotation-dependent depletion score for BRCA1/2 variants in patients with breast and/or ovarian cancer. Cancer Sci. *109*, 453–461.
53. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M., et al. (2017). GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*.
54. Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J.D., and Lin, X. (2013). Sequence kernel association tests for the combined effect of rare and common variants. Am. J. Hum. Genet. *92*, 841–853.
55. Lee, S., Fuchsberger, C., Kim, S., and Scott, L. (2016). An efficient resampling method for calibrating single and gene-based rare variant association analysis in case-control studies. Biostatistics *17*, 1–15.
56. Vincent, S.D., and Buckingham, M.E. (2010). How to make a heart: The origin and regulation of cardiac progenitor cells. Curr. Top. Dev. Biol. *90*, 1–41.
57. Kirby, M.L., and Waldo, K.L. (1995). Neural crest and cardiovascular patterning. Circ. Res. *77*, 211–215.
58. Lindsay, E.A., Goldberg, R., Jurecic, V., Morrow, B., Carlson, C., Kucherlapati, R.S., Shprintzen, R.J., and Baldini, A. (1995). Velo-cardio-facial syndrome: frequency and extent of 22q11 deletions. Am. J. Med. Genet. *57*, 514–522.
59. Morrow, B., Goldberg, R., Carlson, C., Das Gupta, R., Sirotkin, H., Collins, J., Dunham, I., O'Donnell, H., Scambler, P., Shprintzen, R., et al. (1995). Molecular definition of the 22q11 deletions in velo-cardio-facial syndrome. Am. J. Hum. Genet. *56*, 1391–1403.
60. Carlson, C., Sirotkin, H., Pandita, R., Goldberg, R., McKie, J., Wadey, R., Patanjali, S.R., Weissman, S.M., Anyane-Yeboa, K., Warburton, D., et al. (1997). Molecular definition of 22q11 deletions in 151 velo-cardio-facial syndrome patients. Am. J. Hum. Genet. *61*, 620–629.
61. Lees, J.A., Zhang, Y., Oh, M.S., Schauder, C.M., Yu, X., Baskin, J.M., Dobbs, K., Notarangelo, L.D., De Camilli, P., Walz, T., and Reinisch, K.M. (2017). Architecture of the human PI4KIIIα lipid kinase complex. Proc. Natl. Acad. Sci. USA *114*, 13720–13725.
62. Nakatsu, F., Baskin, J.M., Chung, J., Tanner, L.B., Shui, G., Lee, S.Y., Pirruccello, M., Hao, M., Ingolia, N.T., Wenk, M.R., and De Camilli, P. (2012). PtdIns4P synthesis by PI4KIIIα at the plasma membrane and its impact on plasma membrane identity. J. Cell Biol. *199*, 1003–1016.
63. Pagnamenta, A.T., Howard, M.F., Wisniewski, E., Popitsch, N., Knight, S.J., Keays, D.A., Quaghebeur, G., Cox, H., Cox, P., Balla, T., et al. (2015). Germline recessive mutations in PI4KA are associated with perisylvian polymicrogyria, cerebellar hypoplasia and arthrogryposis. Hum. Mol. Genet. *24*, 3732–3741.
64. Ma, H., Blake, T., Chitnis, A., Liu, P., and Balla, T. (2009). Crucial role of phosphatidylinositol 4-kinase IIIalpha in development of zebrafish pectoral fin is linked to phosphoinositide 3-kinase and FGF signaling. J. Cell Sci. *122*, 4303–4310.
65. Sie, P., Dupouy, D., Pichon, J., and Boneu, B. (1985). Constitutional heparin co-factor II deficiency associated with recurrent thrombosis. Lancet *2*, 414–416.
66. Tran, T.H., Marbet, G.A., and Duckert, F. (1985). Association of hereditary heparin co-factor II deficiency with thrombosis. Lancet *2*, 413–414.
67. Vicente, C.P., He, L., Pavão, M.S., and Tollefsen, D.M. (2004). Antithrombotic activity of dermatan sulfate in heparin cofactor II-deficient mice. Blood *104*, 3965–3970.
68. Aihara, K., Azuma, H., Akaike, M., Ikeda, Y., Sata, M., Takamori, N., Yagi, S., Iwase, T., Sumitomo, Y., Kawano, H., et al. (2007). Strain-dependent embryonic lethality and exaggerated vascular remodeling in heparin cofactor II-deficient mice. J. Clin. Invest. *117*, 1514–1526.
69. Steegmaier, M., Yang, B., Yoo, J.S., Huang, B., Shen, M., Yu, S., Luo, Y., and Scheller, R.H. (1998). Three novel proteins of the syntaxin/SNAP-25 family. J. Biol. Chem. *273*, 34171–34179.
70. Sprecher, E., Ishida-Yamamoto, A., Mizrahi-Koren, M., Rapaport, D., Goldsher, D., Indelman, M., Topaz, O., Chefetz, I., Keren, H., O'brien, T.J., et al. (2005). A mutation in SNAP29, coding for a SNARE protein involved in intracellular trafficking, causes a novel neurocutaneous syndrome characterized by cerebral dysgenesis, neuropathy, ichthyosis, and palmoplantar keratoderma. Am. J. Hum. Genet. *77*, 242–251.
71. Moon, A.M., Guris, D.L., Seo, J.H., Li, L., Hammond, J., Talbot, A., and Imamoto, A. (2006). Crkl deficiency disrupts Fgf8 signaling in a mouse model of 22q11 deletion syndromes. Dev. Cell *10*, 71–80.
72. Dixon, J.R., Xu, J., Dileep, V., Zhan, Y., Song, F., Le, V.T., Yardımcı, G.G., Chakraborty, A., Bann, D.V., Wang, Y., et al. (2018). Integrative detection and analysis of structural variation in cancer genomes. Nat. Genet. *50*, 1388–1398.