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], Daniel McGinn[4], T. Blaine Crowley[4], Gabriela M. Repetto[5], Ann Swillen[6], Jeroen Breckpot[6], Joris R. Vermeesch[6], Wendy R. Kates[7], M. Cristina Digilio[8], Marta Unolt[8,9], Bruno Marino[9], Maria Pontillo[10], Marco Armando[10,11], Fabio Di Fabio[9], Stefano Vicari[10,12], Marianne van den Bree[13], Hayley Moss[13], Michael J. Owen[13], Kieran C. Murphy[14], Clodagh M. Murphy[15], Declan Murphy[15], Kelly Schoch[16], Vandana Shashi[16], Flora Tassone[17], Tony J. Simon[17], Robert J. Shprintzen[18], Linda Campbell[19], Nicole Philip[20], Damian Heine-Suñer[21], Sixto García-Miñaúr[22], Luis Fernández[22], International 22q11.2 Brain and Behavior Consortium[23], Carrie E. Bearden[24], Claudia Vingerhoets[25], Therese van Amelsvoort[25], Stephan Eliez[11], Maude Schneider[11], Jacob Vorstman[26], Doron Gothelf[27], Elaine Zackai[4], A.J Agopain[3], Raquel E. Gur[28], Anne S. Bassett[29], Beverly S. Emanuel[4], Elizabeth Goldmuntz[30], Laura E. Mitchell[3], Tao Wang[31], Bernice E. Morrow[1]*

[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 and Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, 19104, USA

[5]Center for Genetics and Genomics, Facultad de Medicina Clinica Alemana-Universidad del Desarrollo, Santiago, 7710162, Chile

[6]Center for Human Genetics, University of Leuven (KU Leuven), Leuven, 3000, Belgium

[7]Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, NY, 13202, USA

[8]Department of Medical Genetics, Bambino Gesù Hospital, Rome, 00165, Italy

[9]Department of Pediatrics, Gynecology and Obstetrics La Sapienza, University of Rome, Rome, 00185, Italy

[10]Department of Neuroscience, Bambino Gesù Hospital, Rome, 00165, Italy

[11]Developmental Imaging and Psychopathology Lab, University of Geneva, Geneva, 1211, Switzerland

[12]Department of Psychiatry, Catholic University, Rome, 00153, Italy

[13] MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, Cardiff University, Wales, CF24 4HQ, UK

[14]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin, 505095 Ireland

[15] Department of Forensic and Neurodevelopmental Sciences, King's College London, Institute of Psychiatry, Psychology & Neuroscience, London, SE5 8AF, UK and Behavioural and Developmental Psychiatry Clinical Academic Group, Behavioural Genetics Clinic, National Adult Autism and ADHD Service, South London and Maudsley Foundation NHS Trust, London, SE5 8AZ, UK

[16]Department of Pediatrics, Duke University, Durham, NC, 27710, USA

[17]M.I.N.D. Institute & Department of Psychiatry and Behavioral Sciences, University of California, Davis, CA, 95817, USA

[18]Virtual VCFS Center, Syracuse, NY, 13206, USA

[19]School of Psychology, University of Newcastle, Newcastle, 2258, Australia

[20]Department of Medical Genetics, Aix-Marseille University, Marseille, 13284, France

[21]Genomics of Health and UDMGC, Son Espases University Hospital, Balearic Islands Health Research Institute (IDISBA), Palma de Mallorca, 07120, Spain

[22]Institute of Medical and Molecular Genetics (INGEMM), University Hospital La Paz, Madrid, 28046, Spain

[23]International 22q11.2 Brain and Behavior Consortium: Table S1

[24]Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles, Los Angeles, CA, 90095, USA

[25]Department of Psychiatry and Psychology, Maastricht University, Maastricht, 6200 MD, the Netherlands

[26]Program in Genetics and Genome Biology, Research Institute, and Department of Psychiatry, The Hospital for Sick Children, Toronto, Ontario, Canada; Department of Psychiatry, University of Toronto, Toronto, Ontario, M5S 1A1, Canada; Department of Psychiatry, University Medical Center Utrecht Brain Center, Utrecht, 3584 CG, the Netherlands

[27]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

[28]Department of Psychiatry, Perelman School of Medicine of the University of Pennsylvania, and Children's Hospital of Philadelphia, Philadelphia, 19104, USA

[29]Dalglish Family 22q Clinic, Clinical Genetics Research Program and Department of Psychiatry, Toronto General Hospital, Centre for Addiction and Mental Health, and the University of Toronto, Toronto, M5T 1L8, Canada

[30]Division of Cardiology, Children's Hospital of Philadelphia and Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, 19104, USA

[31]Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, 10461, USA

***For correspondence:** Bernice.Morrow@einstein.yu.edu

Abstract:

The 22q11.2 deletion syndrome (22q11.2DS) results from non-allelic homologous recombination between low copy repeats termed LCR22. About 60-70% of patients with the typical 3 Mb deletion from LCR22A-D have congenital heart disease, mostly of the conotruncal type (CTD), while others have normal cardiac anatomy. In this study, we tested whether variants in the hemizygous LCR22A-D region are associated with risk for CTDs based upon sequence of the 22q11.2 region from 1,053 22q11.2DS subjects. 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 (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 cases without 22q11.2DS. One of the replicated variants, rs178252, is located in an open chromatin region and resides in the double-elite enhancer, GH22J020947, 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, implicating this gene as a modifier. Rs178252 and rs6004160 are eQTLs of *CRKL*. Further, set-based tests identified an enhancer predicted to target *CRKL* that is significantly associated with CTD risk (GH22J020946, SKAT test $P= 7.21\times10^{-5}$) in the 22q11.2DS cohort. These findings suggest that variance in CTD penetrance in the 22q11.2DS population may be explained in part by variants affecting *CRKL* expression.

Introduction

The vast majority of 22q11.2 deletion syndrome (22q11.2DS [MIM: 192430]) patients have a 3 million base pair (Mb) hemizygous deletion of chromosome 22q11.2 [1]. This syndrome is the most frequent chromosomal microdeletion disorder, occurring in 1/4,000 live births [2; 3] and 1/1,000 fetuses [4; 5]. The 22q11.2DS results from *de novo* non-allelic homologous recombination events between four low copy repeats (LCR) termed LCR22A, B, C and D [6; 7]. Over 85% of affected individuals carry a 3 Mb hemizygous deletion between LCR22A-D [8]. However, nested proximal (LCR22A-B, 1.5 Mb, 5%; LCR22A-C, 2 Mb, 2%; [6; 7]) and distal (LCR22B-D, 1.5 Mb, 4%; LCR22C-D 0.7 Mb,1%) deletions are present in some individuals with 22q11.2DS [1]. In patients with the LCR22A-D and proximal nested LCR22A-B and LCR22A-C deletions, the prevalence of congenital heart disease (CHD) is approximately 65% [9; 10]. A somewhat lower prevalence (~32%) is observed in patients 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-B region, *TBX1* (*T-box 1* [MIM: 602054]), encoding 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, while inactivation of both alleles resulted in a PTA and perinatal lethality [16-18]. Furthermore, missense variants in *TBX1* have been found in certain non-deleted individuals that 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]. Of interest, a genetic interaction was observed between *Tbx1* and *Crkl* in mouse models, suggesting that they may participate in the same functional pathway during embryogenesis [26].

In contrast to the prevalence of CHD at 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 likely due to the existence of additional genetic and/or environmental modifiers. Identification of modifiers may provide insight into the biological mechanism of heart development and disease. While some insights of the genetic architecture of CHD in 22q11.2DS have been gained by array genotyping or 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 there is only one functional copy of genes that are present. To test the hypothesis that common and/or rare single nucleotide variants (SNVs; including small indels) on the remaining allele may be associated with CHD, we used WGS data from 1,053 22q11.2DS subjects, all with the same typical 3 Mb LCR22A-D deletion. We performed a case-control association study using 22q11.2 subjects with CHD or subtypes within, such as CTDs and controls with 22q11.2DS but with a normal heart and/or aortic arch. Further, to determine whether associations identified in the 22q11.2DS cohort were also observed in CTD cases from the general population, we interrogated 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]. Briefly, subjects with a known 22q11.2 deletion, existing DNA samples, and approval by institutional research ethics boards (Albert Einstein College of Medicine; Committee of 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 termed CHD cases. 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 bicuspid aortic valve, were considered as controls. For the CTD subset, any of the following cardiac defects were considered as CTD cases in the present study: TOF, PTA, IAAB, RAA or abnormal origin of the right or left subclavian artery. The difference between CHD and CTD cases was that CTD cases excluded isolated VSD or ASD, but both had LCR22A-D deletions. CTD cases can be separated into two different groups based upon differences in embryological origin, which are 1) cardiac OFT defects that include TOF, PTA, PS and/or PA and 2) aortic arch defects or arterial branching defects from the aortic arch that 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]. Briefly, samples were sequenced using the Illumina HiSeq X Ten for the first 100 samples and the Illumina HiSeq 2500 platform for all other samples at Hudson Alpha Corp. (Huntsville, AL). Sequence reads were mapped to genome build hg38 (December 2013; GRCh38/hg38) with PEMapper (90% stringency for 2x100bp reads and 95% stringency for 2 x 150bp reads; [37]). Deletion sizes were confirmed by the coverage at the 22q11.2 region. Variants on the remaining 22q11.2 allele (LCR22A-D 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 of relatives and mismatched gender or samples with poor-quality sequence were removed. For all samples with the LCR22A-D 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 based on the dataset of genome-wide diploid variants as well as Hapmap 3 r3 (International HapMap project Phase III Release 3) data using plink/1.90b. First, shared variants in this dataset and the Hapmap 3 r3 dataset were extracted and combined into one dataset. Of note, variants with A>T, T>A, G>C and C>G allele types were removed to avoid any potential strand flip issues. Second, variants with minor allele frequency < 0.05 and variants in sex chromosome were excluded. After this, autosomal common variants were pruned using the –indep function to ensure only independent variants were used for PCA. Lastly, PCA was conducted using the –pca function. European Caucasian ancestry of the subjects was determined by the Multidimensional Outlier Detection method as

implemented in SVS Golden Helix software. Firstly, a median centroid vector was calculated as [median (column1), median (column2), median (column3)] based on top 3 PCs for all the samples plus Hapmap Caucasian (CEU) and Tuscans in Italy (TSI) 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 is calculated as follows:

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

Where Q3 and IQR are the third quartile and inner quartile range of each sample (1…N) and M is a user-specified multiplier; in this study, 2 was adopted. Outliers of Caucasian samples were examined in the scatter plot of PC1 versus PC2; samples clustered with HapMap Gujarati Indians in Houston, Texas (GIH) and Mexican ancestry in Los Angeles, California (MEX) population were grouped as Hispanics. Populations dispersed towards Yoruban in Ibadan and Nigeria (YRI) Africans were grouped as African-admixed populations. PCA was further conducted in the three subpopulations respectively, in order to obtain the top several PCs that can be used as covariants to 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 the Caucasian, African-admixed, and Hispanic subsets, respectively, with adjustment of sex and corresponding number of PCs for CHD, CTD, TOF and TOF-PTA-IAAB risk. False discovery rate (FDR) was employed to correct for multiple testing issues. For rare variants, we conducted the Fisher's exact test in the Caucasian population. Considering most genome-wide association studies 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. SNV based analyses were done using plink/1.90 [38]. For the top variants in the LCR22C-D region with evidence of association with CTD risk in the 22q11.2DS cohort, we interrogated existing data from a meta-analysis of three published genome wide association studies (GWAS) 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 using Bystro [39], snpEff [40] and dbNSFP [41]. We adopted the definition of loss of function (LoF) variants as previously described [42]. Briefly, LoF variants included those with 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 predictions for LoF variants by Bystro and snpEff were included. As an ensemble annotation tool, Bystro also has additional annotations for possible function, including 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 prediction as “damaging” or “deleterious” or passing the suggested threshold for scores by at least half of the 29 algorithms compiled in dbNSFP for SNVs (including SIFT [44], SIFT4G [45], Polyphen2-HDIV [46], Polyphen2-HVAR [46], LRT [47], MutationTaster2 [48] and additional algorithms [41]). Probabilities of SNVs affecting splicing were annotated by ada and random forest scores by dbscSNV (database for splicing consensus variants), which is a companion database in dbNSFP (database for functional prediction for non-synonymous variants). 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 to 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 that follows priority of exon > UTR > intron > intergenic and the RefSeq exonic allele function following a priority of indel-frameshift = start loss = stop gain = stop loss > indel-nonFrameshift > nonsynonymous > synonymous.

LINSIGHT measures potential non-coding variants that have effects on fitness based on primate evolution [49]. The LINSIGHT score of top associated variants in the LCR22C-D 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

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

Set-based analyses

Set-based analyses were conducted for common and rare variants, respectively. Sets in the LCR22A-D region including RefSeq genes between the TSS and transcriptional termination site (TTS) end as well as both 2 kb upstream and downstream, promoter regions (both 2 kb upstream and 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). There are 72 genes in the 3 Mb region 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 *P* value was set at 2.1×10^{-4}. Suggestive significant threshold was set at $P=1.0\times10^{-3}$. We used the burden test [35] to evaluate if there were any rare variants in any set that were enriched and we tested whether the effects are in the same direction. We used SKAT to test for common variants to consider 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 with 22q11.2DS and WGS data, who had a 3 Mb LCR22A-D deletion (Figure 1A and Table S2). A total of 14,158 SNVs within the LCR22A-D 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 is 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-D 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, while 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 for logistic regression analyses of common variants among all subjects and stratified analyses for the three subpopulations, respectively.

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, for simplification, they are referred to herein as one group, named common variants throughout this study.

The study design is presented in Figure 3. Briefly, 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 and splicing variants. Then different analytical strategies were applied to the three groups of variant categories that will be described in detail in the next sections.

SNV-based common variant analyses

We performed an association study of common variants, comparing 584 CHD cases to 469 controls with 22q11.2DS and found suggestive evidence for association with several

variants in the LCR22C-D region (Figure S2A). We repeated these analyses, restricting the cases to include only those with CTDs (N=424) and identified significant associations in the same region. The association signal was in a cluster of 62 SNVs in a 350 kb region largely within LCR22C-D (chr22: 20607741-2095814; hg38; Figure 4A and 4B) that are in strong LD (Figure 4C). The SNVs have an AAF >0.05 (Figure 4D), with *P* values ranging from 10^{-3} to 2.59×10^{-5}, of which 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), respectively. These variants were not significantly associated with CTDs in the African-admixed or Hispanic population, possibly because the sample sizes were very small (Figure S2B-C). Of note, none of the SNVs in the coding and non-coding regions of *TBX1* showed association with risk for CTDs (all *P*>0.05) in any subpopulation (Figure 4B and S2B-C).

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-D (Figure S2D-E). In fact, adding 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.

We next tested for evidence whether association differs within subgroups, with the caveats that group sizes vary. We first compared the subgroup of TOF cases with the same controls and then the subgroup that included TOF, PTA or IAAB, to controls, in the Caucasian subpopulation. Variants with a *P* value $<1.0\times10^{-3}$ for all of the categories were compiled that included variants occurring from two or more subgroups, with the smallest *P* value, totaling 69

variants in LCR22C-D region Table S4). We did not observe a difference in the association test for the TOF or 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, while the aortic arch and arterial branches develop from the pharyngeal arch arteries containing vascular endothelium that is mesoderm derived as well as smooth muscle cells that are neural crest derived [56; 57]. The CTD category consists of subjects with cardiac OFT defects and/or aortic arch defects. Because of their different embryological origins, we compared the two. As shown in Figure S3, the group with cardiac OFT defects (Figure S3A) contributes much more to the strength of the association signal in the 350 kb region, with the top variants passing FDR for multiple testing correction, as compared to the group with aortic arch defects (Figure S3B). This result supports the findings that these have different developmental and anatomical origins, therefore their genetic control may also be distinct.

Set-based analysis for common variants and odds ratios

We performed a set-based SKAT test to determine the genetic risk of genes, promoters and enhancers focusing on CTDs in the 22q11.2DS population. Results from the set-based test for common variants is presented in Figure 5A and Table S5. We found 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).

Double elite enhancers have functional genomic evidence for regulatory activity and for particular gene targets [53]. There are eleven double elite enhancers in the 350 kb region (GH22J020748, GH22J020775, GH22J020855, GH22J020866, GH22J020883, GH22J020916, GH22J020936, GH22J020939, GH22J020940, GH22J020946, GH22J020947) as indicated in

Table S6. These double elite enhancers have not been assayed as of yet, for function in human or mouse cardiac development but are of interest as harboring 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 OR as well as the corresponding 95% CIs were plotted in this 350 kb region (Figure 5D-E). Most of the variants were associated with increased odds of CTDs with a median OR of 2.96, ranging from 1.64 to 4.75, while alternate alleles of 3 variants were associated with an OR less than 1 (OR range: 0.48 from 0.52) (Table S4).

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 non-deleted individuals [34]. Four of these variants had meta-analysis *P* values <0.05 and the direction of association was the same as 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*, based on the GeneHancer database (Figure 5B). Moreover, the SNV, rs178252, is one of two variants among the 69 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-D region

One of the challenges is to find possible causal variants in the associated 350 kb region because of the presence of a large LD block. To identify possible functional variants, CADD, LINSIGHT and phastCons scores were generated for the 69 SNVs. Detailed associations of the 69 variants are presented in Table S4. Of note, 12 of the SNVs are eQTLs of *CRKL* as identified in GTEx data (Muscle-Skeletal), suggesting that they may be functional. Two SNVs reside in an open chromatin region based upon data from human stem cells[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 similar AAFs in both datasets, indicating the high quality of our data and that subsets of the variants have similar AAFs in either dataset, which 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 cases, seven in the CTD subset and ten in controls, with no enrichment in CTD cases, 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, it is less likely 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-D interval for CTDs is driven by rare variants in the same region. Due to issues related to population stratification for rare variant association analysis, we focused on the largest sized, 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 cases and controls was not significantly different (Figure S6). Moreover, there was no significant difference between CTD cases versus controls for the joint effect of rare variants when aggregated into genes, promoter and enhancer regions based on 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 less likely that rare variants account for the significant association found for common variants of CTD risk in LCR22C-D. Therefore, the association signals of common variants are less likely 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-D 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. Based upon complete sequence data, this association was not driven by rare variants individually or jointly in the same region, suggesting that associated common variants are amongst 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 resulting in haploinsufficiency of different genes. While early data suggested that the LCR22A-B region was the critical region for the syndrome [6; 7; 58-60], patients were identified recently with non-overlapping LCR22B-D and C-D deletions [11; 13]. Despite having non-overlapping deletions, these individuals had a similar CTDs with half the frequency [14]. The existence of patients with LCR22B-D or C-D deletions and CTDs, coupled with the data presented in this report, support the likelihood that the LCR22C-D region contains modifiers of CTD risk. Individuals with the LCR22A-C deletion have a similar frequency of CTDs as in the LCR22A-B or LCR22A-D region. Thus, we suggest that *CRKL* acts as a risk factor, but that the 22q11.2 region is quite complex and understanding 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-D region. Since 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 did not indicate that there were cardiovascular anomalies present [63]. In studies in zebrafish, knockdown 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 if 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, they 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* 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 to 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 suggests 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 they genetically interact during embryogenesis [26]. The connection between these genes is based upon the hypothesis that *Tbx1* acts upstream and *Crkl* acts downstream in a FGF8 (Fibroblast growth factor 8 [MIM: 600483]) through 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* may 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 were identified that included CTDs [14].

Strengths and limitations in examination of sequence in the 22q11.2 allele

The availability of WGS is a strength of the study, since 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 may 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 may affect expression levels of *CRKL.* Another strength lies the availability of summary *P* value statistics from non-deleted subjects with CTDs that provided supportive data. 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-D and C-D deletions have CHD at 20-30%, suggest 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-D region on the remaining allele of 22q11.2 is associated with the risk for CTDs in subjects with 22q11.2DS. Haploinsufficiency of this region alone is associated with CTDs, and when taken

together with mouse genetic studies, implicate, mostly plausibly, *CRKL* as a possible target of non-coding putative regulatory variants.

Description of Supplemental Data

Supplemental Data include seven figures and six tables.

Declaration of Interests

The authors declare none conflicts of interest exist.

Data availability

All data used for this publication is presented in Tables S2 and S3.

Acknowledgements

We would like to thank families with 22q11.2DS who provided DNA and clinical information for this study. We acknowledge the Genomics and Molecular Cytogenetics Cores at Einstein. 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 never be possible. Dr. Morrow was supported by NIH R01HL132577, R01HL084410, U01MH101720, U54HD090260 and P01HD070454. Other funding sources are detailed in Supplemental Information.

Web Resources

OMIM, https://www.omim.org/

SVS Golden Helix software,

http://doc.goldenhelix.com/SVS/latest/svsmanual/numeric_data_quality.html

Hapmap 3 r3 data, ftp://ftp.ncbi.nlm.nih.gov/hapmap/

UCSC genome browser, https://genome.ucsc.edu/

GTEx, https://gtexportal.org/home

gnomAD, https://gnomad.broadinstitute.org/

plink, https://www.cog-genomics.org/plink/1.9/

UCSC Genome Browser, http://genome.ucsc.edu/

Bystro, https://bystro.io/

dbNSFP, https://sites.google.com/site/jpopgen/dbNSFP

SnpEff, http://snpeff.sourceforge.net/

GeneHancer database, https://www.genecards.org/

GENCODE, https://www.encodeproject.org/)

FANTOM5, http://fantom.gsc.riken.jp/5/

Ensembl Regulatory Build,

https://useast.ensembl.org/info/genome/funcgen/regulatory_build.html

EPDnew, https://epd.epfl.ch//index.php

DbSUPER, http://asntech.org/dbsuper/

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, 101-107.
3. Oskarsdottir, 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., Garcia-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., et al. (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., et al. (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 U S A 103, 7729-7734.
22. Torres-Juan, L., Rosell, J., Morla, M., Vidal-Pou, C., Garcia-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. (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. (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. (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. (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., Hawula, 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, C. (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. (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 Pluss et al.: The strength of PEMapper/PECaller lies in unbiased calling using large sample sizes. Proc Natl Acad Sci U S A 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., et al. (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 le, 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. (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, 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, W452-457.
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. (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., et al. (2017). Architecture of the human PI4KIIIalpha lipid kinase complex. Proc Natl Acad Sci U S A 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., et al. (2012). PtdIns4P synthesis by PI4KIIIalpha 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., Pavao, 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., Yardimci, 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.

Figure legends:

Figure 1. The 22q11.2 deletion syndrome cohort. (A) Illustrative map of the 22q11.2 region depicting the locations of LCR22A, B, C and D (box) and the 3 Mb deleted region (gray box) downstream from 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 subjects that comprise 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 abnormal of the right or left subclavian artery in the absence of other cardiac or aortic arch anomalies.

Figure 2. Ethnicity and constituency of variants in the 1,053 22q11.2DS cohort. (A) Scatter plot 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 Northern and Western European ancestry ; 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), Yoruban in Ibadan, Nigeria (West Africa). (B) Frequency distribution of 14,158 variants that passed QC measures. There are in total 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).

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.

Figure 4. Regional plot of logistic regression analyses identifies significant association to a 350 kb LCR22C-D interval. (A, B) Regional plot for common variants versus CTD risk in 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). Grey blocks represent LCR22A, B, C and D in the 22q11.2 region. Blue horizontal lines represent 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 Genome Assembly, hg38). Red bar indicates LD block. Variants in red survived false discovery rate (FDR) for multiple testing correction. (C) LD structure in the Caucasian population based on 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.

Figure 5. Set-based SKAT test for common variants identifies non-coding variants in gene regulatory regions in the 350 kb LCR22C-D region. (A) SKAT test with adjustment of sex and top four PCs was applied to common variants in the cohort of 669 Caucasian samples. 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) between LCR22A-D, 72 putative promoters of these genes (green; both 2 kb upstream and downstream of TSS) and 96 curated double elite set of enhancers (red) in the 22q11.2 region. 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 significant gene (*SERPIND1*) and enhancer (GH22J020946). Four gray blocks represent LCR22A, B, C

and D. (B) Snapshot of the UCSC Genome Browser in hg38 assembly showing the genomic context in the 350 kb top associated region including Segmental Duplication track (LCR22C), RefSeq genes, H3K4Me3 (promoter) and H2K27Ac (enhancer) marks from cell lines, double elite set 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 there are five double elite enhancers regulating *CRKL* (GH22J020947, GH22J020946, GH22J020940, GH22J020939, GH22J020936) indicated below the regulatory region interactions. (C) Gene alignment in the top associated region based on RefSeq gene track in UCSC genome browser. (D, E) Distribution of ORs of 69 top associated variants ($P < 1.0\times10^{-3}$); 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. Enlarged image is below.

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 main studied CTD case-control 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%)	
TOF-PTA-IAAB	284 (26.9%)	
TOF	194 (18.4%)	
ASD alone	55 (5.2%)	
VSD alone	105 (10.0%)	
Population origin		
Caucasian	790 (75.2%)	669 (74.9%), 312:357 cc[c]
AA[b] and admixed	161 (15.3%)	135 (15.1%), 68:67cc
Hispanic	102 (9.7%)	89 (10.0%), 44:45 cc

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

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 subjects	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 cc[d]	0.720	1.67(1.24-2.25)	6.90E-4	0.393	1.62(1.23-2.14)	6.03E-4	0.939	1.65(1.25-2.18)	3.66E-4	0.392	1.64(1.24-2.16)	4.80E-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 full model with inherited genotype modeled as an additive effect and maternal genotype unrestricted, to the model with just the unrestricted maternal genotype. [c]meta-analyses of three CTD cohorts without 22q11.2 deletion. [d]cc, case-control. [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]

Figure 1

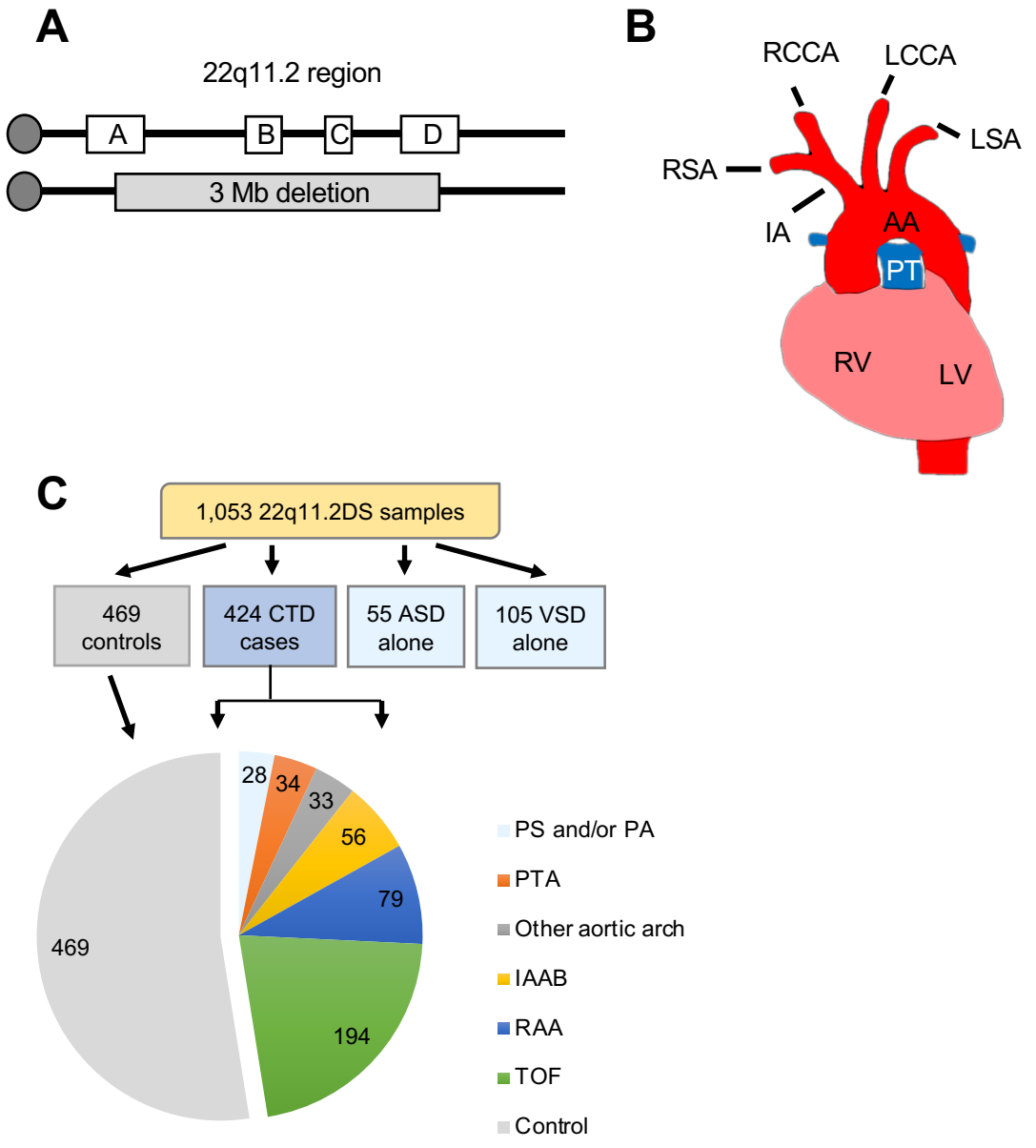
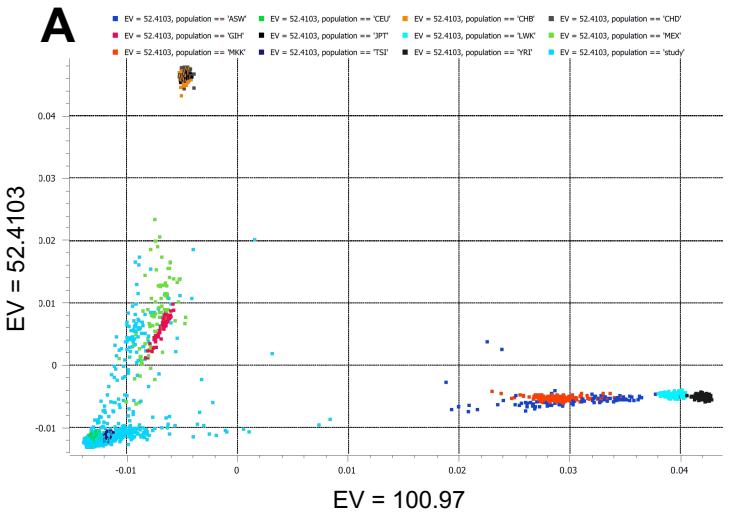

A
22q11.2 region
A
B
C
D
3 Mb deletion
B
RCCA
LCCA
LSA
RSA
IA
AA
PT
RV
LV
C
1,053 22q11.2DS samples
469 controls
424 CTD cases
55 ASD alone
105 VSD alone
28
34
33
56
79
194
469
PS and/or PA
PTA
Other aortic arch
IAAB
RAA
TOF
Control

Figure 2


A
EV = 52.4103, population == 'ASW'
EV = 52.4103, population == 'CEU'
EV = 52.4103, population == 'CHB'
EV = 52.4103, population == 'CHD'
EV = 52.4103, population == 'GIH'
EV = 52.4103, population == 'JPT'
EV = 52.4103, population == 'LWK'
EV = 52.4103, population == 'MEX'
EV = 52.4103, population == 'MKK'
EV = 52.4103, population == 'TSI'
EV = 52.4103, population == 'YRI'
EV = 52.4103, population == 'study'
0.04
0.03
0.02
0.01
0
-0.01
EV = 52.4103
-0.01
0
0.01
0.02
0.03
0.04
EV = 100.97

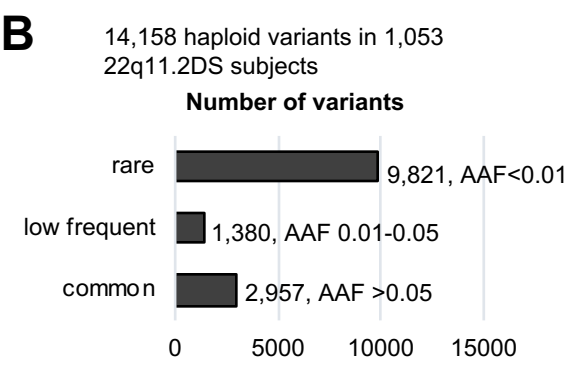

B
14,158 haploid variants in 1,053
22q11.2DS subjects
Number of variants
rare
9,821, AAF<0.01
low frequent
1,380, AAF 0.01-0.05
common
2,957, AAF >0.05
0
5000
10000
15000

Figure 3

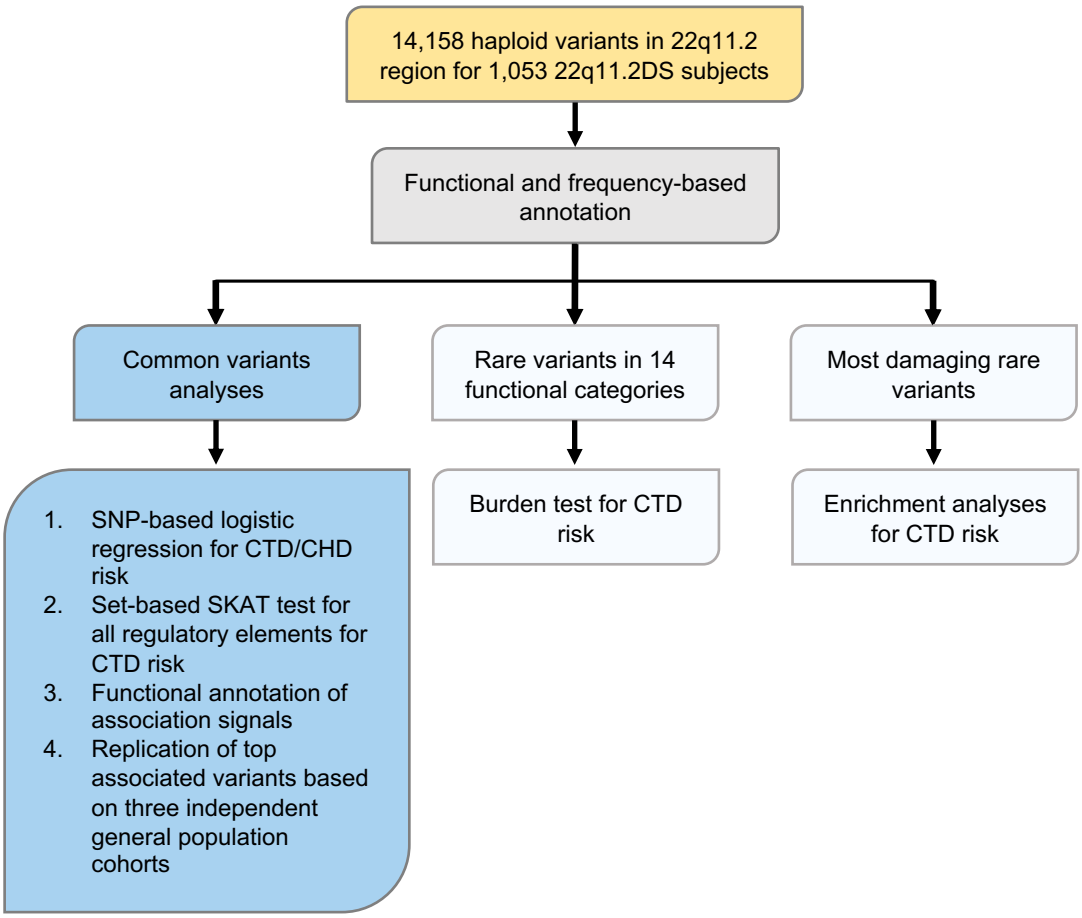

14,158 haploid variants in 22q11.2 region for 1,053 22q11.2DS subjects
Functional and frequency-based annotation
Common variants analyses
Rare variants in 14 functional categories
Most damaging rare variants
1. SNP-based logistic regression for CTD/CHD risk
2. Set-based SKAT test for all regulatory elements for CTD risk
3. Functional annotation of association signals
4. Replication of top associated variants based on three independent general population cohorts
Burden test for CTD risk
Enrichment analyses for CTD risk

Figure 4

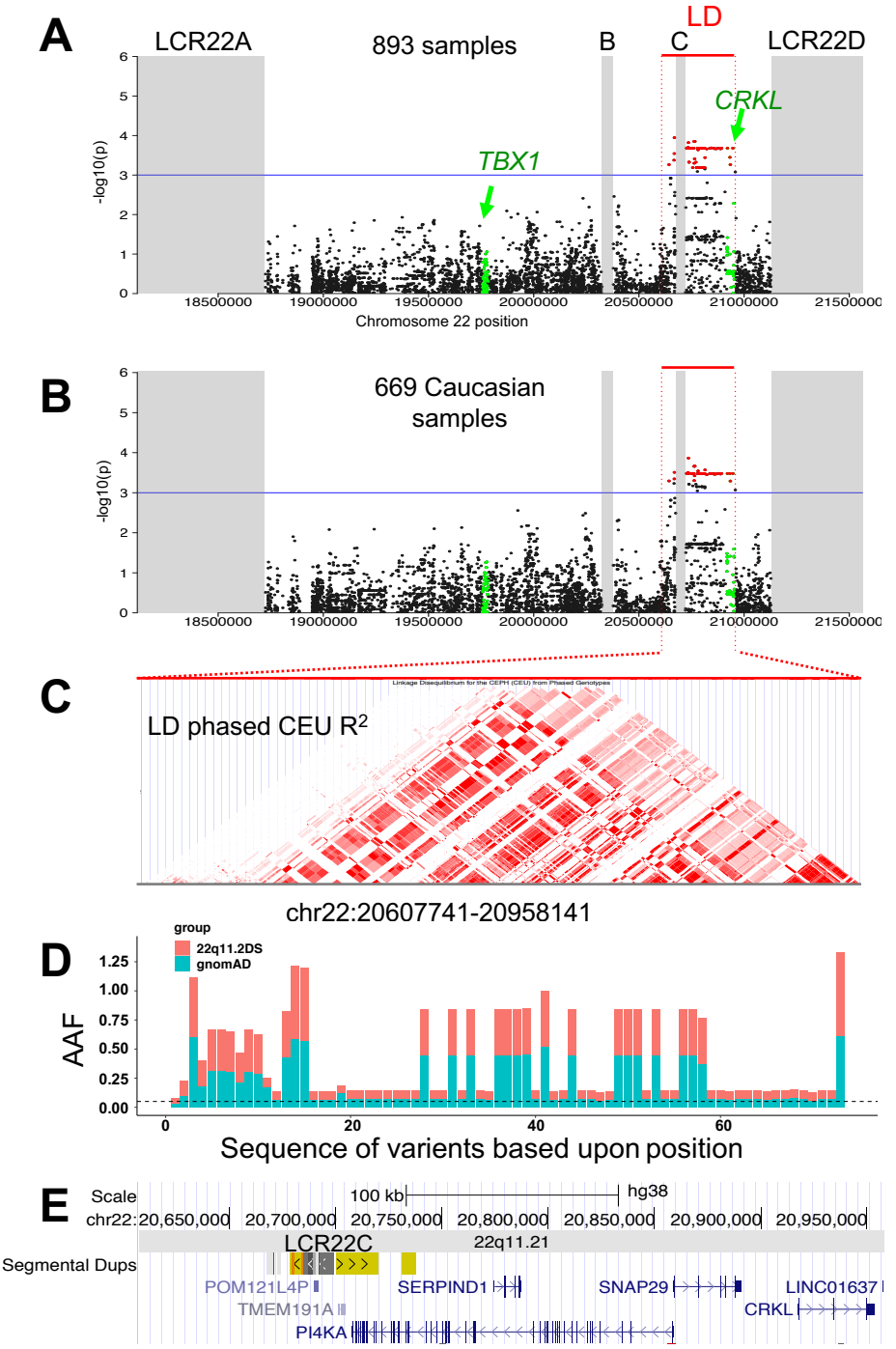
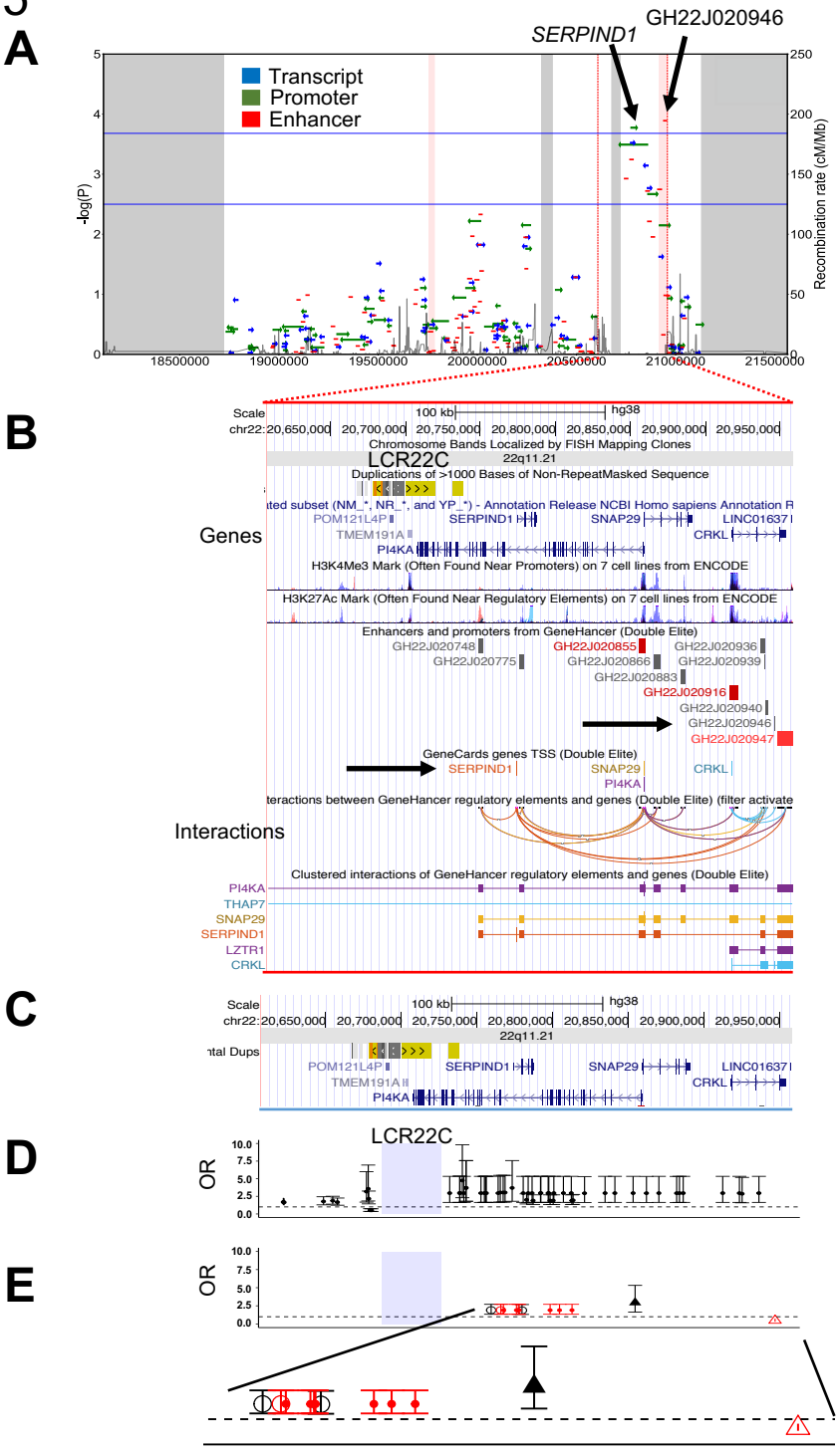

A
LCR22A
893 samples
B
C
LD
LCR22D
CRKL
TBX1
-log10(p)
Chromosome 22 position
B
669 Caucasian samples
C
LD phased CEU R2
chr22:20607741-20958141
D
group
22q11.2DS
gnomAD
AAF
Sequence of varients based upon position
E
Scale
100 kb
hg38
chr22:20,650,000
20,700,000
20,750,000
20,800,000
20,850,000
20,900,000
20,950,000
LCR22C
22q11.21
Segmental Dups
POM121L4P
SERPIND1
SNAP29
LINC01637
TMEM191A
CRKL
PI4KA

Figure 5


A
SERPIND1
GH22J020946
Transcript
Promoter
Enhancer
-log(P)
Recombination rate (cM/Mb)
18500000
19000000
19500000
20000000
20500000
21000000
21500000
B
Scale
100 kb
hg38
chr22:20,650,000
20,700,000
20,750,000
20,800,000
20,850,000
20,900,000
20,950,000
Chromosome Bands Localized by FISH Mapping Clones
LCR22C
22q11.21
Duplications of >1000 Bases of Non-RepeatMasked Sequence
Genes
POM121L4P
SERPIND1
SNAP29
LINC01637
TMEM191A
CRKL
PI4KA
H3K4Me3 Mark (Often Found Near Promoters) on 7 cell lines from ENCODE
H3K27Ac Mark (Often Found Near Regulatory Elements) on 7 cell lines from ENCODE
Enhancers and promoters from GeneHancer (Double Elite)
GH22J020748
GH22J020855
GH22J020936
GH22J020775
GH22J020866
GH22J020939
GH22J020883
GH22J020916
GH22J020940
GH22J020946
GH22J020947
GeneCards genes TSS (Double Elite)
Interactions
Clustered interactions of GeneHancer regulatory elements and genes (Double Elite)
PI4KA
THAP7
SNAP29
SERPIND1
LZTR1
CRKL
C
D
LCR22C
OR
E
OR