Deletion size and background genetic variation shape congenital heart disease phenotypes in 3,016 individuals with 22q11.2 deletion syndrome

Authors:

Jhih-Rong Lin[1], Daniella Miller[1], Dana Luong[1], Tanner Nelson[1], T. Blaine Crowley[2], Oanh T. Tran[2], Bhooma Thiruvahindrapuram[3], Amirhossein Hajianpour[3], Linda Campbell[4], Tiffany Busa[5], Damian Heine-Suñer[6], Sixto García-Miñaúr[7], Luis Fernández[7], Kieran C. Murphy[8], Declan Murphy[9], Wanda Hawula[10], Kathleen Angkustsiri[11], Vandana Shashi[12], Kelly Schoch[13], Carrie E. Bearden[13], Aoy Tomita Mitchell[14], Michael E. Mitchell[14], Miri Carmel[15], Abraham Weizman[15], Elena Michaelovsky[16], Doron Gothelf[16], Marianne B.M. van den Bree[17], Michael J. Owen[17], Jacob A.S. Vorstman[18], Erik Boot[19,20], Claudia Vingerhoets[20], Therese van Amelsvoort[20], Ann Swillen[21], Jeroen Breckpot[21], Joris R. Vermeesch[21], Koen Devriendt[21], Maude Schneider[22], Stephan Eliez[23], M. Cristina Digilio[24], Marta Unolt[24,25], Carolina Putotto[26], Bruno Marino[27], Maria Pontillo[28], Marco Armando[29], Stefano Vicari[30,31], Gabriela M. Repetto[32], Wendy R. Kates[33], Robert J. Shprintzen[34], Raquel E. Gur[35], ,Elaine H. Zackai[2], Elizabeth Goldmuntz[36], Tao Wang[37], Srilakshmi Raj[1], Beverly S. Emanuel[2], Donna M. McDonald-McGinn[2], Stephen C. Scherer[3,38], Anne S. Bassett[22,39,40], Zhengdong D. Zhang[1*], Bernice E. Morrow[1,41,42]*

[1]Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, 10461, USA.

[2]Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, 19104, USA; Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, 19104, USA.

[3]Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON Canada; The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON Canada

[4]School of Psychology, University of Newcastle, Newcastle 2258, Australia.

[5]Genetics Service, University Hospital of Nimes, Nimes, 30900, France.

[6]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.

[7]Institute of Medical and Molecular Genetics, University Hospital La Paz, Madrid 28046, Spain.

[8]Department of Psychiatry, Royal College of Surgeons in Ireland, Dublin D02 YN77 , Ireland.

[9]Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE5 8AZ, UK.

[10]Department of Genetics, Polish Mother's Memorial Hospital-Research Institute, 93-338 Lodz, Poland; wanda.hawula@iczmp.edu.pl (W.H.);

[11]Developmental Behavioral Pediatrics, UC Davis MIND Institute, University of California, Davis, CA 95817, USA.

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

[13]Departments of Psychiatry and Biobehavioral Sciences and Psychology, Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles, Los Angeles, CA, 90095, USA.

[14]Department of Surgery, Medical College of Wisconsin, Milwaukee, WI 53226, USA

[15]Felsenstein Medical Research Center, Gray Faculty of Medicine and Health Sciences, Tel Aviv University and Felsenstein Medical Research Center

[16]The Department of Child & Adolescent Psychiatry, Edmond and Lily Sapfra Children's Hospital, Sheba Medical Center and Gray Faculty of Medicine and Health Sciences and Sagol School of Neuroscience, Tel Aviv University, Ramat Gan, 5262000, Israel

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

[18]Department of Psychiatry and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, M5G 0A4, Canada

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

[20]The Dalglish Family 22q Clinic, Toronto General Hospital, University Health Network, Toronto, ON, Canada.

[21]Center for Human Genetics, University Hospital Leuven, Department of Human Genetics, University of Leuven (KU Leuven), Leuven 3000, Belgium.

[22]Clinical Psychology Unit for Intellectual and Developmental Disabilities, Faculty of Psychology and Educational Sciences, University of Geneva, Switzerland

[23]Developmental Imaging and Psychopathology Laboratory, Fondation Pôle Autisme, University of Geneva School of Medicine, Geneva, Switzerland; Department of Genetic Medicine and Development, University of Geneva School of Medicine, Geneva, Switzerland; Synapsy Center for Neuroscience and Mental Health Research, University of Geneva School of Medicine, Geneva, Switzerland.

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

[25]Department of Pediatric Cardiology, Bambino Gesù Hospital, Rome, 00165, Italy.

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

[27]Child and Adolescence Neuropsychiatry Unit, Department of Neuroscience, IRCSS Bambino Gesù Children's Hospital of Rome, Rome, Italy.

[28]Department of Life Sciences and Public Health, Catholic University and Child & Adolescent Psychiatry Unit at Bambino Gesù Hospital, Rome, 00165, Italy.

[29]Division of Child and Adolescent Psychiatry, Department of Psychiatry, University Hospital of Vaud Canton and University of Lausanne, CH-1015 Lausanne Switzerland

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

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

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

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

[34]The Virtual Center for Velo-Cardio-Facial Syndrome, Inc, Syracuse, NY, 13104, USA.

[35]Department of Psychiatry, Perelman School of Medicine of the University of Pennsylvania Philadelphia, PA 19104, USA; Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

[36]Division of Cardiology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

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

[38]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada and McLaughlin Centre and Department of Molecular Genetics, University of Toronto.

[39]Clinical Genetics Research Program and Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada

[40]Department of Psychiatry, University of Toronto, Toronto, Ontario, M5T 1R8, Canada.

[41]Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY 10461, USA

[42]Department of Obstetrics & Gynecology and Family Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

*Corresponding authors:

zhengdong.zhang@einsteinmed.edu, bernice.morrow@einsteinmed.edu

ABSTRACT

Congenital heart disease (CHD) occurs in over half of individuals with 22q11.2 deletion syndrome (22q11.2DS) and the types of lesions range from mild to severe. To determine the basis of variation in cardiac phenotypes we analyzed demographic data from 3,016 unrelated individuals with 22q11.2DS from centers in the Northeast US, Canada, Europe, South America, Israel and Australia. Most individuals in this cohort had a 3 million base pair hemizygous deletion between low copy repeat, LCR22 A-D (87.2%), while a minority had nested proximal deletions (1.5 Mb, A-B, 5.4%; 1.8 Mb, A-C, 1.8%). By univariable analysis methods, we found several demographic features contributed to phenotypic variation, which may in part be due to collection site heterogeneity. To disentangle these effects, we performed multivariable mixed-effects logistic regression and uncovered significant differences between CHD phenotypes and these features. Individuals with the A-D deletion had a lower risk of persistent truncus arteriosus (OR = 0.37, 95% CI 0.18-0.75) but a higher risk of septal defects (OR = 4.7, 95% CI 1.7-12.8) compared to those with the smaller A-B deletion, suggesting distinct developmental pathways sensitive to 22q11.2 gene dosage. In addition, genome-wide genetic principal components (PCs) were associated with specific CHD subtypes, including reduced risk of pulmonary stenosis or atresia with other heart lesions with increasing PC2 (OR = 0.73, 95% CI 0.61-0.87) and increased risk of abnormal origin of the subclavian arteries with PC4 (OR = 2.6, 95% CI 1.4-4.9), indicating that background genetic variation may modify lesion-specific susceptibility. Together, these results suggest that both deletion size and background genetic variation shape the highly variable cardiac phenotypes in 22q11.2DS.

INTRODUCTION

The chromosome 22q11.2 deletion syndrome (22q11.2DS; DiGeorge/velo-cardio-facial/conotruncal anomaly face syndrome) is a congenital anomaly disorder that occurs in 1/1,000 fetuses [1;2] and 1/4,000 live births [3,4] constituting the most frequent genomic disorder in humans. This syndrome results from non-allelic homologous recombination events between flanking low copy repeats on chromosome 22q11.2 termed LCR22 [5-8] resulting in *de novo* deletions of different sizes (**Figure 1**). Molecular genetic testing such as multiplex-ligation dependent probe amplification (MLPA) or chromosomal microarray is typically used for diagnostic testing of patients with clinical features reminiscent of 22q11.2DS. The most frequent deletion is a 3 million base pair (Mb) hemizygous deletion flanked by LCR22 A-D (A-D). Nested A-B (1.5 Mb) and A-C (1.8 Mb) deletions occur with lower frequency but with similar overall phenotypes [7-10].

The clinical presentation of this syndrome ranges from mild to severe including characteristic facial differences, immune deficits, and hypocalcemia as well as neurocognitive or psychiatric disorders [11,12]. One of the most frequent and serious medical findings is congenital heart disease (CHD) that occurs in 50-80% of individuals with 22q11.2DS [13]. The types of CHD that occur range from isolated, mild aortic arch or arterial branching defects (right sided aortic arch, RAA; abnormal origin of the right or left subclavian arteries, ASA) with little clinical symptoms, ventricular and/or atrial septal defects (septal defects, SD; VSD, ASD) to severe, complex and lethal conotruncal heart defects (CTDs) including tetralogy of Fallot (TOF), interrupted aortic arch type B (IAAB) and persistent truncus arteriosus (PTA) [14,15]. These severe CTDs require surgery for survival beyond the neonatal or early infancy period. Overall, there is extensive phenotypic heterogeneity in the prevalence and types of CHD in 22q11.2DS. The role of demographic features on phenotypic heterogeneity of CHD is not well established due to the rarity of the syndrome and lack of a comprehensive and systematic assessment of basic medical data.

In this study, we retrospectively analyzed basic demographic information and detailed cardiac phenotypes from 3,016 unrelated individuals with 22q11.2DS. They were ascertained at clinical centers that specialize in care of individuals of 22q11.2DS in the US, Canada, Europe, South America and Australia. This large cohort size makes it possible to identify significant but subtle risks or protection for CHD types. Univariable analysis identified different points of heterogeneity of the data. This necessitated multivariable analyses to adjust for confounding factors in the analysis. This multivariable approach revealed that both deletion size and genome-wide genetic principal components are associated with specific CHD phenotypes.

METHODS

Ethics approval statement

Our research on 22q11.2 disorders is part of an Internal Review Board approved program at Albert Einstein College of Medicine (IRB #1999-201). Individuals ascertained at Albert Einstein College of Medicine or Montefiore Medical Center obtained written informed consent (and/or assent for minors) for all participants as part of this IRB protocol. Co-authors at different sites have their own individual institutional IRB approval with their own informed consent procedures. This includes, Children's Hospital of Philadelphia (IRB 07-5352), University of Pennsylvania (IRB #832678), University of California Los Angeles (IRB#10-001071-CR-00012), Medical College of Wisconsin (MCW PRO 47811; v. 2025.07.30), Duke University (IRB Pro00010602), University of Toronto affiliated hospitals (IRB CAMH 114-2001, CAMH 154-2002, CAMH 150-2002, CAMH 151-2002 and UHN 98-E156), Sapienza (Protocol Ref. 5803) and Rome, IT (Prot. 468/12), University of Geneva (IRB Project ID: 2020-02296), Geneva Ethic Committee (Marco Armando). Further, this study was approved by ethics committees in the UK, The Netherlands (Utrecht; 08-345 Vorstman) and Belgium (Cardiff, NHS Wales Research Ethics Committee; Maastricht, Maastricht University Medical Centre-IRB NL70681.068.19; KU Leuven; Research Medical Ethics Committee UZ KU Leuven, IRB S52418), University of Newcastle, Newcastle,

AU, Son Espases University Hospital, Mallorca-IRB# IB 4084/20, University Hospital La Paz, Madrid, Royal College of Surgeons in Ireland, Dublin, Ireland, King's College London, London, Polish Mother's Memorial Hospital-Research Institute, Tel Aviv University-Sheba Medical Center Israel (IRS#SMC-7783-10). The study was approved by the local ethics committee at the Institute of Psychiatry, South London and Maudsley Trust (067/00; L. Campbell). All individuals were de-identified for this study.

Study population

De-identified demographic data were evaluated from 3,746 unrelated individuals with 22q11.2DS recruited from clinical centers for 22q11.2DS in the US, Canada, Northern and Southern Europe, Chile and Australia, including and in addition to the International 22q11.2 Brain and Behavior Consortium [16] (Full de-identified dataset is presented in **Supplementary Table S1**). Each had a clinical diagnosis of 22q11.2DS (velo-cardio-facial syndrome-MIM#192430, DiGeorge syndrome-MIM#188400, conotruncal anomaly face syndrome). The type of the 22q11.2 deletion was determined using molecular diagnostic testing including chromosomal microarray, multiplex ligation-dependent probe amplification (MLPA; (SALSA MLPA kit P250 DiGeorge; MRC Holland, The Netherlands [17]) or research based whole exome/genome sequencing [18-20]. We included deletions between LCR22 A-B, A-C and A-D in this study. Rare subjects with nested distal deletions such as B-D and C-D deletions were excluded. Subjects with other deletion types on 22q11.2, or those extending centromeric or telomeric to this region, were extremely rare and were also excluded. Both types of deletions (B-D, C-D, or atypical deletions) were excluded due to the paucity of these samples.

Most individuals with 22q11.2DS were unrelated probands and had a *de novo* deletion. Transmitting parents and relatives with 22q11.2DS were excluded from this study for consistency. A total of 3,016 subjects fulfilled the criteria for the study and had cardiology and echocardiography reports as infants and young children. Babies and children were included,

provided there was a clinical cardiac record of lesions at birth or the presence of a normal heart. Adults were included when there were pediatric records of cardiac evaluation including echocardiography that were available. We did not include prenatal cases. We obtained sex and self-reported race and/or ethnicity (race/ethnicity) for most and recruitment site information. We were missing data on sex in less than 0.5% and self-identified race/ethnicity in about 15%. A portion of these data in a subset of the 3,016 subjects was previously described [18-20].

Association between sex or self-reported race/ethnicity and deletion type

Whole genome sequence (WGS) was available for 1,868 of the 3,016 individuals included in the study (below). Exclusion of first-degree degree kinship among these 1,868 subjects was confirmed using KING software[21]. For those with WGS, we could identify the genetic sex and those with mismatches were already excluded. For the 1,148 without WGS data, we used patient identified sex for the analysis. Proband sex (male, female) was compared across 22q11.2DS subjects using a 2×3 contingency table. Each proband was classified into one of three 22q11.2 deletion types: A-B, A-C, and A-D. Independence between sex and deletion-type group was evaluated using the Pearson's χ^2 test. Statistical significance was defined as adjusted $P < 0.05$ for multiple comparisons. Self-reported race/ethnicity was defined as White, Black or African American, Asian, American Indian or Alaska Native, Natie Hawaiian or Other Pacific Islander, Mixed race or Other (not belonging to any of the other categories). Mixed ethnicity is considered a combination of more than one race/ethnicity such as White with Black or Asian. We were missing self-reported race/ethnicity for 15% of the cohort. Self-reported race/ethnicity analyses were restricted to the two largest self-reported race/ethnicity categories in the cohort, "White" and "Mixed". A 2×3 contingency table of deletion size by race/ethnicity was constructed and data were analyzed as described for sex and deletion size.

Co-occurrence of cardiac phenotypes in 22q11.2DS

The types of cardiac lesions that were identified in this cohort were organized to best parse individual and combined, complex lesions. For simplicity, the cardiac phenotypes were segregated into classes (Normal, CHD is any heart defect, CTD is any conotruncal defect minus SD, SD includes ASD-only, VSD-only and ASD+VSD) and among the classes they were parsed to phenotype subgroups (TOF, SD, RAA+ that is RAA with/out other defects, VSD-only is isolated VSD, IAAB, RAA-only that is RAA without other defects, PTA, ASA without other defects, PS/PA+ is PS or PA not part of TOF but can be with other defects, ASD-only is isolated ASD, ASD+VSD is ASD with VSD but no other defects, PS/PA-only is PS or PA but no other defects). The phenotypes are described in more detail in **Supplementary Table S2**. For visualization of the CHD types, we paired each lesion with data retained containing non-missing values and a 2×2 contingency table was generated. Pairs in which two traits were observed in ≥ 1 subject and absent in ≥ 1 subject (i.e., all marginal totals > 0) were considered testable. Association was quantified with Fisher's exact test (two-tailed), yielding an odds ratio (OR) and *P* value. False-discovery-rate adjustment across all 91 testable pairs was done using the Benjamini–Hochberg procedure (FDR *q*). To visualize the strength and direction of association, the signed score$_{ij}$ = $\mathrm{sign}\left(\log_2 OR_{ij}\right) \times \left(-\log_{10} q_{ij}\right)$ was calculated for each pair, giving positive values for co-occurrence (OR > 1; blue) and negative values for mutual exclusivity (OR < 1; red). Scores were assembled into a symmetric matrix in phenotype order, capped at the most extreme finite values, and zeros assigned to non-testable pairs. Rows and columns were hierarchically clustered (Euclidean distance, complete linkage; R package pheatmap). https://cran.r-project.org/package=pheatmap.

Principal component analysis following removal of batch effects in whole genotype sequence

We derived the top five principal components (PCs) from principal component analysis (PCA) to capture variance in population structure or other common genetic features, based on common single nucleotide variants (SNPs, minor allele frequency, MAF > 0.05) from 1,868 subjects with available short-read whole genome sequence (WGS; Batch 1 = International Brain and Behavior Consortium-NIMH Data Archive-NDA64; Batch 2 = dbGAP; NIH Center for Inherited Disease Research; phs002514.v1.p1). PLINK was then used to select 300,384 linkage disequilibrium (LD)-pruned common DNA variants on the autosomes with "indep-pairwise" function (window size = 50 SNPs, step size = 10, r^2 threshold = 0.1) and performed PCA. A total of 1,868 22q11.2DS subjects were processed using the DRAGEN unified variant calling pipeline [22]. The cohort was sequenced on two different sequencing platforms. About half the samples were sequenced at the Center of Inherited Disease Research (phs002514.v1.p1) and the other half was sequenced as part of the International 22q11.2 Brain and Behavior Consortium (IBBC; NIMH Data Archive 22q11.2DS). Both sets of data are publicly available. These technical differences were suggested by the top PCs (**Supplementary Figure S1A**) derived from the raw sequencing data. These platform differences can introduce batch effects and biases due to technical but not true genetic differences. To resolve this issue of technical batch affects, we performed a variant quality control process (QC) before PCA aimed at removing the observed batch effects. Specifically, we set genotypes with depth (DP) < 5 or genotype quality (GQ) < 10 to missing in the project-level VCF (pVCF) for the WGS data. We then excluded variants with a missingness rate > 10%, in addition to variants with a QUAL score < 30. Finally, we used PLINK to select 300,067 linkage disequilibrium (LD)-pruned common variants on the autosomes with "indep-pairwise" function (window size = 50 SNPs, step size = 10, r^2 threshold = 0.1) and performed PCA. After QC, the batch effect previously observed in PC3 was no longer present, and none of the top five PCs used in our analysis were associated with WGS platform status (**Supplementary Figure S1B, S1C,** and **S1D**).

Ancestry analysis

The global ancestry proportions, or proportions of genetic ancestry largely derived from continental superpopulations, of each individual were estimated using ADMIXTURE software (v1.3.0)[23]. First, PLINK (v1.90b6.21) was used to remove SNPs and individuals with >2% missingness (--geno 0.02, –mind 0.02), SNPs with low minor allele frequency (--maf 0.05), SNPs that violated Hardy-Weinberg equilibrium (--hwe 1e-10), individuals with extreme heterozygosity rates (>3SD from mean), and SNPs in high linkage disequilibrium (--indep-pairwise 50 5 0.2). After applying these filters, 1,797 individuals and ~4.2 million SNPs remained. An initial ADMIXTURE run with k=5 (5 different populations) was used to identify reference individuals with more than 90% ancestry from a single continent. The cohort was merged with the 1000 Genomes reference[24], resulting in ~250k variants and 3,215 individuals. ADMIXTURE was rerun with k=5 to determine global ancestry estimates for the cohort. Global ancestry estimates were visualized using pong (v1.5).

Univariable screening

Each CHD phenotype class and subgroup was first explored in relation to one explanatory variable at a time. For the categorical variables (recruitment site, sex, self-reported race/ethnicity, deletion type), a χ^2 test of independence provided a global *P* value for the phenotype-predictor association. In the same data subset, a crude (unadjusted) logistic-regression model of the form $\text{logit}\{\, P(phenotype = 1)\,\} = \beta_0 + \sum_{j=1}^{J-1} \beta_j \, 1(predictor = j)$ was fitted. Wald statistics generated odds-ratio (OR) estimates with 95% Wald confidence intervals (CI) for each non-reference category. For principal components (PC1-PC5, which together accounted for 74% of the genetic variance, **Supplementary Figure S1**), every phenotype was regressed simultaneously on the five Z-standardized PC scores, and a likelihood-ratio test compared this full model with an intercept-only model to provide a joint "PC1-5" *P* value. For each variable, we used the Benjamini-

Hochberg procedure for false-discovery-rate control and $q < 0.05$ for statistical significance. Analyses were carried out in R with glm, broom, and lmtest.

Multivariable mixed-effects logistic regression

For each binary phenotype group, we fitted a separate mixed-effects logistic-regression model with a logit link. Fixed effects were the 22q11.2 deletion category (reference: recurrent A-B, contrasts: recurrent A-D and A-C), sex, and the first five principal components (PC1-PC5, previously Z-scaled). Recruitment site (23 sites) was modeled as a random intercept to account for center-level differences in diagnosis and ascertainment, stabilize estimates for sites with small sample number, and avoid overfitting that would arise if each site were treated as a fixed covariate:

$$\text{logit}\, P(Y_{ij} = 1) = \beta_0 + \beta_{\text{del}} + \beta_{\text{sex}} + \sum_{k=1}^{5} \beta_k \text{PC}_{k,ij} + u_j, \quad u_j \sim N(0, \sigma^2_{\text{site}})$$

Models were fitted with lme4::glmer (optimizer = bobyqa) in R.

Wald z-tests provided a two-sided *P* value for every non-intercept coefficient. We first controlled the false-discovery rate (FDR) within each phenotype (column *q* in Supplementary Tables) and then applied a second Benjamini-Hochberg adjustment across the entire set of 112 fixed-effect estimates (14 phenotypes × 8 predictors), reported as *q*_global. Results with *q*_global < 0.05 were considered study-wide significant.

Bootstrap sensitivity analysis

To assess the robustness of key associations identified in the mixed-effects logistic regression models, we performed a two-stage nonparametric bootstrap procedure that preserved the hierarchical structure of the data. First, recruitment sites were resampled with replacement. Second, within each resampled site, individuals were resampled with replacement. For each

bootstrap replicate, we refit the same mixed-effects logistic regression model used in the primary analysis, including deletion group, sex, and ancestry principal components as fixed effects and study site as a random intercept. We repeated this procedure for 500 iterations for selected significant associations (pulmonary stenosis/atresia with/out other defects excluding TOF-PS/PA+ with PC2; persistent truncus arteriosus-PTA and septal defects-SDs with deletion group A-D; abnormal origin of the subclavian arteries-ASA with PC4). From the bootstrap distributions of the fixed-effect coefficients, we estimated median effects, empirical 95% confidence intervals (2.5th-97.5th percentiles), and the proportion of replicates with effect estimates in the same direction as the original model. This approach evaluates the stability of the findings while accounting for clustering by recruitment site.

RESULTS

Demographic features in 3,016 subjects with 22q11.2DS

A retrospective cohort study was performed on 3,016 subjects with 22q11.2DS to determine if we can identify contributing factors to variation of CHD phenotypes at birth or infancy. The 22q11.2 region and LCR22s with deletion types (A-B, A-C and A-D) with genes included for context is illustrated in **Figure 1** with details of the demographic features collected on each subject in **Supplementary Table S1**. Most of the individuals had a 3Mb, A-D deletion (n=2,788, 87.2%) while a small subset had nested A-B (n=172, 5.4%) and fewer had A-C deletions (n=56, 1.8%) as shown in **Figure 2A**. There was missing deletion type data from 4% of the cohort (**Figure 2A**). These subjects were included because they were tested by clinical fluorescence *in situ* hybridization with probes in the A-B region (**Figure 1**). The proportions of deletion types were similar to what has been previously described for 22q11.2DS [6-8,10,25], but with greater overall numbers of subjects for each deletion type in this cohort to increase the resolution of the analysis.[26,27] We found that half the individuals of this cohort were reported as male and half as

female, which were confirmed in 1,868 subjects with WGS data available (**Figure 2B**). The majority of subjects were self-reported as White (n=2124, 66.4%), while the next largest group was of self-reported Mixed race/ethnicity (n=393, 12.3%). Fewer were Black, Asian or 'Other' categories of different self-reported ancestries (**Figure 2C**). Self-reported race/ethnicity was missing in 15% of the total set (**Figure 2C**). Subjects were recruited from 23 different clinical centers for 22q11.2DS (**Figure 2D**). The largest number of subjects were recruited from the Northeast US followed by Toronto CA, and then other sites in the US, Europe, South America and Australia (**Figure 2D**).

Association analysis between deletion type with sex and self-reported race/ethnicity

For the first analysis, reported sex (including genetically validated sex from WGS) of each individual was compared across the A-B, A-C and A-D deletion types. We found that the proportion of males was similar across groups (**Supplementary Table S3**). Pearson's χ^2 test confirmed that the distribution of sex did not vary by deletion type (χ^2 = 1, df = 2, P = 0.61). We then performed an association analysis between self-reported race/ethnicity and deletion types. We did not find a significant difference in the distribution within the two largest race/ethnicity groups consisting of White and Mixed-ethnicity individuals showed a virtually identical pattern (**Supplementary Table S4**). The other groups were too small in size to do this comparison. Pearson's χ^2 test indicated no evidence that deletion-size distribution differs between the two race/ethnicity groups (χ^2 = 2.3, df = 2, P = 0.32). Thus, in this cohort, the type of 22q11.2 deletion is not associated with self-reported sex and race/ethnicity.

Ancestry in the 22q11.2DS cohort from WGS

To more precisely identify genetic ancestry in the 22q11.2DS population, the 1000 Genomes project data was used as a reference[24]. The results for the 1000 Genomes project are shown in

the pong plot and segmented to different ancestries as shown in **Figure 3A**. The software ADMIXTURE was then used on the 22q11.2DS cohort with available WGS to estimate global ancestry proportions, using the 1000 Genomes data. We found significant representations of different continental ancestries in our sample with the majority being of European ancestry (**Figure 3B**). This is consistent with the origin of most of the samples, but it provides for the first time a quantitative measure of the proportions of continental ancestries, and degrees of admixture, represented in this large sized cohort.

Architecture of CHD types in 22q11.2DS

In this cohort of 3,016 individuals, we found that 59% had a type of CHD (n=1895/3016). Most individuals with CHD had complex defects comprising combinations of lesions. In order to count each individual subject once for further analysis, we generated two main classes of CHD types defined in **Supplementary Table S2** that are: CTDs (n=1,193/3,016; 40% of total or 67% of CHD; **Figure 4**), and ASD and/or VSD (septal defects; SD; 424/3,016; 14% of the total of 23% of CHD; **Figure 4**). The 10% not belonging to these two classes had isolated aortic arch branching alone or pulmonic stenosis or atresia (PS/PA) without other defects or other rare lesions that weren't assigned to a class. Many individuals had more than one lesion (**Supplementary Table S1**). Phenotype subgroups were generated from the CHD types (**Supplementary Table S2**), according to the main and most severe anatomical defect such as the lesion requiring the earliest intervention for survival [28]. We generated 11 phenotype subgroups (**Figure 4**) and compared the phenotype network relationships between the two classes of CHD and phenotype subgroups as shown in **Figure 5**. For example, a given individual with TOF had a CTD, but this person could not have both TOF and a PTA, because they are mutually exclusive and cannot co-occur (illustrated in **Figure 5**; pair-wise analysis; **Supplementary Table S5**). Together, the heatmap and network shown in **Figure 5** provide a coherent architecture of CHD in 22q11.2DS except for other isolated defects comprising three

subgroups. Other isolated defects included abnormal branching of the left or right subclavian arteries, right sided aortic arch (RAA) and PS/PA, when added together was a total of 181 subjects (**Figure 4**). Additionally, these defects frequently co-occurred with CTDs or when co-occurring with a VSD, they are considered a CTD. Of interest, we found that RAA showed enrichment with TOF (OR = 2.50, q = 1.2E-13) and PTA (OR = 1.98, q = 3.5E-3). RAA is rare in the general population especially when they co-occur with CTDs[14,29]. Therefore, we can infer that the presence of RAA with CTDs further implicates a diagnosis of 22q11.2DS, Overall, these two classes and 11 subgroups were used for our further analyses.

Deletion size and genetic background contribute to heterogeneity in CHD risk

Preliminary univariable screening analyses indicated that the prevalence of CHD classes and phenotype subgroups varied across recruitment sites and showed associations with sex, deletion size, and genetic principal components for selected lesions (**Supplementary Tables S6-S10**). In particular, strong site-to-site heterogeneity for CTD and other lesions suggested that simple unadjusted models were insufficient to disentangle correlated demographic and genetic effects. These observations motivated the use of a multivariable mixed-effects framework to jointly model covariates while accounting for recruitment-site heterogeneity.

To understand how different factors jointly influence the occurrence of CHD types in 22q11.2DS, we applied the multivariable mixed-effects logistic regression method. This approach models each class and phenotype subgroup as the outcome while simultaneously adjusting for deletion size, sex, and population structure (PC1-PC5), with recruitment site included as a random effect. This approach accounted for both measured covariates and unmeasured differences across sites, such as diagnostic practices or ascertainment patterns, and provided more reliable estimates of independent effects than univariable models.

Across 14 CHD outcomes and 126 fixed-effect contrasts, four associations remained significant after stringent study-wide false discovery rate control (q_global < 0.05; **Table 1, Supplementary Table S11**). These involved a subset of CHD phenotypes associated with 22q11.2DS.

Table 1. Significant fixed-effect coefficients in the multivariable mixed-effects logistic regression.

CHD	*Predictor*	*OR (95% CI)*	*P-value*	*q_global*
PTA	Deletion A-D vs A-B	0.37 (0.18 - 0.75)	5.74E-3	4.52E-2
SD	Deletion A-D vs A-B	4.68 (1.71 - 12.84)	2.70E-3	2.43E-2
PS/PA+	PC2	0.73 (0.61 - 0.87)	5.35E-4	5.19E-3
ASA	PC4	2.59 (1.37 - 4.89)	3.37E-3	2.83E-2

Notes: The statistical significance is set at the global FDR threshold (*q*_global) < 0.05. PTA, persistent truncus arteriosus; SD, septal defect including atrial or ventricular septal defect; PS/PA+, pulmonary stenosis or atresia with other heart lesions excluding TOF; ASA, abnormal origin of the subclavian arteries. PC, principal component. OR, odds ratio.

Deletion size emerged as the dominant determinant for selected CHD subtypes. Carriers of the largest (A-D) deletion had roughly one-third the risk of having a PTA defect compared to carriers of the smaller A-B deletion (**Table 1**; OR = 0.37, 95% CI: 0.18-0.75, q_global = 0.042). A-B was the reference deletion size, and each of A-D and A-C deletion sizes was compared with it (but not between A-D and A-C). Only the A-D deletion showed a significant association, suggesting that genes within the B-D interval may exert a protective effect against this severe CTDs. Conversely, the A-D deletion increased the risk of combined septal defects more than four-fold (**Table 1**; OR = 4.68, 95% CI: 1.71-12.84, q_global = 0.023). Thus, while the A-D deletion appears to reduce susceptibility to PTA, it simultaneously increases risk of SDs (VSD, ASD), indicating that deletion size has divergent effects across developmental pathways. In two-stage bootstrap sensitivity analyses that resampled recruitment sites and individuals within sites, the

direction of effect for A-D deletion was consistently preserved for both PTA (median OR = 0.37; P[OR < 1] = 0.90) and SD (median OR = 5.2; P[OR > 1] = 0.99), supporting the stability of these associations despite variability in resampled datasets.

We observed a significant association between pulmonary stenosis/atresia with/out other defects such as TOF (PS/PA+) and PC2, whereby each one-standard-deviation increase along PC2 was associated with a 27% reduction in risk (OR = 0.73, 95% CI: 0.61-0.87, q_global = 5.2E-3). Bootstrap analyses supported the stability of the PC2 effect on PS/PA+, with the resampled OR similar to the original estimate and approximately 92% of bootstrap samples retaining an OR < 1. We also identified an association between abnormal origin of the subclavian arteries (ASA) and PC4 (OR = 2.59, 95% CI: 1.37-4.89, q_global = 0.028). Bootstrap analyses showed a median OR of approximately 3.1, with ~96% of bootstrap samples yielding OR > 1, although the confidence intervals were wide, indicating greater variability in this estimate. Sensitivity analyses using ADMIXTURE-derived ancestry proportions did not identify statistically significant direct associations with these CHD subtypes. Together, these findings suggest that principal components capturing genome-wide variation are associated with selected cardiac phenotypes, while direct ancestry proportion estimates did not show independent effects in this cohort.

DISCUSSION

In this report, we demonstrate that deletion size at chromosome 22q11.2 , but not sex, influences risk for specific cardiac lesions in 22q11.2DS. Individuals carrying the common 3Mb A-D deletion had a reduced risk of PTA but an increased risk of SDs compared with individuals carrying the smaller nested 1.5Mb A-B deletion. These opposing effects suggest that haploinsufficiency of genes within the B-D region differentially modifies susceptibility to distinct cardiac phenotypes. In addition, principal components reflecting genome-wide variation were

associated with pulmonic stenosis/atresia with/out other defects excluding TOF (PS/PA+) and abnormal origin of the right or left subclavian arteries (ASA). Direct analyses using ADMIXTURE-derived ancestry proportions did not identify significant independent effects. These findings suggest that broader genetic background may contribute to lesion-specific variability[26], though confirmation in more diverse cohorts will be necessary. The large size of this cohort (n = 3,016) and the use of multivariable mixed-effects logistic regression allowed us to account for study site heterogeneity and other covariates, providing a more robust assessment of demographic influences on cardiac phenotype variability.

Clinical implications for cardiac phenotypes in 22q11.2DS. In this cohort, we found that 59% had CHD, which is similar to the range of 60-80% that has been reported[14], This variation in percentage is in part due to ascertainment biases (see limitations of the study, below). The main class of CHD types was CTDs including TOF, PTA and IAAB. These three CTDs as well as RAA or PS/PA with VSDs, are common defects in 22q11.2DS. They are highly suggestive of a diagnosis of 22q11.2DS as has been previously reported [15]; [1,29,30] but this larger cohort provides more detailed percentages for each. It is also known that certain CTDs including double outlet right ventricle (0.2%; 7/3016) or transposition of the great arteries (0.06%; 2/3016 are rarely seen in 22q11.2DS [1,15,29,30] and these were also extremely rare in this cohort . Similarly, lesions such as atrioventricular septal defects did not occur in any subjects with 22q11.2DS in our cohort but are common in individuals affected with Down syndrome.

The clinical implications of having a 22q11.2 deletion include the known risk for CHD that require immediately medical intervention in the newborn period [1,4,12,31,32]. There are also concerns because minor defects might not be detected on a medical exam due to the absence of clinical signs of a known lesion [33]. Specifically, RAA that co-occurs with abnormal branching of the left subclavian artery can cause a vascular ring resulting in constriction of the trachea and esophagus requiring surgical repair. In this report, we emphasize that isolated atrial or

ventricular septal defects, are relatively common in 22q11.2DS as well as for sporadic CHD, and therefore genetic testing should be considered in the presence of septal defects with syndromic features including mild craniofacial differences, developmental delay and other medical problems. Therefore, vigilant monitoring for CHD by echocardiography and cardiology evaluation through the pediatric period is required. Adults with the syndrome further require continued monitoring because of the co-occurrence of other medical problems that can affect outcomes, due to the deletion [13,14,29,34].

Genes that contribute to CHD types in 22q11.2DS. The most frequent class of defects in 22q11.2DS are CTDs; they occurred in 40% of this cohort. There are over 47 protein coding genes that are hemizygously deleted in the 3Mb A-D region, and there also are additional non-coding genes in this region. A subset of the coding genes is required for cardiac development based upon studies of mouse models, particularly for CTDs. These include *TBX1* (T-box transcription factor 1; A-B region), *CRKL* (CRK-like proto-oncogene adaptor; C-D region) *DGCR8* (DiGeorge syndrome critical region 8; A-B region) *SLC25A1* (Solute carrier family 25 member 1; mitochondria; A-B region) and *HIRA* (Histone cell cycle regulator; LCR22-D/distal). The most extensively studied is *TBX1*, which encodes a T-box containing transcription factor. Heterozygous mutations of *TBX1* in the absence of a deletion result in a phenocopy of 22q11.2DS [35-39]. Although the prevalence of *TBX1* mutations is rare in the general population, they are recurrent and affect protein structure or function [36,37,39-43]. Further, mouse models demonstrate its importance in cardiac development. Inactivation of one copy of the gene in the mouse results in mild lethality after birth with aortic arch anomalies (ASA and IAAB) at reduced penetrance [44]. However, inactivation of both copies results in a perinatal lethal PTA at full penetrance [44]. Mice with an A-B equivalent hemizygous deletion have increased lethality compared to *Tbx1* heterozygous mice and have aortic arch anomalies at higher penetrance [45]. Nonetheless, partial rescue of aortic arch anomalies occurred by crossing in mice

overexpressing *TBX1*, implicating *TBX1* as an important gene in the A-B region but that haploinsufficiency of multiple genes contribute to the overall phenotypes [46], [47]. The aortic arch anomalies that occur result from hypoplasia or absent 4th pharyngeal arch arteries in early embryos (4th PAA). The 4th PAA derives from the endothelium of the second heart field mesoderm and surrounding neural crest cells that form resulting smooth muscle [27,48,49]. The origin of PTA results from disruption of these two populations as well. The second heart field mesoderm is required to expand the cardiac outflow tract allowing for proper alignment and the neural crest cells are needed to migrate into the outflow tract to result in its septation [50,51]. This failure can be due to failed migration or function of neural crest cells into the outflow tract and/or failed signaling or function of adjacent second heart field mesoderm cells [50-52]. *Tbx1* is expressed and is required in different tissues in the pharyngeal apparatus including the second heart field mesoderm [53] and epithelia endoderm; [54] and ectoderm; [55,56] but it is not expressed in the neural crest cells and not significantly in the heart itself [57]. It is known to signal from the second heart field mesoderm and epithelia to direct neural crest cells to the PAAs and outflow tract [58,59].

The *CRKL* gene is in the C-D region and encodes a ubiquitously expressed cytoplasmic adaptor protein to cell surface receptor tyrosine kinase signaling. Inactivation of one copy of the *CRKL* gene does not affect cardiac development and this was uncovered about 25 years ago [60]. Inactivation of both alleles [60] or hypomorphs [61] of both alleles of *Crkl* in mice results in mid-gestational lethality with outflow tract alignment defects such as TOF or overriding aorta with a VSD among other additional intracardiac defects. Since *Tbx1* and *Crkl* affect fibroblast growth factor signaling mice that are heterozygous for *Tbx1* and *Crkl* were crossed and compound heterozygous knockout embryos had aortic arch defects at high penetrance, implicating that haploinsufficiency of these two genes might contribute to CHD in human patients with 22q11.2DS [62].

Two other genes of note are *SLC25A1* and *DGCR8*, both in the A-B region. The *SLC25A1* gene encodes a mitochondrial citrate transporter that can indirectly affect histone modification. Recently, it was discovered that inactivation of both alleles of the *Slc25a1* gene in the trophoblast in mouse models affects formation of the placenta and secondarily results in CHD with mid-gestational lethality and a severe VSD, abnormal alignment of the outflow tract and intracardiac muscle defects [63,64]. This sheds new light onto the fact that placental defects can secondarily affect embryonic development including heart formation and remodeling, that when defective causes of CHD in 22q11.2DS. The *DGCR8* gene encodes an essential micro-RNA processing enzyme that is ubiquitously expressed in all cells. Inactivation of mouse *Dgcr8* in the mesoderm results in early embryonic lethality but inactivation in the second heart field resulted in mid-gestational lethality with a PTA and thin ventricular walls with hemorrhage [65,66]. Inactivation of *Dgcr8* in neural crest cells results in perinatal lethality with a PTA with VSD [67]. Therefore, *Dgcr8* and perhaps micro-RNAs are required also for formation of the outflow tract/conotruncal region of the heart.

HIRA (Histone cell cycle regulator) encodes a multifunctional protein functioning as a chaperone to bring histones to DNA independent of replication. Inactivation of *Hira* in mice results in pre-implantation lethality and studies in chick showed that knockdown results in a PTA by affecting neural crest function [68]. Since its not deleted in most 22q11.2DS patients, its direct role in human patients is yet to be established. There are other genes in the 22q11.2 region that may contribute to CHD when hemizygously deleted, but they have not been directly implicated in cardiac development as of yet.

All the genes described above (*TBX1*, *SLC25A1*, *HIRA* and *DGCR8*) besides *CRKL* lie in the LCR22 A-B region, while *CRKL* is in the C-D region. It is possible that haploinsufficiency of these genes, together with those affecting basic homeostatic functions such as those involved in mitochondrial functions [69,70] affecting genetic networks of other genes in rest of the

genome, and these may serve as modifiers, potentially explaining differences in deletion type for various CHD phenotypes, but this is not proven as of yet.

Limitations of this study. Despite the large cohort size, several limitations should be considered. Participants were recruited at multiple clinical sites across the US, Europe, South America, and Australia. Although we observed ancestral heterogeneity within the cohort, the majority of individuals were of European ancestry. This imbalance may limit the generalizability of findings to more diverse populations. Broader representation will be important for evaluating potential population-specific modifiers in 22q11.2DS. Further, clinical ascertainment also varied across sites [15]. Some centers were pediatric tertiary care hospitals, whereas others were adult psychiatry outpatient services [14], resulting in differences in age at evaluation and referral patterns. Mild cardiac lesions, such as isolated aortic arch branching defects, might therefore have been under-detected at certain sites[14]. Nonetheless, all subjects were recruited from centers of clinical excellence for 22q11.2DS and awareness of the importance of cardiac evaluation. Finally, bicuspid aortic valve defect was not included in this study due to its high prevalence in the general population, although it has been reported in association with 22q11.2 deletions [71].

Conclusions: These findings demonstrate that deletion size and genome-wide genetic background contribute to variation in cardiac phenotypes in 22q11.2DS. The largest A–D deletion was associated with reduced risk of persistent truncus arteriosus but increased risk of septal defects, indicating that deletion size influences the spectrum of cardiac outcomes in divergent ways. In addition, principal components capturing genome-wide population structure were associated with selected CHD subtypes, although direct ancestry proportion estimates did not show independent effects. Together, these results suggest that both regional gene dosage within 22q11.2 and broader genetic background may influence lesion-specific susceptibility.

Further studies in more diverse cohorts will be required to clarify the biological mechanisms underlying these associations.

ACKNOWLEDGEMENTS

We warmly thank the families who participated in the study, and the family associations (Generation 22, Connect 22, and Relais 22) for their support. This work was supported by NIH grants P01HD070454, R01HL157157, R01GM125757 and U01MH101720 (BEM, ZDZ, REG, DMM, BSE, PIs). This work was also supported by R21HD060309 (AOM, MEM). This work was supported in part by the Canadian Institutes of Health Research (CIHR) [MOP-313331 and MOP-111238] and from the Inaugural Dalglish Chair in 22q11.2 Deletion Syndrome at the University of Toronto and University Health Network (ASB). The work was supported by ANID-Chile Fondecyt grants #1100131, 1130392, 1171014, 1211411 to GR. Grants from the Swiss National Science Foundation (#320030_179404 and #320030_144260 to SE and MS), The National Center of Competence in Research (NCCR) “Synapsy - The Synaptic Bases of Mental Diseases” (#51NF40-185897 to SE) and NeuroNA-Synapsy Clinical Cohort Grant (to SE and MS). Grants U01MH119736, R01 MH085953 and SFARI Explorer Award provided support (CB); as well as from KU Leuven (C14/24/135) and FWO grants G0A2622N and G0A7T26N (JRV, JB, AS). JASV is holder of the SickKids Psychiatry Associates Chair in Developmental Psychopathology and acknowledges funding support (U01MH101722).

CONFLICTS OF INTEREST

JASV serves as a consultant for NoBias Therapeutics Inc. and has received speaker fees for Henry Stewart Talks Ltd. DMM serves on the National Medical Advisory Board for Natera. The other authors declare no relevant conflicts of interest.

DATA AVAILABILITY

Subject demographic data is provided in the supplementary tables in the manuscript. WGS is available at the International Brain and Behavior Consortium-NIMH Data Archive-NDA64. Additional WGS is available at dbGAP; NIH Center for Inherited Disease Research; phs002514.v1.p1).

REFERENCES

1. Freud, L.R., Galloway, S., Crowley, T.B., Moldenhauer, J., Swillen, A., Breckpot, J., Borrell, A., Vora, N.L., Cuneo, B., Hoffman, H., et al. (2024). Prenatal vs postnatal diagnosis of 22q11.2 deletion syndrome: cardiac and noncardiac outcomes through 1 year of age. Am J Obstet Gynecol *230*, 368 e361–368 e312. 10.1016/j.ajog.2023.09.005.
2. 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. 10.1002/pd.4613.
3. 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. 10.1542/peds.112.1.101.
4. 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. 10.1038/nrdp.2015.71.
5. McDonald-McGinn, D.M., Tonnesen, M.K., Laufer-Cahana, A., Finucane, B., Driscoll, D.A., Emanuel, B.S., and Zackai, E.H. (2001). Phenotype of the 22q11.2 deletion in individuals identified through an affected relative: cast a wide FISHing net! Genet Med *3*, 23–29. 10.1097/00125817-200101000-00006.
6. 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. 10.1002/ajmg.1320570339.
7. 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. 10.1093/hmg/8.7.1157.
8. 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. 10.1093/hmg/9.4.489.
9. Rauch, A., Zink, S., Zweier, C., Thiel, C.T., Koch, A., Rauch, R., Lascorz, J., Huffmeier, U., Weyand, M., Singer, H., and Hofbeck, M. (2005). Systematic assessment of atypical deletions reveals genotype-phenotype correlation in 22q11.2. J Med Genet *42*, 871–876. 10.1136/jmg.2004.030619.

10. 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. 10.1086/302343.
11. Boot, E., Oskarsdottir, S., Loo, J.C.Y., Crowley, T.B., Orchanian-Cheff, A., Andrade, D.M., Arganbright, J.M., Castelein, R.M., Cserti-Gazdewich, C., de Reuver, S., et al. (2023). Updated clinical practice recommendations for managing adults with 22q11.2 deletion syndrome. Genet Med *25*, 100344. 10.1016/j.gim.2022.11.012.
12. Oskarsdottir, S., Boot, E., Crowley, T.B., Loo, J.C.Y., Arganbright, J.M., Armando, M., Baylis, A.L., Breetvelt, E.J., Castelein, R.M., Chadehumbe, M., et al. (2023). Updated clinical practice recommendations for managing children with 22q11.2 deletion syndrome. Genet Med *25*, 100338. 10.1016/j.gim.2022.11.006.
13. 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. 10.1002/ajmg.a.38662.
14. Goldmuntz, E. (2020). 22q11.2 deletion syndrome and congenital heart disease. Am J Med Genet C Semin Med Genet *184*, 64–72. 10.1002/ajmg.c.31774.
15. Momma, K. (2010). Cardiovascular anomalies associated with chromosome 22q11.2 deletion syndrome. Am J Cardiol *105*, 1617–1624. 10.1016/j.amjcard.2010.01.333.
16. 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. 10.1038/mp.2017.161.
17. Vorstman, J.A., Jalali, G.R., Rappaport, E.F., Hacker, A.M., Scott, C., and Emanuel, B.S. (2006). MLPA: a rapid, reliable, and sensitive method for detection and analysis of abnormalities of 22q. Hum Mutat *27*, 814–821. 10.1002/humu.20330.
18. 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. 10.1002/humu.21568.
19. 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. 10.1093/hmg/ddy028.
20. Zhao, Y., Guo, T., Fiksinski, A., Breetvelt, E., McDonald-McGinn, D.M., Crowley, T.B., Diacou, A., Schneider, M., Eliez, S., Swillen, A., et al. (2018). Variance of IQ is partially dependent on deletion type among 1,427 22q11.2 deletion syndrome subjects. Am J Med Genet A *176*, 2172–2181. 10.1002/ajmg.a.40359.
21. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics *26*, 2867–2873. 10.1093/bioinformatics/btq559.
22. Behera, S., Catreux, S., Rossi, M., Truong, S., Huang, Z., Ruehle, M., Visvanath, A., Parnaby, G., Roddey, C., Onuchic, V., et al. (2025). Comprehensive genome analysis and variant detection at scale using DRAGEN. Nat Biotechnol *43*, 1177–1191. 10.1038/s41587-024-02382-1.
23. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res *19*, 1655–1664. 10.1101/gr.094052.109.

24. Genomes Project, C., Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. 10.1038/nature15393.
25. Morrow, B., Goldberg, R., Carlson, C., Das Gupta, R., Sirotkin, H., Collins, J., Dunham, I., O'Donnell, H., Scambler, P., Shprintzen, R., and et al. (1995). Molecular definition of the 22q11 deletions in velo-cardio-facial syndrome. Am J Hum Genet *56*, 1391–1403.
26. Egbe, A., Uppu, S., Stroustrup, A., Lee, S., Ho, D., and Srivastava, S. (2014). Incidences and sociodemographics of specific congenital heart diseases in the United States of America: an evaluation of hospital discharge diagnoses. Pediatr Cardiol *35*, 975–982. 10.1007/s00246-014-0884-8.
27. Phillips, H.M., Stothard, C.A., Shaikh Qureshi, W.M., Kousa, A.I., Briones-Leon, J.A., Khasawneh, R.R., O'Loughlin, C., Sanders, R., Mazzotta, S., Dodds, R., et al. (2019). Pax9 is required for cardiovascular development and interacts with Tbx1 in the pharyngeal endoderm to control 4th pharyngeal arch artery morphogenesis. Development *146*. 10.1242/dev.177618.
28. Lindinger, A., Schwedler, G., and Hense, H.W. (2010). Prevalence of congenital heart defects in newborns in Germany: Results of the first registration year of the PAN Study (July 2006 to June 2007). Klin Padiatr *222*, 321–326. 10.1055/s-0030-1254155.
29. 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. 10.1007/s00246-013-0694-4.
30. Sauter, C., Hofbeck, M., Franz, P., Kettenstock, L., Steger, K., Linhardt, M., Reiter, A., Stork, S., Romanos, M., and Radtke, F. (2025). Congenital heart disease in 22q11.2 deletion syndrome: a meta-analysis and systematic review of the literature. J Med Genet *62*, 700–708. 10.1136/jmg-2025-110624.
31. Putotto, C., Pulvirenti, F., Pugnaloni, F., Isufi, I., Unolt, M., Anaclerio, S., Caputo, V., Bernardini, L., Messina, E., Moretti, C., et al. (2022). Clinical Risk Factors for Aortic Root Dilation in Patients with 22q11.2 Deletion Syndrome: A Longitudinal Single-Center Study. Genes (Basel) *13*. 10.3390/genes13122334.
32. Putotto, C., Pugnaloni, F., Unolt, M., Maiolo, S., Trezzi, M., Digilio, M.C., Cirillo, A., Limongelli, G., Marino, B., Calcagni, G., and Versacci, P. (2022). 22q11.2 Deletion Syndrome: Impact of Genetics in the Treatment of Conotruncal Heart Defects. Children (Basel) *9*. 10.3390/children9060772.
33. Putotto, C., Unolt, M., Lambiase, C., Marchetti, F., Anaclerio, S., Favoriti, A., Tancredi, G., Mastromoro, G., Pugnaloni, F., Liberati, N., et al. (2023). Cardiac function in adolescents and young adults with 22q11.2 deletion syndrome without congenital heart disease. Eur J Med Genet *66*, 104651. 10.1016/j.ejmg.2022.104651.
34. Ron, H.A., Crowley, T.B., Liu, Y., Unolt, M., Schindewolf, E., Moldenhauer, J., Rychik, J., Goldmuntz, E., Emanuel, B.S., Ryba, D., et al. (2022). Improved Outcomes in Patients with 22q11.2 Deletion Syndrome and Diagnosis of Interrupted Aortic Arch Prior to Birth Hospital Discharge, a Retrospective Study. Genes (Basel) *14*. 10.3390/genes14010062.
35. 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. 10.1136/jmg.38.12.e45.
36. Zweier, C., Sticht, H., Aydin-Yaylagul, I., Campbell, C.E., and Rauch, A. (2007). Human TBX1 missense mutations cause gain of function resulting in the same phenotype as 22q11.2 deletions. Am J Hum Genet *80*, 510–517. 10.1086/511993.

37. 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. 10.1016/s0140-6736(03)14632-6.
38. Zhang, M., Li, F.X., Liu, X.Y., Hou, J.Y., Ni, S.H., Wang, J., Zhao, C.M., Zhang, W., Kong, Y., Huang, R.T., et al. (2018). TBX1 loss-of-function mutation contributes to congenital conotruncal defects. Exp Ther Med *15*, 447–453. 10.3892/etm.2017.5362.
39. Pan, Y., Wang, Z.G., Liu, X.Y., Zhao, H., Zhou, N., Zheng, G.F., Qiu, X.B., Li, R.G., Yuan, F., Shi, H.Y., et al. (2015). A Novel TBX1 Loss-of-Function Mutation Associated with Congenital Heart Disease. Pediatr Cardiol *36*, 1400–1410. 10.1007/s00246-015-1173-x.
40. Xu, Y.J., Chen, S., Zhang, J., Fang, S.H., Guo, Q.Q., Wang, J., Fu, Q.H., Li, F., Xu, R., and Sun, K. (2014). Novel TBX1 loss-of-function mutation causes isolated conotruncal heart defects in Chinese patients without 22q11.2 deletion. BMC Med Genet *15*, 78. 10.1186/1471-2350-15-78.
41. Stoller, J.Z., and Epstein, J.A. (2005). Identification of a novel nuclear localization signal in Tbx1 that is deleted in DiGeorge syndrome patients harboring the 1223delC mutation. Hum Mol Genet *14*, 885–892. 10.1093/hmg/ddi081.
42. Castellanos, R., Xie, Q., Zheng, D., Cvekl, A., and Morrow, B.E. (2014). Mammalian TBX1 preferentially binds and regulates downstream targets via a tandem T-site repeat. PLoS One *9*, e95151. 10.1371/journal.pone.0095151.
43. Jiang, X., Li, T., Liu, S., Fu, Q., Li, F., Chen, S., Sun, K., Xu, R., and Xu, Y. (2021). Variants in a cis-regulatory element of TBX1 in conotruncal heart defect patients impair GATA6-mediated transactivation. Orphanet J Rare Dis *16*, 334. 10.1186/s13023-021-01981-4.
44. Jerome, L.A., and Papaioannou, V.E. (2001). DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat Genet *27*, 286–291. 10.1038/85845.
45. Lindsay, E.A., Botta, A., Jurecic, V., Carattini-Rivera, S., Cheah, Y.C., Rosenblatt, H.M., Bradley, A., and Baldini, A. (1999). Congenital heart disease in mice deficient for the DiGeorge syndrome region. Nature *401*, 379–383. 10.1038/43900.
46. 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. 10.1016/s0092-8674(01)00247-1.
47. 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. 10.1038/35065105.
48. Ramirez, A., Vyzas, C.A., Zhao, H., Eng, K., Degenhardt, K., and Astrof, S. (2024). Buffering Mechanism in Aortic Arch Artery Formation and Congenital Heart Disease. Circ Res *134*, e112–e132. 10.1161/CIRCRESAHA.123.322767.
49. Wang, X., Chen, D., Chen, K., Jubran, A., Ramirez, A., and Astrof, S. (2017). Endothelium in the pharyngeal arches 3, 4 and 6 is derived from the second heart field. Dev Biol *421*, 108–117. 10.1016/j.ydbio.2016.12.010.
50. Yamagishi, H. (2024). Human Genetics of Truncus Arteriosus. Adv Exp Med Biol *1441*, 841–852. 10.1007/978-3-031-44087-8_51.
51. Gill, E., and Bamforth, S.D. (2024). Molecular Pathways and Animal Models of Truncus Arteriosus. Adv Exp Med Biol *1441*, 853–865. 10.1007/978-3-031-44087-8_52.
52. Kelly, R.G. (2023). The heart field transcriptional landscape at single-cell resolution. Dev Cell *58*, 257–266. 10.1016/j.devcel.2023.01.010.

53. Zhang, Z., Huynh, T., and Baldini, A. (2006). Mesodermal expression of Tbx1 is necessary and sufficient for pharyngeal arch and cardiac outflow tract development. Development *133*, 3587–3595. 10.1242/dev.02539.
54. Jackson, A., Kasah, S., Mansour, S.L., Morrow, B., and Basson, M.A. (2014). Endoderm-specific deletion of Tbx1 reveals an FGF-independent role for Tbx1 in pharyngeal apparatus morphogenesis. Dev Dyn *243*, 1143–1151. 10.1002/dvdy.24147.
55. Calmont, A., Ivins, S., Van Bueren, K.L., Papangeli, I., Kyriakopoulou, V., Andrews, W.D., Martin, J.F., Moon, A.M., Illingworth, E.A., Basson, M.A., and Scambler, P.J. (2009). Tbx1 controls cardiac neural crest cell migration during arch artery development by regulating Gbx2 expression in the pharyngeal ectoderm. Development *136*, 3173–3183. 10.1242/dev.028902.
56. Randall, V., McCue, K., Roberts, C., Kyriakopoulou, V., Beddow, S., Barrett, A.N., Vitelli, F., Prescott, K., Shaw-Smith, C., Devriendt, K., et al. (2009). Great vessel development requires biallelic expression of Chd7 and Tbx1 in pharyngeal ectoderm in mice. J Clin Invest *119*, 3301–3310. 10.1172/JCI37561.
57. De Bono, C., Liu, Y., Ferrena, A., Valentine, A., Zheng, D., and Morrow, B.E. (2023). Single-cell transcriptomics uncovers a non-autonomous Tbx1-dependent genetic program controlling cardiac neural crest cell development. Nat Commun *14*, 1551. 10.1038/s41467-023-37015-9.
58. Kodo, K., Shibata, S., Miyagawa-Tomita, S., Ong, S.G., Takahashi, H., Kume, T., Okano, H., Matsuoka, R., and Yamagishi, H. (2017). Regulation of Sema3c and the Interaction between Cardiac Neural Crest and Second Heart Field during Outflow Tract Development. Sci Rep *7*, 6771. 10.1038/s41598-017-06964-9.
59. Kodo, K., Uchida, K., and Yamagishi, H. (2021). Genetic and Cellular Interaction During Cardiovascular Development Implicated in Congenital Heart Diseases. Front Cardiovasc Med *8*, 653244. 10.3389/fcvm.2021.653244.
60. 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. 10.1038/85855.
61. 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. 10.1016/j.ajhg.2014.12.025.
62. 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. 10.1016/j.devcel.2005.12.002.
63. Fan, W., Li, Z., He, X., Wang, X., Sun, M., and Yang, Z. (2024). SLC25A1 regulates placental development to ensure embryonic heart morphogenesis. Development *151*. 10.1242/dev.204290.
64. Ohanele, C., Peoples, J.N., Karlstaedt, A., Geiger, J.T., Gayle, A.D., Ghazal, N., Sohani, F., Brown, M.E., Davis, M.E., Porter, G.A., Jr., et al. (2024). The mitochondrial citrate carrier SLC25A1 regulates metabolic reprogramming and morphogenesis in the developing heart. Commun Biol *7*, 1422. 10.1038/s42003-024-07110-8.
65. Chen, X., Wang, L., Huang, R., Qiu, H., Wang, P., Wu, D., Zhu, Y., Ming, J., Wang, Y., Wang, J., and Na, J. (2019). Dgcr8 deletion in the primitive heart uncovered novel microRNA regulating the balance of cardiac-vascular gene program. Protein Cell *10*, 327–346. 10.1007/s13238-018-0572-1.
66. Racedo, S.E., Liu, Y., Shi, L., Zheng, D., and Morrow, B.E. (2024). Dgcr8 functions in the secondary heart field for outflow tract and right ventricle development in mammals. Dev Biol *506*, 72–84. 10.1016/j.ydbio.2023.12.005.

67. Chapnik, E., Sasson, V., Blelloch, R., and Hornstein, E. (2012). Dgcr8 controls neural crest cells survival in cardiovascular development. Dev Biol *362*, 50–56. 10.1016/j.ydbio.2011.11.008.
68. Farrell, M.J., Stadt, H., Wallis, K.T., Scambler, P., Hixon, R.L., Wolfe, R., Leatherbury, L., and Kirby, M.L. (1999). HIRA, a DiGeorge syndrome candidate gene, is required for cardiac outflow tract septation. Circ Res *84*, 127–135. 10.1161/01.res.84.2.127.
69. Devaraju, P., and Zakharenko, S.S. (2017). Mitochondria in complex psychiatric disorders: Lessons from mouse models of 22q11.2 deletion syndrome: Hemizygous deletion of several mitochondrial genes in the 22q11.2 genomic region can lead to symptoms associated with neuropsychiatric disease. Bioessays *39*. 10.1002/bies.201600177.
70. Maynard, T.M., Meechan, D.W., Dudevoir, M.L., Gopalakrishna, D., Peters, A.Z., Heindel, C.C., Sugimoto, T.J., Wu, Y., Lieberman, J.A., and Lamantia, A.S. (2008). Mitochondrial localization and function of a subset of 22q11 deletion syndrome candidate genes. Mol Cell Neurosci *39*, 439–451. 10.1016/j.mcn.2008.07.027.
71. DiGregorio, H., Mansoorshahi, S., Carlisle, S.G., Tovar Pensa, C., Watts, A., McNeely, C., Sabate-Rotes, A., Yetman, A., Michelena, H.I., De Backer, J.F.A., et al. (2025). Contribution of rare chromosome 22q11.2 copy number variants to non-syndromic bicuspid aortic valve. Heart *111*, 221–229. 10.1136/heartjnl-2024-324669.

SUPPLEMENTARY TABLE LEGENDS

Supplementary Table S1. Demographic and CHD phenotype data for 22q11.2DS cohort of 3,016 total subjects. Column A is the record identification of each subject. Column B is the study site listed by city, state and country. Column C is sex of each subject (1 is male and 2 is female). Column D is self-reported ancestry (0 is unknown, 1 is American Indian or Alaska Native, 2 is Asian, 3 is Black or African American, 4 is Native Hawaiian or Other Pacific Islander, 5 is White 6 is mixed of different ancestries, 7 is other ancestry not fitting into the other categories. Column E is deletion size (1 is A-D deletion, 2 is A-B deletion, 3 is A-C deletion, NA is FISH or unknown deletion size). Columns F-J are the principal component results for each subject. Columns K-Y are phenotypes in which 0 means unaffected with that phenotype and 1 means affected with that phenotype, while NA means there is missing data for that defect. CHD is any heart defect, CTD is any conotruncal defect minus SDs (ASD-only, VSD-only and ASD+VSD). Phenotype subgroups are listed in separate columns (see definitions in

Supplementary Table S2; TOF, SD, RAA+ that is RAA with/out other defects, VSD-only is isolated VSD, IAAB, RAA-only that is RAA without other defects, PTA, ASA without other defects, PS/PA+ is PS or PA not part of TOF but can be with other defects, ASD-only is isolated ASD, ASD+VSD is ASD with VSD but no other defects, PS/PA-only is PS or PA but no other defects. Columns Z-AD are genetic admixture ancestry information for each subject, AMR is Admixed American, EAS is East Asian, SAS is South Asian, EUR is European and AFR is African.

Supplementary Table S2. Definition of cardiac phenotypes observed in the 22q11.2DS cohort. This file contains detailed text information about each of the cardiac phenotypes that occurred in the 22q11.2DS cohort.

Supplementary Table S3. There is no difference in the percentage of sex among 22q11.2DS subjects with different deletion types. Sex versus deletion types were compared and percentages are provided.

Supplementary Table S4. There is no difference between self-reported White versus mixed ancestry and deletion type in the 22q11.2DS cohort. Self-reported ancestry of the two largest groups, of White and mixed ancestry is provided with deletion type. There were no significant differences in deletion type.

Supplementary Table S5. Mutually exclusive CHD phenotypes in the 22q11.2DS cohort. Each phenotype was compared to each other (Columns A and B) and examined in the total number of pairs possible (Column C) and tested if they co-occur. The odds ratio (OR; Column E) of their co-occurrence is shown along with the p-value (Column F) and q value (Column G) for the cohort.

Supplementary Table S6. Univariable logistic regression analysis for study site. Each CHD phenotype (Column A) was compared to the recruitment study site location (Column C). The odds ratio (OR; Column D), lower and upper confidence limit (lcl, ucl; Columns E and F), p-value (column G), chi-square p value (chisq p; Column H) and q values for the p-value adjusted for false positives, are indicated (Column I). Significant q values are shown in red font in Column I.

Supplementary Table S7. Univariable logistic regression analysis for sex. Each CHD phenotype (Column A) was compared to the subject sex (Column B). The odds ratio (OR; Column D), lower and upper confidence limit (lcl, ucl; Columns E and F), p-value (column G), chi-square p value (chisq p; Column H) and q values for the p-value adjusted for false positives, are indicated (Column I). Significant q values are shown in red font in Column I.

Supplementary Table S8. Univariable logistic regression analysis for self-reported ancestry. Each CHD phenotype (Column A) was compared to the self-reported ancestry (Column B) and the specific ancestry, defined as (2 is Asian, 3 is Black or African American, 4 is Native Hawaiian or Other Pacific Islander, 5 is White 6 is mixed of different ancestries, 7 is other ancestry not fitting into the other categories) (Column C). The odds ratio (OR; Column D), lower and upper confidence limit (lcl, ucl; Columns E and F), p-value (column G), chi-square p value (chisq p; Column H) and q values for the p-value adjusted for false positives, are indicated (Column I). There are no significant q values.

Supplementary Table S9. Univariable logistic regression analysis for deletion type. Each CHD phenotype (Column A) was compared to the deletion type (Column B) with types indicated (Column C). The odds ratio (OR; Column D), lower and upper confidence limit (lcl, ucl;

Columns E and F), p-value (column G), chi-square p value (chisq p; Column H) and q values for the p-value adjusted for false positives, are indicated (Column I). Significant q values are shown in red font in Column I.

Supplementary Table S10. Univariable logistic regression analysis for principal components. Each CHD phenotype (Column A) was compared to the PC number (Column B) with each PC indicated (Column C). The odds ratio (OR; Column D), lower and upper confidence limit (lcl, ucl; Columns E and F), p-value (column G), chi-square p value (chisq p; Column H) and q values for the p-value adjusted for false positives, are indicated (Column I). Significant q values are shown in red font in Column I.

Supplementary Table S11. Multivariable mixed-effects logistic regression results for CHD phenotypes. Each congenital heart disease phenotype is listed in Column A. Column B indicates the model effect type (fixed effect), and Column C specifies the predictor term. Columns D and E report the regression coefficient (Estimate) and its standard error (STD.error) on the log-odds scale. The corresponding odds ratio (OR) is shown in Column F, with the 95% confidence interval given in Columns G–H. Column I reports the Wald z-statistic, and Column J provides the two-sided p-value. Column K shows the false discovery rate (FDR)–adjusted p-value within phenotype, while Column L reports the study-wide FDR–adjusted q-value across all phenotypes and predictors. Recruitment site was modeled as a random effect in all analyses. Associations meeting study-wide FDR significance are highlighted.

Supplementary Table S12. Correlation between principal components and ancestry proportions. Pearson correlation coefficients (r) and corresponding two-sided p-values are shown for the association between the first five principal components (PC1–PC5; rows) and ADMIXTURE-derived ancestry proportions (columns: East Asian [EAS], South Asian [SAS],

European [EUR], African [AFR], and Admixed American [AMR]). Correlations were computed using individuals with non-missing values for both variables. This table is used to aid interpretation of ancestry-related principal components identified in multivariable analyses.

Figure 1

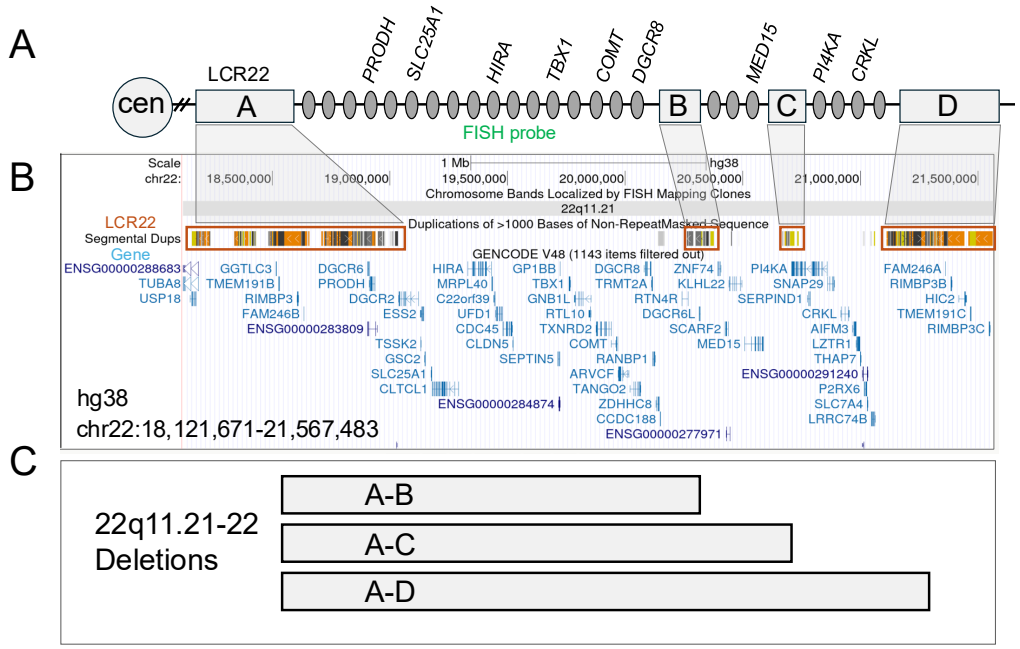


Figure 1. Map of 22q11.2 region. (A) Diagram of the 22q11.21-22 region with the centromere (cen) shown on the left. The position of the LCR22s (low copy repeats on 22q11.2), -A, -B, -C, -D are shown as light gray boxes. Genes are indicated (gray ovals) with representative known coding genes, between the LCR22s. (B) Screenshot from the UCSC browser, hg38 assembly (coordinates are indicated). The position of the LCR22s (segmental duplications are shown (orange boxes) as are the genes spanning the region (Genecode track). (C) The position of the deletions in the 22q11.2 region, A-B, A-C and A-D, correspond to chromosome breakpoints in the respective LCR22s.

Figure 2

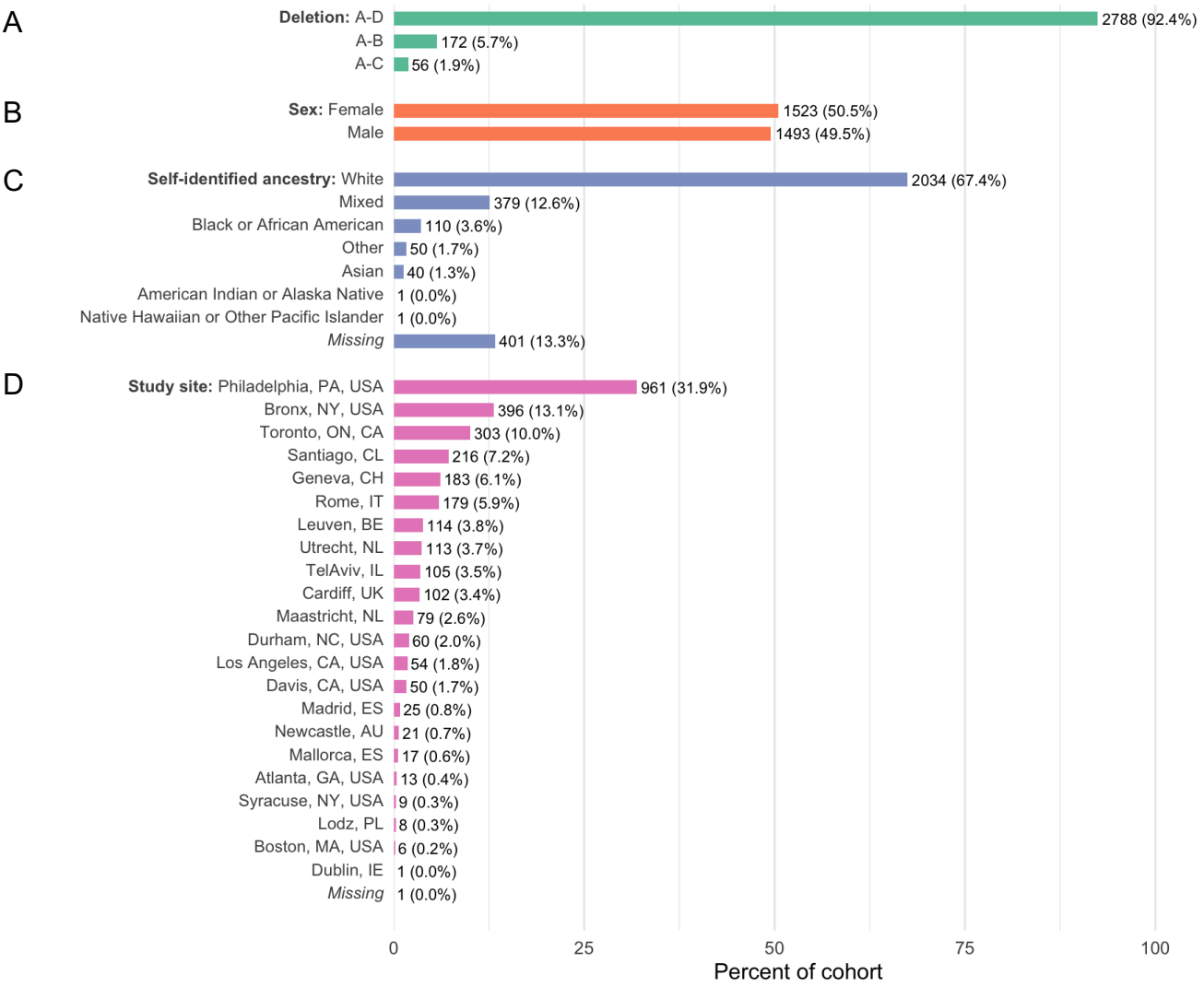


Figure 2. Cohort composition across key demographic and genomic variables. Horizontal bar charts display the proportion of the study cohort that falls into each category. (A) 22q11.2 deletion size, (B) self-reported sex, gender, (C) self-identified ancestry and (D) recruiting study site. Bars are colored by variable, scaled to the percentage of the full cohort, and labelled with the corresponding count and percent.

Figure 3

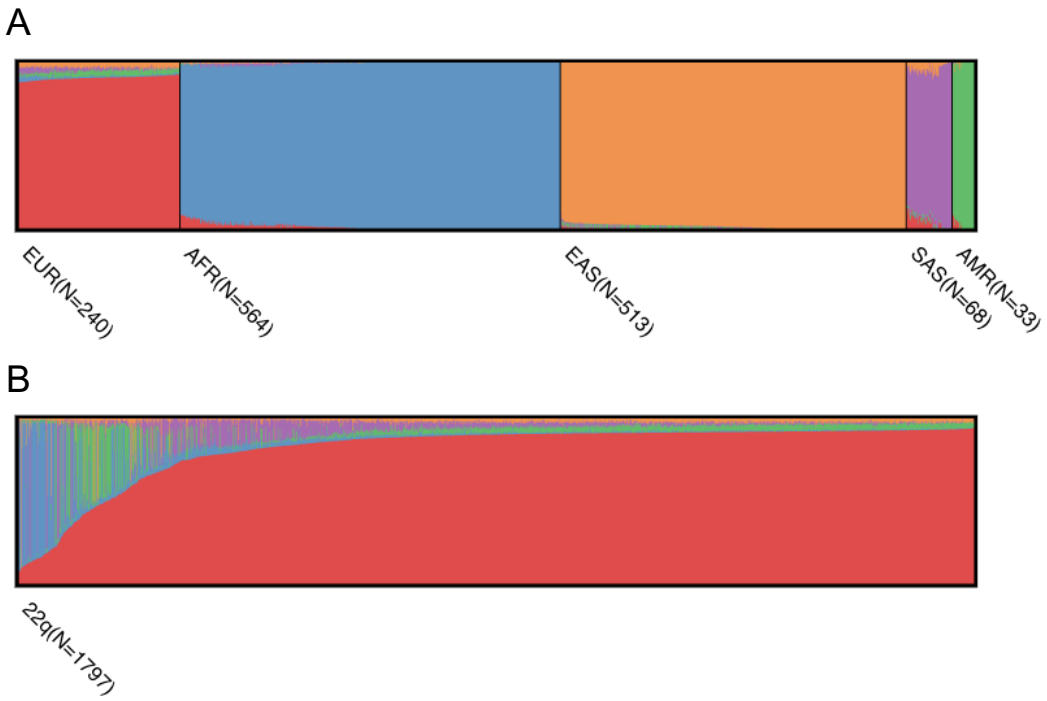


Figure 3. Analysis of 22q population structure. (A) Admixture plot (k=5) for 1000 Genomes reference individuals, defining the global ancestry color mapping. (B) Admixture plot (k=5) all 1,797 cohort individuals, where each bar represents the estimated global ancestry of an individual. Colors correspond to predominant global ancestry as defined in (A): red represents European ancestry (EUR), purple represents South Asian ancestry (SAS), blue represents African ancestry (AFR), green represents American ancestry (AMR), and orange represents East Asian ancestry (EAS).

Figure 4

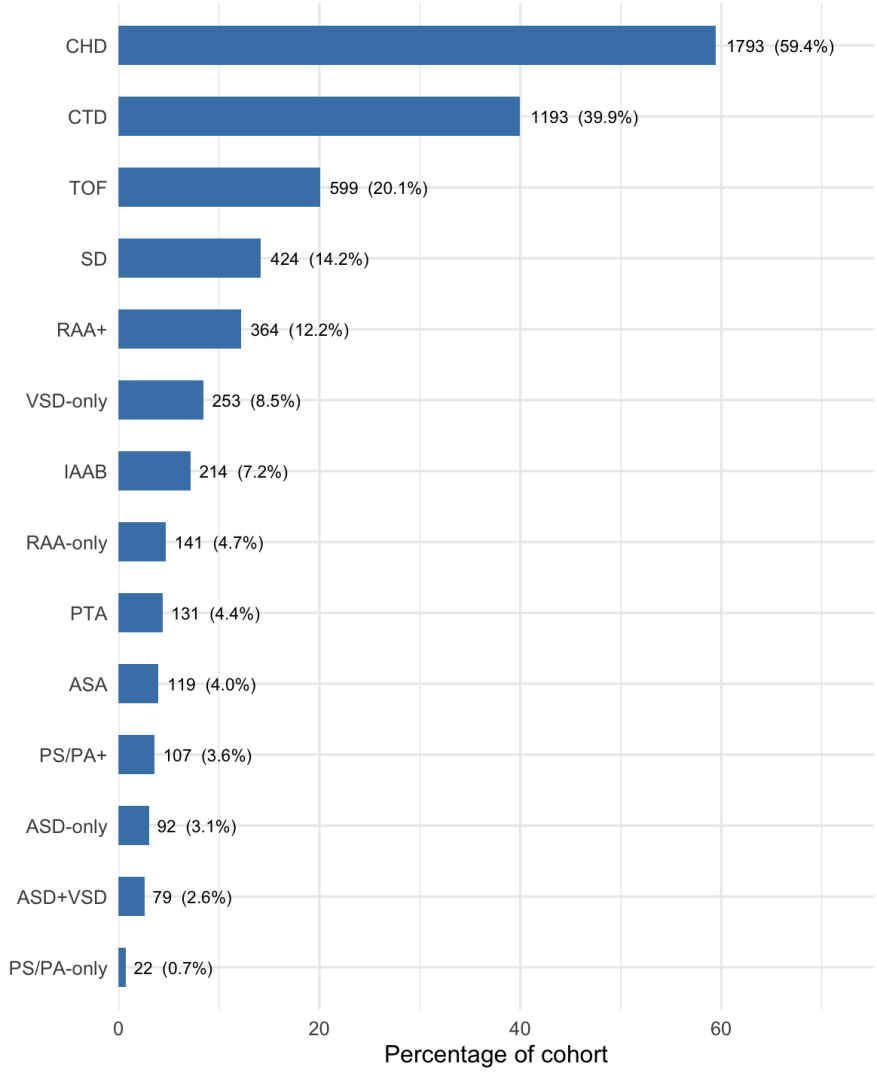
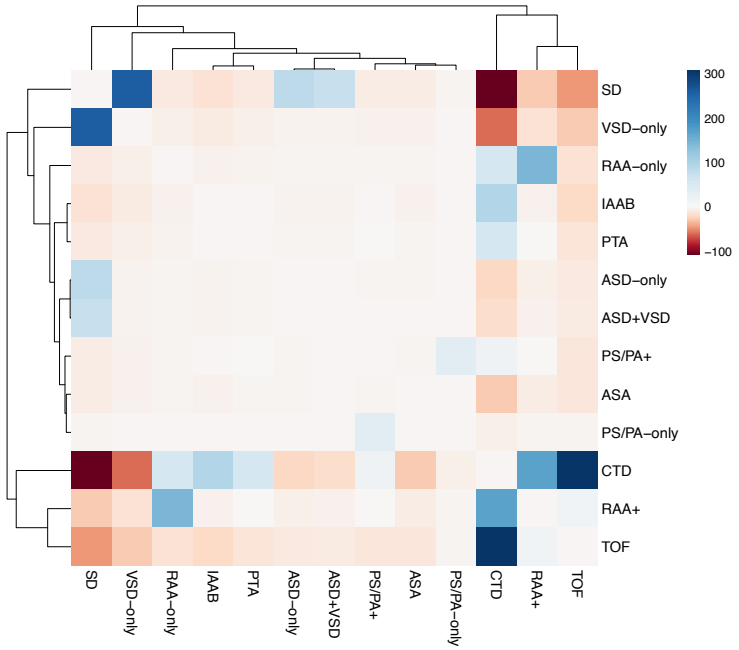


Figure 4. Prevalence of congenital heart defects in the 22q11.2 deletion cohort. Bar graph with x-axis indicates the percentage of the cohort and indicated separately for each phenotype class and subgroup with the number of individuals per group (percentage is indicated). Congenital heart disease-CHD; conotruncal heart defect-CTD; tetralogy of Fallot-TOF; septal defects-SD, atrial septal defects, only-ASD-only, right sided aortic arch with other defects-RAA+; interrupted aortic arch type B-IAAB; pulmonic stenosis or atresia excluding TOF but with other defects-PS/PA+; persistent truncus arteriosus-PTA; abnormal origin of the right or left subclavian artery-ASA; atrial septal and ventricular septal defects-ASD+VSD; right sided aortic arch with no other defects-RAA-only; ventricular septal defects with no other defects-VSD-only; pulmonic stenosis or atresia and no other defects-PS/PA-only.

Figure 5

A



B

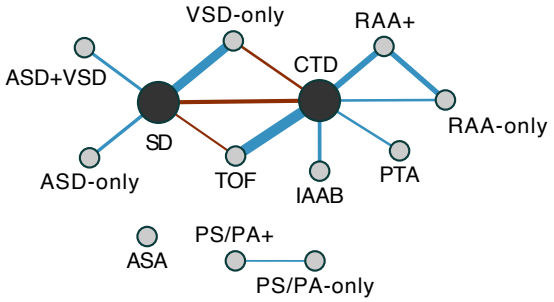


Figure 5. Pairwise associations among CHD phenotypes in 22q11.2DS. (A) The heat-map shows the signed $-\log_{10}(q\text{-value})$ from Fisher's exact tests for every pair of CHD subtypes. Positive values (blue) indicate phenotypes that co-occur more often than expected, whereas negative values (red) denote mutual exclusivity; the color intensity is proportional to statistical significance after Benjamini-Hochberg correction. White cells represent non-significant comparisons ($|-\log_{10}q| < 1$). Rows and columns are ordered by average-linkage hierarchical clustering (dendrograms). (B) Network of significant pair-wise relationships. Each node represents a CHD type; edges connect phenotype pairs with an FDR-adjusted $-\log_{10}(q\text{-value}) > 5$. Blue edges denote positive associations (co-occurrence), red edges denote negative associations (mutual exclusivity). Edge width is proportional to $-\log_{10}(q\text{-value})$.