Original Article

OPEN

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

Tingwei Guo, PhD; Gabriela M. Repetto, MD; Donna M. McDonald McGinn, MS, CGC; Jonathan H. Chung, PhD; Hiroko Nomaru, PhD; Christopher L. Campbell, PhD; Anna Blonska, MD, PhD; Anne S. Bassett, MD; Eva W.C. Chow, MD; Elisabeth E. Mlynarski, PhD; Ann Swillen, PhD; Joris Vermeesch, PhD; Koen Devriendt, MD, PhD; Doron Gothelf, MD; Miri Carmel, PhD; Elena Michaelovsky, PhD; Maude Schneider, PhD; Stephan Eliez, MD, PhD; Stylianos E. Antonarakis, MD, PhD; Karlene Coleman, MS, CGC; Aoy Tomita-Mitchell, PhD; Michael E. Mitchell, MD; M. Cristina Digilio, MD; Bruno Dallapiccola, MD; Bruno Marino, MD; Nicole Philip, MD; Tiffany Busa, MD; Leila Kushan-Wells, MS; Carrie E. Bearden, PhD; Małgorzata Piotrowicz, MD, PhD; Wanda Hawuła, MS; Amy E. Roberts, MD; Flora Tassone, PhD; Tony J. Simon, PhD; Esther D.A. van Duin, MS; Thérèse A. van Amelsvoort, MD, PhD; Wendy R. Kates, PhD; Elaine Zackai, MD; H. Richard Johnston, PhD; David J. Cutler, PhD; A.J. Agopian, PhD; Elizabeth Goldmuntz, MD; Laura E. Mitchell, PhD; Tao Wang, PhD; Beverly S. Emanuel, PhD; Bernice E. Morrow, PhD; on behalf of the International 22q11.2 Consortium/Brain and Behavior Consortium*

Background—The 22q11.2 deletion syndrome (22q11.2DS; DiGeorge syndrome/velocardiofacial syndrome) occurs in 1 of 4000 live births, and 60% to 70% of affected individuals have congenital heart disease, ranging from mild to severe. In our cohort of 1472 subjects with 22q11.2DS, a total of 62% (n=906) have congenital heart disease and 36% (n=326) of these have tetralogy of Fallot (TOF), comprising the largest subset of severe congenital heart disease in the cohort.

Methods and Results—To identify common genetic variants associated with TOF in individuals with 22q11.2DS, we performed a genome-wide association study using Affymetrix 6.0 array and imputed genotype data. In our cohort, TOF was significantly associated with a genotyped single-nucleotide polymorphism (rs12519770, P=2.98×10^{-8}) in an intron of the adhesion *GPR98* (G-protein–coupled receptor V1) gene on chromosome 5q14.3. There was also suggestive evidence of association between TOF and several additional single-nucleotide polymorphisms in this region. Some genome-wide significant loci in introns or noncoding regions could affect regulation of genes nearby or at a distance. On the basis of this possibility, we examined existing Hi-C chromatin conformation data to identify genes that might be under shared transcriptional regulation within the region on 5q14.3. There are 6 genes in a topologically associated domain of chromatin with *GPR98*, including *MEF2C* (Myocyte-specific enhancer factor 2C). *MEF2C* is the only gene that is known to affect heart development in mammals and might be of interest with respect to 22q11.2DS.

Conclusions—In conclusion, common variants may contribute to TOF in 22q11.2DS and may function in cardiac outflow tract development. **(*Circ Cardiovasc Genet*. 2017;10:e. DOI: 10.1161/CIRCGENETICS.116.001690.)**

Key Words: chromosomes ■ DiGeorge syndrome ■ genotype ■ ivelo-cardio-facial syndrome ■ tetralogy of Fallot

One of the greatest challenges in the area of human genetics is to understand the basis of phenotypic heterogeneity in known diseases. The 22q11.2 deletion syndrome (22q11.2DS; velocardiofacial syndrome/DiGeorge syndrome; Mendelian Inheritance in Man No. 192430, 188400) is one of the most common genomic disorders, occurring in 1 of 4000 live births.[1] Over 90% of affected individuals have a de novo, hemizygous 3 million base pair (Mb) deletion on chromosome 22q11.2.[2–4] All subjects with the deletion have features of the syndrome,

See Editorial by McBride and Ware
See Clinical Perspective

Received December 16, 2016; accepted June 29, 2017.

*A list of the International 22q11.2 Consortium and IBBC is listed in Table 1 in the Data Supplement.

The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.116.001690/-/DC1.

Correspondence to Bernice E. Morrow, PhD, Department of Genetics, Albert Einstein College of Medicine, 1301 Morris Park Ave, Bronx, NY 10461. E-mail Bernice.morrow@einstein.yu.edu

but the clinical presentation is quite variable. For example, 60% to 70% of patients have congenital heart disease (CHD) involving the cardiac outflow tract (OFT) and aortic arch, whereas the rest have apparently normal cardiac structures.[5] Among the most serious defect observed in individuals with the 22q11.2DS is tetralogy of Fallot (TOF), which is defined by the presence of a ventricular septal defect, pulmonary stenosis, overriding aorta, and right ventricular hypertrophy. TOF is caused in part by failed migration or differentiation of second heart field mesodermal cells from the pharyngeal apparatus in embryos, needed to form or remodel the cardiac OFT.[6] TOF occurs in 1 of 2500 live births in the general population (Center for Disease Control and Prevention). Among individuals with the 22q11.2DS, 36% in our cohort has TOF, and among individuals with TOF, ≈15% have a 22q11.2 deletion.[7,8]

Among the genes in the deleted region on 22q11.2, *TBX1*, which encodes a T-box transcription factor, is the major candidate for CHD.[9–11] *Tbx1* is expressed in the second heart field mesoderm, which is disrupted in 22q11.2DS.[9–11] Global inactivation[9–11] or second heart field–specific inactivation[12] of *Tbx1* results in neonatal lethality with severe cardiac OFT defects. One hypothesis to explain variable phenotypic expression in the 22q11.2DS is the presence of pathogenic variants in *TBX1* on the haploid allele of 22q11.2. Previously, we tested whether common or rare single-nucleotide variants (SNVs) in the coding region of *TBX1* on the remaining allele of 22q11.2 were associated with CHD in 22q11.2DS subjects, but we did not find an association.[13] Another hypothesis is that there are copy number variations elsewhere in the genome that could explain differences in phenotypes. We previously found that a commonly occurring genomic duplication encompassing the glucose transporter gene, *SLC2A3*, was associated with CHD ($P=2.68\times10^{-4}$).[14] This copy number variation occurred in 5.8% of individuals with 22q11.2DS and CHD and 1.1% of those with 22q11.2DS and no CHD. Recently, a partial duplication of a chromatin modifier, *KANSL1*, was associated with CHD in a Chilean 22q11.2DS cohort.[15] However, these copy number variations occurred in only some deleted subjects with CHD and thus do not explain the basis of phenotypic variability in the majority of patients.

Our goal was to identify common single-nucleotide polymorphisms (SNPs) that are associated with TOF in individuals with 22q11.2DS. We restricted our analyses to TOF because it is the largest single phenotypic category of severe CHD in our cohort. Restricting our analyses in this way may reduce heterogeneity in the genes that contribute to CHD in individuals with 22q11.2DS and thus may increase the power of a GWAS.

Methods

Human Subjects and Phenotype Data

We assembled a cohort of subjects with 22q11.2DS (Tables I and II in the Data Supplement). Subjects were previously recruited by the International Chromosome 22q11.2DS Consortium, the International 22q11.2 Brain Behavior Consortium (http://22q11-ibbc.org), and clinical groups that specialize in the treatment of individuals with 22q11.2DS. All subjects within the cohort had a clinical diagnosis of 22q11.2DS that was confirmed by the presence of a 22q11.2 deletion using fluorescence in situ hybridization or multiplex ligation-dependent probe amplification (SALSA MLPA kit P250 DiGeorge; MRC Holland, The Netherlands). Informed consent was obtained for all participants, and this study was conducted under an Internal Review Board-approved protocol at the Albert Einstein College of Medicine (CCI 1999-201).

For this study, we used previously collected genomic DNA and phenotypic and demographic information. We obtained echocardiogram and cardiology reports to confirm the specific CHD diagnosis (eg, TOF).

SNP Array Genotype and Data Quality Control

Genomic DNA from 1244 study subjects was array genotyped using Affymetrix GeneChip Genome-Wide SNP 6.0 array. The majority of samples were genotyped was at the Genomics Facility core laboratory of Albert Einstein College of Medicine. However, 37 samples were genotyped in the Advanced Genomics laboratory core at the Children's Research Institute (Milwaukee, WI) for clinical purposes,[16] and 191 Chilean samples were genotyped in the Center for Human Genetics, Clínica Alemana Universidad del Desarrollo, Santiago, Chile.[15]

The raw data from all arrays were processed through the same pipeline using the same criteria. Genotype data from arrays with contrast quality control scores ≤0.4 per sample, contrast quality control <1.7 per batch, and Median Absolute Pairwise Difference metric >0.35, were excluded. Genotypes were called using the Birdseed V2 Genotyping Algorithm (call rate: 99.02±0.02%). To account for batch effects, BEAGLECALL Version 1.0.1 software was used to rescore genotypes.[17] SNPs with call rates <95%, minor allele frequency <1%, or Hardy–Weinberg equilibrium *P* value $<10^{-5}$ were excluded. In addition, samples that showed second-degree relatedness or closer, based on identity by state, were removed. For each subject, deletion size was determined using the $\log_2$ intensity ratio as estimated by the Copy Number Analysis Module of Golden Helix Powerseat Package. The CEL files and genotype data are being deposited to National Center for Biotechnology Information database of Genotypes and Phenotypes phs001339.v1.p1.

We performed imputation to increase the number of SNPs available for analysis. Only genotyped SNPs with minor allele frequency >1% were used for imputation. Haplotypes were prephased using SHAPEIT software,[18,19] and imputation was performed using IMPUTE2 with the 1000 Genomes Phase I data set as the reference panel.[20] Imputed SNPs with minor allele frequency ≤1% or imputation quality (INFO) scores ≤0.8 were excluded from the GWAS.

Statistical Methods

We used a case–control approach, in which individuals with 22q11.2DS and TOF were considered cases and individuals with 22q11.2DS without CHD were considered controls. We conducted principal component analyses to identify the PCs of race/ethnicity. Potential associations between TOF, sex, and deletion size were assessed using logistic regression adjusted for the first 4 PCs. A *P* value <0.05 was considered significant.

The association between TOF and each SNP was assessed by logistic regression analysis under an additive genetic model using data from all study subjects. These analyses were performed using SNPTEST v2.5.2 https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html) and accounted for the genotyping accuracy and first 4 PCs of race/ethnicity.[21,22] A *P* value of 5×10^{-8} was used as the genome-wide significance cutoff for single association tests. For a meta-analysis, the cohort was split into groups determined by principal component analyses. Each group was analyzed separately using logistic regression, and the results were meta-analyzed using the inverse-variance method. Power for these analyses was assessed using QUANTO (http://biostats.usc.edu/Quanto.html). Manhattan plots and quantile–quantile (Q–Q) plots were generated using Golden Helix Powerseat. For regions of interest identified in the GWAS, regional association plots were generated using LocusZoom software (http://locuszoom.sph.umich.edu/locuszoom/).[23]

Conditional logistic regression analyses were performed to determine whether multiple variants within a region are independently associated with TOF in individuals with 22q11.2DS. Specifically, within a region, we conditioned on the genotyped SNP with the lowest P value and the first 4 principal components of race/ethnicity and individually evaluated the association of TOF with each additional SNP in the region. Conditional analyses were conducted using SNPTEST v2.5.2.

Linkage Disequilibrium Analysis of Whole-Genome Sequence to Identify Variants in Linkage Disequilibrium With GWAS Findings

Whole-genome sequencing of a subset (n=397) of our 22q11.2DS samples was performed using the Illumina HiSeq2000 and HiSeq X Ten platform at Hudson Alpha Institute for Biotechnology (Huntsville, Alabama) as part of the International 22q11.2 Brain and Behavior Consortium to find genes for schizophrenia. Variant calling was performed using PEMapper software for read mapping to the hg38 (GRCh38) reference genome and PECaller software for variant calling.[24] CrossMap (http://crossmap.sourceforge.net/) was used to convert genome coordinates between hg38(GRCh38) and hg19 (GRCh37).[25] To follow-up on the top result from the GWAS (Results), the genomic region (chromosome 5 [chr5]:88 703 723–91 409 593) from *MEF2C* through *ARRDC3* was extracted from the whole-genome sequencing data. Functional annotation of SNVs was performed using the Variant Classification and the Annotate and Filter tools in the Golden Helix software. Nonexonic SNVs (intronic, intergenic) were removed, and predicted functional SNVs were used to generate an linkage disequilibrium (LD) matrix using Haploview. LD measurements of r^2 >0.8 were used to define LD haplotype blocks.

Mouse Embryo Analysis and Whole-Mount RNA In Situ Hybridization

Gene expression profiling was previously performed to identify differentially expressed genes in the second heart field mesoderm of wild-type mouse embryos.[12] Data from wild-type embryos were extracted for evaluation of specific expression levels in this tissue. For in situ hybridization, RNA probes were generated from mouse embryo cDNA using digoxigenin-uridine triphosphate (Roche Diagnostic Corp, Indianapolis, IN; Table III in the Data Supplement). Swiss Webster strain wild-type mouse embryos were isolated at day (E)9.5 or E10.5 and used for experiments.

Results

Description of the 22q11.2DS Population

A total of 1472 unrelated subjects with 22q11.2DS were ascertained through multiple sources (Tables I and II in the Data Supplement) and were genotyped on Affymetrix 6.0 arrays. Genotypes (n≈6.6 million SNPs with minor allele frequency >0.01 and INFO>0.8) were also imputed.

The characteristics of the study subjects are shown in Table 1. These subjects predominantly self-reported as white and non-Hispanic (Table 1). This was confirmed by principal component analyses using 4 PCs (we did not observe a difference between usage of 4 or 10 PCs) to estimate genetic ancestry (Figure I in the Data Supplement). The majority (191/217) of Hispanic subjects were recruited at the collection site in Santiago, Chile. Approximately 62% of the subjects (n=906) had CHD at birth. There are 4 sets of low copy repeats, termed LCR22A, B, C, and D, that span the 22q11.2 region. All of the subjects had a deletion of 1 allele of *TBX1*, which is located in the LCR22A-B interval, and the majority of subjects had the typical 3 Mb deletion flanked by LCR22A-D (Table 1; Table IV in the Data Supplement). In this cohort, neither sex nor deletion size were significantly associated with CHD in general or with TOF specifically (P>0.05).

GWAS to Identify Genetic Loci for TOF

The TOF phenotype comprised the largest individual group of subjects with severe intracardiac anomalies in our 22q11.2DS population (36%; Table 1; Figure 1). We conducted a GWAS to identify genetic variants associated with TOF. This analysis was based on data from 326 subjects with 22q11.2DS and TOF and 566 subjects with 22q11.2DS and normal cardiac anatomy. This study had a power of 80% to detect an odds ratio of >1.9 for a common SNP with an allele frequency >0.3 under a log-additive model at $P < 5\times10^{-8}$.

Subjects of all races and ethnicities were included, and associations were assessed using logistic regression adjusted with the first 4 PCs of race/ethnicity. As neither deletion size nor sex was significantly associated with TOF in this cohort, these variables were not included in the logistic models. The genomic inflation factor (λ=1.02) and the Q–Q plot (Figure II in the Data Supplement) provided little evidence of a systematic deviation from the expected distribution of the test statistic.

Three SNPs mapping to intron 61 of *GPR98* (G-protein–coupled receptor 98) were significantly associated with TOF. The genotyped SNP, rs12519770 ($P=2.98\times10^{-8}$),

Table 1. Characteristics of Study Subjects

Subject Characteristics	n (%)
Self-reported race	
White	1229 (83)
Black	34 (2)
Asian	11 (0.7)
Native American	1 (<0.1)
Mixed Ancestry	20 (1)
Unknown	177 (12)
Self-reported ethnicity	
Non-Hispanic	1255 (85)
Hispanic	217 (15)
Sex	
Male	717 (49)
Female	754 (51)
Unknown	1 (<0.1)
Congenital heart defect	
TOF	326 (22)
Other CHD	580 (39)
Normal (controls)	566 (38)
Deletion size	
Typical 3 Mb (LCR22A-D)	1366 (93)
Nested LCR22A-B	73 (5)
Nested LRC22A-C	26 (2)
Other	12 (1)

The cohort of 1472 samples is shown categorized based on self-reported race, ethnicity, and sex. The numbers and percentages in each category are indicated. A total of 906 subjects have congenital heart disease (CHD), whereas the rest have normal structures. The deletion sizes are indicated. LCR indicates low copy repeats; and TOF, tetralogy of Fallot.

and imputed SNPs, rs7720206 ($P=2.22\times10^{-8}$) and chr5: 90 067 043:D ($P=2.10\times10^{-8}$), showed the strongest association (Figure 2A, Table 2). These three SNPs seem to be in complete LD (Figure 2B and 2C). For rs12519770, the A allele was the risk allele with a frequency of 0.58 in TOF cases and 0.45 in controls, conferring an odds ratio of 1.69 ($P=3.2\times10^{-8}$) per copy of the A allele in the 22q11.2DS cohort (Table 2; Figure 2). There was also suggestive evidence of association between TOF and 2 additional groups of SNPs in *GPR98*. The top genotyped SNP in each cluster (rs6889138, rs6893710) is listed in Table 2 and illustrated in Figure 2B and 2C.

Although our initial GWAS adjusted for the first 4 PCs of race/ethnicity, the observed associations may still reflect bias because of uncontrolled confounding resulting from population stratification. Consequently, we repeated our analyses for the top SNP, rs12519770, after separating the cohort into 3 groups: white, Admixed, and African, as determined by principal component analyses (Figure I in the Data Supplement). The *P* value for this SNP was significant in the meta-analysis ($P=4.43\times10^{-8}$; Table V in the Data Supplement), suggesting that the observed association is unlikely to be the result of population stratification.

Within the 5q14.3 region, there seemed to be 3 clusters of SNPs that were associated with TOF. We refer to these as clusters 1, 2, and 3, and the clusters are ranked in ascending order based on the *P* value for the top SNP within the cluster. To determine whether >1 variant was independently associated with TOF, we performed conditional analyses in which we conditioned on the genotyped SNP in *GPR98* with the smallest *P* value in cluster 1 (rs12519770; $P=2.98\times10^{-8}$) and individually evaluated the association of TOF with each of the additional SNPs in the 5q14.3 region (n=1344 SNPs, Table VI in the Data Supplement). In these conditional analyses, there was suggestive evidence for association with 1 SNP (rs6893710, $P=3.92\times10^{-5}$). This variant was the top SNP in the third cluster of associated genes (Table 2). The association of

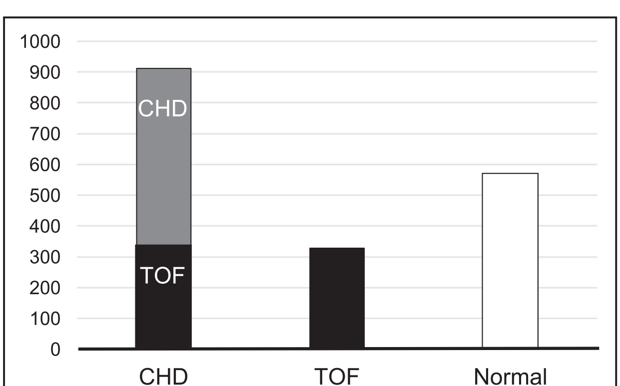

Figure 1. Distribution of cardiovascular phenotypes in 1472, 22q11.2 deletion syndrome (22q11.2DS) subjects. The number of subjects (*y* axis) sorted into phenotypes (*x* axis) is shown in the bar graph. All individuals have a hemizygous 22q11.2 deletion. The most serious cardiovascular diagnoses with the largest number of subjects is tetralogy of Fallot (TOF; n=326; black bar) among a total with congenital heart disease (CHD; n=906; gray bar) when compared with those with no intracardiac or aortic arch anomalies as detected by echocardiogram summary and cardiology report (white bar).

the top SNP in the second cluster (rs6889138) was attenuated in the conditional analysis ($P=0.002$).

Definition of the 5q14.3 Locus

To identify nonsynonymous variants that may be in LD with the rs12519770 and to narrow the region containing the association signal on 5q14.3 based on LD, we performed an LD analysis using existing whole-genome sequence data from 397 individuals with 22q11.2DS (http://22q11-ibbc.org; unpublished data, International 22q11.2 Brain and Behavior Consortium authors in Supplementary Table 1, 2017). These individuals comprise a subset of the samples genotyped on Affymetrix 6.0 arrays and were selected based on psychiatric but not cardiovascular phenotype. There were 9680 SNVs identified in the 2.7 Mb region around *GPR98* (chr5: 8 799 640–90 704 983). There were 161 coding SNVs (102 nonsynonymous, 58 synonymous, and 1 splicing) and 115 SNVs in the 3′ or 5′-untranslated regions for a total of 276 SNVs (Table VII in the Data Supplement). None of these SNVs were in LD with rs12519770. The SNP, rs6893710, was in weak LD with the synonymous variant, rs41304884 (*GPR98*, NM_032119.3, c.16164G>A;D′=0.859, $r^2=0.389$; Figure 2B). There were no nonsynonymous variants related to our association signal.

Most GWAS signals that have been previously discovered are in intergenic regions and may mark transcriptional regulatory regions in the genome rather than genes themselves. To test this for TOF in 22q11.2DS, we examined the LD pattern from available whole-genome sequencing data on the same 397 22q11.2DS subjects to narrow the interval with SNPs showing the strongest association. We narrowed down the association signal to a 104.7 kb region on chromosome 5 (chr5: 90 057 563–90 162 285; Figure 2C and 2D, red block). Most of the common SNPs with *P* values $<10^{-5}$ were located in this region (Figure 2C). A similar LD pattern has been observed in the white subset from the 1000 Genomes Project (Figure III in the Data Supplement). Thus, we were able to narrow the region of the association signal.

Genes Mapping to 5q14.3

Because the SNPs found within the intron of *GPR98* might affect its regulation or, instead, the regulation of other genes in the region, we examined local chromosome conformation forming topologically associated domains (TADs).[26–28] To investigate higher-order chromatin-mediated looping, we extracted data for the 5q14.3 interval from the Hi-C browser.[29] There was chromatin interaction data for 28 different cell lines ranging from H1 embryonic stem cells to cancer cell lines.[29–31] We focused on TAD contact domains in a 3.3 Mb region including *GPR98* (Figure 3; Figure IV in the Data Supplement). Because the 104.7 kb region with genetic association is within the *GPR98* locus, it is possible that variants might affect its expression or that of nearby genes. In addition to *GPR98*, this region includes 6 additional genes (Figure 3; Figure IV in the Data Supplement), including 5 protein coding genes within the 2.3 Mb TAD: *MEF2C* (Myocyte enhancer factor 2C), *CETN3* (Centrin 3), *MBLAC2* (Metallo-β-lactamase domain containing 2), *POLR3G* (RNA polymerase III subunit G), and *LYSMD3* (LysM, putative peptidoglycan-binding, domain containing 3). One gene, *ARRDC3* (Arrestin domain

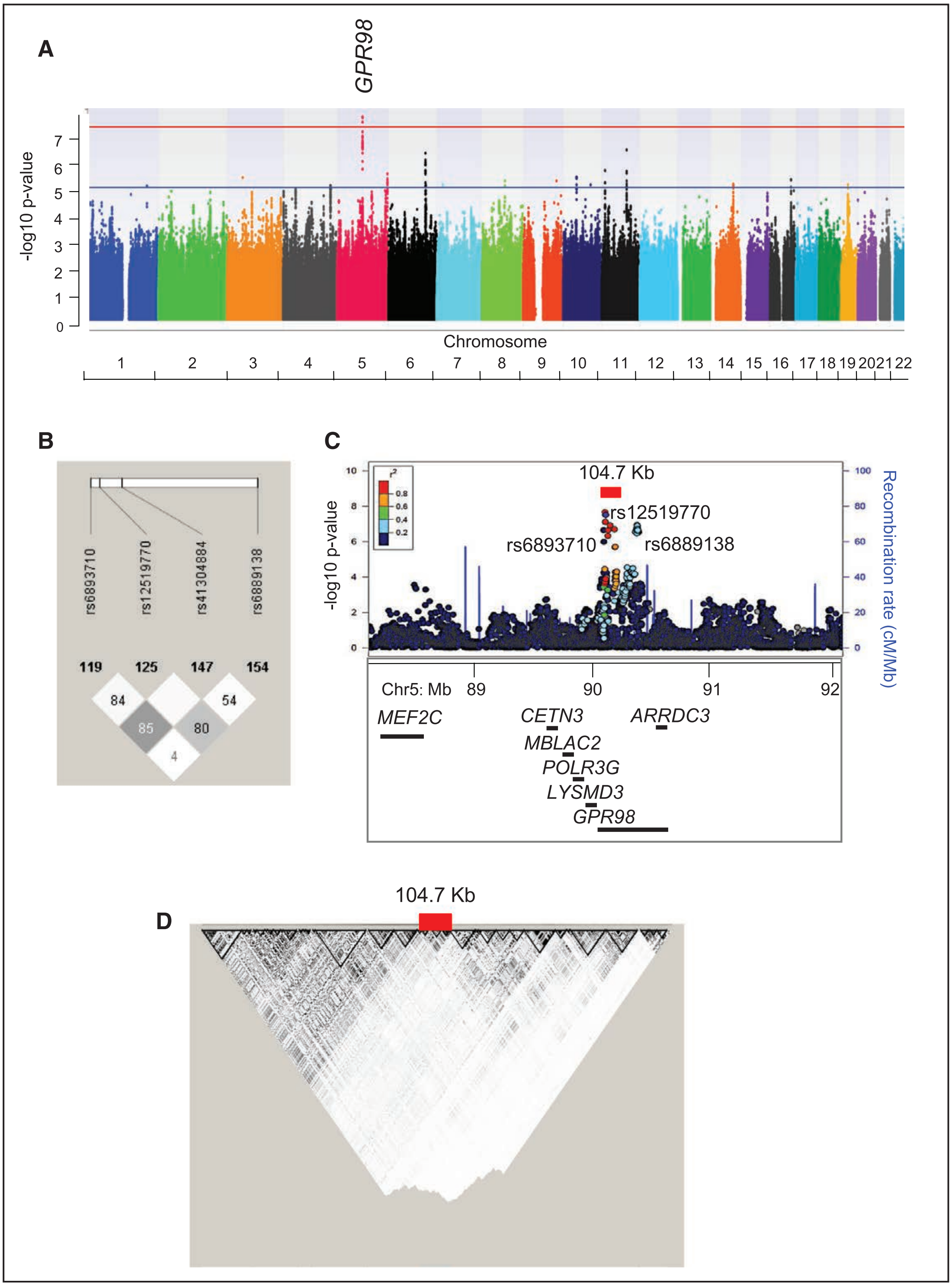

Figure 2. Genome-wide association results for tetralogy of Fallot (TOF) in 22q11.2 deletion syndrome (22q11.2DS). **A**, Values in the Manhattan plot for TOF vs controls were plotted against their respective positions on the autosomal chromosomes. The red line represents the genome-wide significance threshold (P=5×10^{-8}). The blue line represents the threshold for suggestive association (P=1×10^{-5}). A single locus marked by the *GPR98* (G-protein–coupled receptor V1) gene reached genome-wide significance. **B**, LD matrix of selected, predicted functional SNPs with top *P* values on the 5q14.3 region from WGS (Methods). The LD with respect to associated single-nucleotide polymorphisms (SNPs) with highest *P* values in the region is shown. The LD plot is based on r^2 values. Key: r^2=0 is given in white, 0<r^2<1 is given in shades of grey and r^2=1 is given in black. The pairwise D′ values are provided in the boxes. Nine SNPs are in modest LD (D′=0.80, r^2=0.23) with rs12519770. The G allele of the top genotyped SNP, rs6889138, located in intron 74 of *GPR98* and had a minor allele frequency (MAF) of 0.30 in TOF cases, and 0.21 in controls giving an odds ratio (OR) of 1.68 (P=1.72×10^{-7}) per copy in the 22q11.2DS cohort. There were 14 SNPs in the second group that had suggestive association with TOF and were in modest LD with SNP, rs12519770 (D′=0.84, r^2=0.02). The top SNP, rs6893710, is located in intron 47 of *GPR98* and had a MAP of 0.058 in TOF cases, but 0.015 in controls, giving an OR of 4.05 (P=1.04×10^{-6}) per copy of the C allele in the 22q11.2DS cohort (Table 2). **C**, LocusZoom plot of region of association at rs12519770 on 5q14.3 indicating $-\log_{10} P$ values (*y* axis) against the chromosomal positions of SNPs (*x* axis). The genotyped (*Continued*)

Figure 2 Continued. SNP with the strongest association signal in each locus is represented as a purple diamond; the other SNPs are colored according to the extent of LD (correlation r^2 is based on CEU HapMap haplotypes) with this SNP. Estimated recombination rates (GRCh37/hg19, CEU; 1000 Genomes Project 2012) are shown as light blue lines. Genes are indicated below the LocusZoom plot. **D**, Fine mapping of whole-genome sequencing (WGS; MAF>0.05) based on LD and imputed SNPs from the arrays was used to narrow the TOF signal to a 104.7 Kb region (chromosome 5 [chr5]: 90 057 563–90 162 285) as shown. Most of the common SNPs from this region have *P* values <10^{-5}. ARRDC3 indicates Arrestin domain containing 3; CETN3, Centrin 3; CEU, Northern Europeans who are Utah residents part of the CEPH collection; GPR98, G-protein–coupled receptor V1; LYSMD3, LysM, putative peptidoglycan-binding, domain containing 3; MBLAC2, Metallo-β-lactamase domain containing 2; MEF2C, Myocyte-specific enhancer factor 2C; and POLR3G, RNA polymerase III subunit G.

containing 3), maps downstream of *GPR98*, but it is in a different TAD (Figure 3). The same domain structure marked by TAD triangles occurred in most of the cell lines that were examined (Figure IV in the Data Supplement).

We next determined whether any of the genes are expressed in the pharyngeal apparatus or heart in embryos. Probes were generated, and in situ hybridization was successfully performed for *Gpr98, Cetn3*, *Lysmd3*, and *Mef2c* in mouse embryos at E9.5 and E10.5, when the cardiac OFT is expanding (Figure 3). Although *Gpr98* is weakly expressed at E9.5, it is strongly expressed in the neural tube region, particularly the hindbrain[32] (Figure 3). *Cetn3*, important in the cilia[33] for centrosome reproduction,[34] and *Lysmd3*, of unknown function, are ubiquitously expressed, although *Cetn3* has lower expression levels in the heart itself (Figure 3). *Mef2c* encodes a MADS box transcription factor, and it is expressed in the pharyngeal apparatus including the second heart field mesoderm,[35] as indicated in Figure 3. Among the genes in the TAD, *MEF2C* is the only one specifically expressed in cardiac progenitor cells known to be required for heart development.[36]

The second heart field mesodermal progenitor cell populations forming the cardiac OFT lie within the distal pharyngeal apparatus. We then examined expression levels of genes in the 5q14.3 region in existing Affymetrix microarray data from the microdissected distal pharyngeal apparatus.[12] The purpose was to determine whether any of the genes are expressed in this critical tissue. We compared expression levels of genes on 5q14.3 to the highest expressed genes (*Actb, Gapdh*), *Tbx1* expression, and the lowest expressed genes (*Il6, Olfr299*) in this tissue as shown in Figure 3C. All of the genes are expressed in the pharyngeal apparatus, albeit *Gpr98* is expressed at the lowest level, as can be also seen in Figure 3B.

The Encyclopedia of DNA Elements functional genomics data were examined in the 104.7 kb region of LD with SNP, rs12519770, to identify possible regulatory regions. We found 3 possible regulatory regions defined by binding of multiple transcription factors, with one close to rs12519770 (Figure V in the Data Supplement). None of the SNPs that were genotyped in our study lie within these putative regulatory regions; however, critical embryonic regulatory regions could be different, and they have not yet been defined.

Discussion

We identified genome-wide significant associations between TOF and several SNPs in an intron of *GPR98* in our 22q11.2DS cohort. We narrowed the associated region to a 104.7 kb interval. This interval may harbor functional variants that are in LD with the associated SNPs or noncoding variants that regulate the expression of genes within a broad TAD on 5q14.3.

GPR98 contains 90 exons, spans over 610 kb, and encodes a member of the adhesion-G protein-coupled receptor family of receptors. The GPR98 protein binds calcium and is weakly expressed early in mouse embryonic development at E9.5, but it becomes more strongly expressed in the future brain and neural tube by E10.5. It is the largest of the 7-transmembrane receptors and has important functions in hearing and vision.[37,38] Recessive mutations cause Usher syndrome type 2C (Mendelian Inheritance in Man No. 605472), which is characterized by congenital hearing loss and progressive retinitis pigmentosa.[39] There are multiple splice variants present in *GPR98*, and it is not known if all isoforms have similar functions. Thus, it is possible that ≥1 splice variants could have a function in neural crest cells deriving from the neural tube. Neural crest cells are a migratory population of progenitor cells in the pharyngeal apparatus, which contribute to cardiac OFT septation. There are no reports of a possible function of *GPR98* in the cardiovascular system and no known connections to human TOF.

CETN3 is another gene of note in the 5q14.3 region because it encodes a centrin protein that functions in the cytoskeleton of centrosomes and cilia.[40] Cilia are critically important for conferring left right assymetry during embryonic development and when disrupted is associated with human cardiac anomalies[22] Because laterality defects are not commonly found in association with 22q11.2DS, more work would need

Table 2. Top SNPs in *GPR98* Associated With TOF

SNP	Chr:Position (bp)	Cluster	Risk Allele	Other Allele	RAF	OR (95% CI)	*P* Value	Conditional *P* Value*	Conditional *P* Value†
rs12519770	5:90073277	1	A	G	0.5	1.69 (1.39–2.06)	2.98E-08	NA	NA
rs6889138	5:90335635	2	G	A	0.24	1.68 (1.35–2.10)	1.72E-07	0.00199523	NA
rs6893710	5:90058041	3	C	T	0.03	4.05 (2.27–7.24)	1.04E-06	3.92E-05	6.11E-05

The top genotyped single-nucleotide polymorphisms (SNPs) in each cluster of SNPs identified for tetralogy of Fallot (TOF) are shown. Chr:Position (bp), chromosome and positions are indicated according to NCBI36/hg18 (March 2006), and allele coding was based on the positive strand. The cluster numbers that the SNPs belong to are indicated. *GPR98* indicates G-protein–coupled receptor V1; NA, not associated; OR, odds ration; and RAF, risk allele frequency in the whole cohort.

*Conditional *P* value: analysis was conditioned on SNP rs12519770.

†Conditional *P* value: individually evaluated the association of TOF with 1344 SNPs in the 5q14.3 region (Table VI in the Data Supplement).

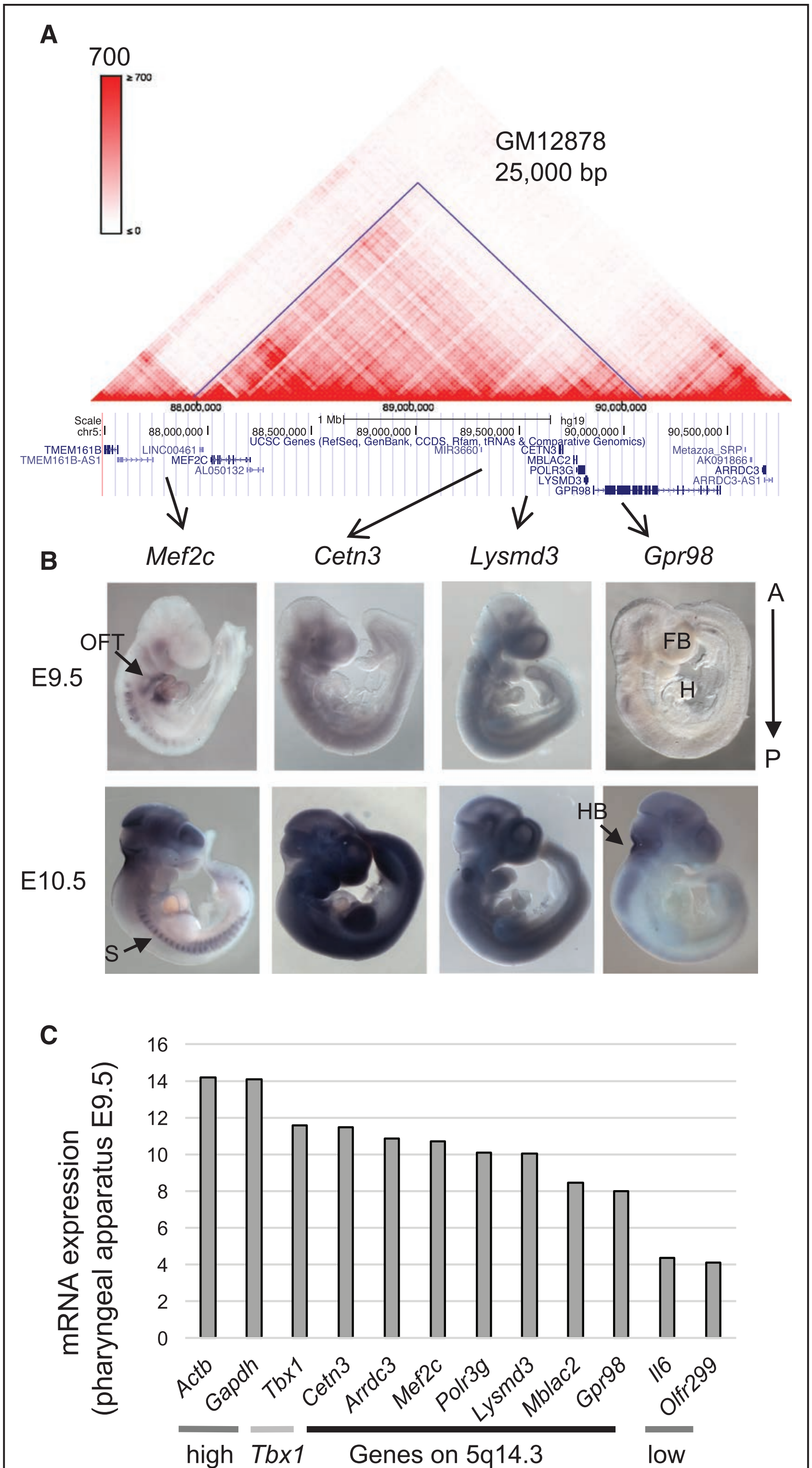

Figure 3. Representative heatmap of chromatin conformation for the *MEF2C* (Myocyte-specific enhancer factor 2C)-*GPR98* (G-protein–coupled receptor V1) interval and in situ hybridization of selected genes in mouse embryos. **A**, Chromatin conformation map extracted from Hi-C data for the human lymphoblastoid cell line, GM12878, from the HapMap sample set.[29] Similar results were found using other cell types (Figure III in the Data Supplement). The intensity of each pixel represents the normalized number of contacts between a pair of loci. The purple line marking the triangle depicts the contact domain[31] between the *GPR98* and the *MEF2C* locus. **B**, Lateral views of E9.5 and E10.5 wild-type embryos following in situ hybridization of antisense probes for *Gpr98*, *Cetn3* (Centrin 3), *Mef2c,* and *Lysmd3* (LysM, putative peptidoglycan-binding, domain containing 3). Purple color indicates mRNA expression. **C**, Bar graph of relative gene expression levels from Affymetrix microarrays for the microdissected distal pharyngeal apparatus from E9.5 stage mouse embryos (*y* axis). Individual genes are shown for comparison, including representative housekeeping genes with high expression (*Actb*, *Gapdh*), genes on chromosome 5q14.3 (*Cetn3, Arrdc3, Mef2c, Polr3g, Lysmd3, Mblac2,* and *Gpr98*) and representative low expressing genes (*Il6, Olfr299*), in descending order according to expression. A indicates anterior; Arrdc3 indicates Arrestin domain containing 3; Cetn3, Centrin 3; FB, forebrain; H, heart; HB, hindbrain; Lysmd3, LysM, putative peptidoglycan-binding, domain containing 3; Mblac2, Metallo-β-lactamase domain containing 2; OFT, cardiac outflow tract; P, posterior; Polr3g, RNA polymerase III subunit G; and S, somites.

to be done to provide additional support of the role of *CETN3* as a modifier of TOF in these individuals.

Among the remaining 5 genes (*MEF2C, POLR3G, MBLAC2, LYSMD3,* and *ARRDC3*), *MEF2C* is of particular interest because it encodes a transcription factor required in the second heart field mesoderm of the pharyngeal apparatus during embryogenesis for cardiac OFT development.[41,42] Because haploinsufficiency of *TBX1* is important for 22q11.2DS, and both *MEF2C* and *TBX1* are expressed in the second heart field progenitor cells, it is possible that *TBX1* might act in the same genetic pathway as *MEF2C*. Further, studies in mouse models indicate that *Tbx1* may be a negative regulator of *Mef2c*.[43] This suggests that genetic variants in the 5q14.3 locus that may be associated with *MEF2C* expression levels could act as genetic modifiers of 22q11.2DS, with the caveat that causation will require direct experimental support. Further, ISL1 transcription factor (ISL LIM homeobox 1 transcription factor) and GATA (transcription factor) proteins bind and regulate both *Mef2c* and *Nkx2-5* (NK2 homeobox 5 transcription factor) cardiac development genes.[44] One hypothesis to test in the future by direct experimentation would be that rare DNA variants in *MEF2C*, *ISL1,* or *NKX2-5* affect cardiac OFT formation in individuals with 22q11.2DS.

In recent years, higher-order chromatin structure technologies have demonstrated that chromatin interactions occur in a nonrandom manner along the chromosome arms, which are separated into regions of highly interacting chromatin.[30,45–47] Relevant to this, the 104.7 kb region found with top associated SNPs to TOF shows a possible regulatory connection with *MEF2C* located 2 Mb upstream of *GPR98* using available Hi-C, chromatin conformation data.[29] One limitation of using Hi-C data to draw conclusions about gene regulation is that these data indicate that the 2 genes may reside within the same topological region but do not prove that there is a definitive regulatory connection. Further chromatin conformation data in cell progenitors relevant to cardiac OFT development, followed by direct experimental approaches, will be required to define this interaction further. In addition, a better understanding of the regulatory landscape in this region will be needed to identify the mechanism(s) responsible for the association to TOF we have observed in the 5q14.3 region.

Further support for *MEF2C,* as a possible modifier gene, comes from a recent GWAS of circulating VEGF (vascular endothelial growth factor) levels in blood in adults.[48] Although this is a study of adults, factors that regulate VEGF levels may be similar throughout life and could affect fetal development. Previously, it was found that absence of one of the VEGF isoforms causes a phenocopy of 22q11.2DS in mouse models.[49] In this recent GWAS of VEGF levels in adults, *MEF2C* and *JMJD1C* (Jumonji domain containing 1C) were found among the 6 loci with significant association to VEGF levels. *JMJD1C* is relevant because we previously found significant enrichment of rare predicted exonic variants in *JMJD1C* in whole-exome sequence from 184 22q11.2DS subjects in which the cases were enriched for TOF.[50] This provides a potential biological connection between common and previous rare variant analyses for 22q11.2DS. However, in regards to the general population, neither haploinsufficiency nor mutation of *MEF2C* in humans has been associated with any type of CHD thus far.[51–53] Further proof that *MEF2C* is a cardiac disease gene in the general population will require future genetic studies.

The number of samples we obtained, even after 25 years of collection, is quite small for a GWAS of a complex trait. Thus, one of the limitations of the study was the lack of a true replication cohort. Another limitation of our study is possible population stratification. The majority of our cohort was self-reported as white. Nevertheless, a subset had different ethnicities requiring statistical correction in the analysis. Further, we examined each ethnicity separately and combined each group by performing a meta-analysis. We found that the top SNP in *GPR98* was still statistically significant. In addition, the odds ratio was in the same direction in all the subcohorts, supporting our findings despite the limitations.

In this report, we used available bioinformatic data to interpret our data. However, this study did not provide proof of causation. This will need to be done by performing functional studies including development of animal models. Despite the limitations of the study, this is the first GWAS to identify common variants that may modify the cardiac phenotype in a large cohort of individuals with 22q11.2DS.

Conclusions

A GWAS of TOF in 22q11.2DS has identified a significant locus on 5q14.3 harboring potential genetic risk factors. Several genes reside in this locus including *MEF2C* that is a known gene for cardiac OFT development in animal models. Further work needs to be done to ascertain whether *MEF2C* or other genes in this region act as modifiers of TOF in 22q11.2DS.

Acknowledgments

We thank the families with 22q11.2DS who provided DNA and clinical information. We acknowledge the Genomics and Molecular Cytogenetics Cores at Einstein. We also acknowledge Mark Zeffren, Nousin Haque, Antoneta Preldakaj, and Francisco Ujueta for project management and John Bruppacher, Dan Arroyo, Michael Gleeson, Dominique Calandrillo, and Frédérique Bena for technical support at Einstein. We also greatly appreciate the effort of Dr Frédérique Bena who works with S.E. and S.E.A. (Institute of Genetics and Genomics of Geneva, Switzerland). URLs: Golden Helix software: http://goldenhelix.com/; EIGENSOFT: http://genetics.med.harvard.edu/reich/Reich_Lab/Software.html; QUANTO: http://biostats.usc.edu/Quanto.html; IMPUTE: https://mathgen.stats.ox.ac.uk/impute/impute.html; HapMap: http://hapmap.ncbi.nlm.nih.gov/; LocusZoom: http://csg.sph.umich.edu/locuszoom/; PLINK 1.07: http://zzz.bwh.harvard.edu/plink/; R statistical software: http://www.r-project.org/; 1000 Genomes Project: http://www.1000genomes.org/; META: https://mathgen.stats.ox.ac.uk/genetics_software/meta/meta.html; SNPTEST v2.5.2: https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html.

Appendix

From the Department of Genetics (T.G., J.H.C., H.N., C.L.C., T.W., B.E.M.) and Department of Epidemiology and Population Health (T.W.), Albert Einstein College of Medicine, Bronx, NY; Center for Human Genetics, Facultad de Medicina Clinica Alemana Universidad del Desarrollo, Santiago, Chile (G.M.R.); Division of Human Genetics (D.M.M.M., E.E.M., E.Z., B.S.E.), Division of Cardiology (E.G.), and Department of Pediatrics (E.G.), Children's Hospital of Philadelphia, Perelman School of Medicine, University of Pennsylvania, Philadelphia; Department of Genetics, Wroclaw Medical University, Poland (A.B.); Clinical Genetics Research Program, Center for Addiction and Mental Health and Department of Psychiatry, University of Toronto (A.S.B., E.W.C.C.); Dalglish Family 22q Clinic, Department of Psychiatry and Toronto General Research Institute, University Health Network, Canada (A.S.B.); Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada (A.S.B.); Center for Human Genetics, University of Leuven (KU Leuven), Belgium (A.S., J.V., K.D.); The Child Psychiatry Division, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel (D.G.); Sackler Faculty of Medicine and Sagol School of Neuroscience, Tel Aviv University, Israel (D.G., M.C., E.M.); Felsenstein Medical Research Center, Sackler Faculty of Medicine, Tel Aviv University, Petah Tikva, Israel (M.C., E.M.); Developmental Imaging and Psychopathology Lab, University of Geneva School of Medicine, Switzerland (M.S., S.E.); Department of Genetic Medicine, UNIGE and iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva Medical Center, Switzerland (S.E.A.); Marcus Autism Center, Children's Healthcare of Atlanta, GA (K.C.); Division of Pediatric Cardiovascular

Surgery, Children's Hospital of Wisconsin, Milwaukee (A.T.-M., M.E.M.); Department of Surgery, Medical College of Wisconsin, Milwaukee (A.T.-M., M.E.M.); Department of Medical Genetics, Bambino Gesù Hospital, Rome, Italy (M.C.D., B.D.); Department of Pediatrics, La Sapienza University of Rome, Italy (B.M.); Department of Medical Genetics, Aix Marseille University, APHM, GMGF, Timone Hospital, France (N.P., T.B.); Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California at Los Angeles (L.K.-W., C.E.B.); Department of Genetics, Polish Mother's Memorial Hospital, Research Institute, Łódź, Poland (M.P., W.H.); Department of Cardiology and Division of Genetics, Boston Children's Hospital, MA (A.E.R.); M.I.N.D. Institute and Department of Psychiatry and Behavioral Sciences (F.T.) and M.I.N.D. Institute and Department of Biochemistry and Molecular Medicine (T.J.S.), University of California, Davis; Department of Psychiatry and Psychology, University of Maastricht, The Netherlands (E.D.A.V.D., T.A.v.A.); Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, NY (T.A.v.A., W.R.K.); Department of Human Genetics, Emory University School of Medicine, Atlanta, GA (H.R.J., D.J.C.); Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Atlanta, GA (H.R.J.); and Human Genetics Center and Department of Epidemiology, Human Genetics and Environmental Sciences, UTHealth School of Public Health, Houston, TX (A.J.A., L.E.M.).

Sources of Funding

This work was supported by National Institutes of Health grants R01 HL084410 (Dr Emanuel, Dr Morrow, D.M. McDonald McGinn, Dr Guo, A.S. Bassett), P01 HD070454 (Dr Goldmuntz, Dr Mitchell, Dr Agopian, Emanuel, D.M. McDonald McGinn, Dr Mlynarski, Dr Guo, Dr Wang, Dr Nomaru, Dr Campbell), U01 MH101720 (Dr Emanuel, Dr Morrow, D.M. McDonald McGinn, Dr Repetto, Dr Bassett, Dr Bassett, Dr Swillen, Dr Gothelf, Dr Eliez, Dr Tassone, Dr Philip, Dr Bearden, Dr Simon, E.D.A. van Duin, Dr van Amelsvoort, Dr Kates, Dr Guo, Dr Wang), R21HL118637 (Drs Wang, Morrow, Guo, Goldmuntz), T32GM007491-41 (JHC). This work was supported by the American Heart Association, grant 14PRE199800006 (Dr Chung). Dr Repetto was supported by the *Fondo National de Desarrollo Cientifico y Tecnologico*-Chile (grants 1100131 and 1130392). Dr Bassett was supported by the Dalglish Chair in 22q11.2 Deletion Syndrome, the Canada Research Chair in Schizophrenia Genetics and Genomic Disorders, Canadian Institutes of Health Research funding (Missionary Orientation Program-97800 and Missionary Orientation Program-89066), and the University of Toronto McLaughlin Centre. Dr Bearden was supported by National Institutes of Health grant R01 MH085903. Dr Simon was supported by National Institutes of Health grant R01 HD042974. Dr Mitchell was supported by National Institutes of Health grant R21 HD060309-01. Dr Eliez was supported by the Swiss National Science Foundation (Swiss National Science Foundation 324730_121996; Swiss National Science Foundation 324730_144260).

Disclosures

None.

References

1. McDonald-McGinn DM, Sullivan KE, Marino B, Philip N, Swillen A, Vorstman JA, et al. 22q11.2 deletion syndrome. *Nat Rev Dis Primers*. 2015;1:15071. doi: 10.1038/nrdp.2015.71.
2. Lindsay EA, Greenberg F, Shaffer LG, Shapira SK, Scambler PJ, Baldini A. Submicroscopic deletions at 22q11.2: variability of the clinical picture and delineation of a commonly deleted region. *Am J Med Genet*. 1995;56:191–197. doi: 10.1002/ajmg.1320560216.
3. Lindsay EA, Goldberg R, Jurecic V, Morrow B, Carlson C, Kucherlapati RS, et al. Velo-cardio-facial syndrome: frequency and extent of 22q11 deletions. *Am J Med Genet*. 1995;57:514–522. doi: 10.1002/ajmg.1320570339.
4. Morrow B, Goldberg R, Carlson C, Das Gupta R, Sirotkin H, Collins J, et al. Molecular definition of the 22q11 deletions in velo-cardio-facial syndrome. *Am J Hum Genet*. 1995;56:1391–1403.
5. Burn J Goodship J. Developmental genetics of the heart. *Curr Opin Genet Dev*. 1996;6:322–325.
6. Kelly RG. The second heart field. *Curr Top Dev Biol*. 2012;100:33–65. doi: 10.1016/B978-0-12-387786-4.00002-6.
7. Goldmuntz E, Clark BJ, Mitchell LE, Jawad AF, Cuneo BF, Reed L, et al. Frequency of 22q11 deletions in patients with conotruncal defects. *J Am Coll Cardiol*. 1998;32:492–498.
8. Chessa M, Butera G, Bonhoeffer P, Iserin L, Kachaner J, Lyonnet S, et al. Relation of genotype 22q11 deletion to phenotype of pulmonary vessels in tetralogy of Fallot and pulmonary atresia-ventricular septal defect. *Heart*. 1998;79:186–190.
9. Lindsay EA, Vitelli F, Su H, Morishima M, Huynh T, Pramparo T, et al. Tbx1 haploinsufficieny in the DiGeorge syndrome region causes aortic arch defects in mice. *Nature*. 2001;410:97–101. doi: 10.1038/35065105.
10. Merscher S, Funke B, Epstein JA, Heyer J, Puech A, Lu MM, et al. TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. *Cell*. 2001;104:619–629.
11. Jerome LA Papaioannou VE. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. *Nature Genet*. 2001;27:286–291.
12. Racedo SE, Hasten E, Lin M, Devakanmalai GS, Guo T, Ozbudak EM, et al. Reduced dosage of β-catenin provides significant rescue of cardiac outflow tract anomalies in a Tbx1 conditional null mouse model of 22q11.2 deletion syndrome. *PLoS Genet*. 2017;13:e1006687. doi: 10.1371/journal.pgen.1006687.
13. Guo T, McDonald-McGinn D, Blonska A, Shanske A, Bassett AS, Chow E, et al; International Chromosome 22q11.2 Consortium. Genotype and cardiovascular phenotype correlations with TBX1 in 1,022 velo-cardio-facial/DiGeorge/22q11.2 deletion syndrome patients. *Hum Mutat*. 2011;32:1278–1289. doi: 10.1002/humu.21568.
14. Mlynarski EE, Sheridan MB, Xie M, Guo T, Racedo SE, McDonald-McGinn DM, et al; International Chromosome 22q11.2 Consortium. Copy-Number Variation of the glucose transporter gene SLC2A3 and congenital heart defects in the 22q11.2 deletion syndrome. *Am J Hum Genet*. 2015;96:753–764. doi: 10.1016/j.ajhg.2015.03.007.
15. León LE, Benavides F, Espinoza K, Vial C, Alvarez P, Palomares M, et al. Partial microduplication in the histone acetyltransferase complex member KANSL1 is associated with congenital heart defects in 22q11.2 microdeletion syndrome patients. *Sci Rep*. 2017;7:1795. doi: 10.1038/s41598-017-01896-w.
16. Tomita-Mitchell A, Mahnke DK, Struble CA, Tuffnell ME, Stamm KD, Hidestrand M, et al. Human gene copy number spectra analysis in congenital heart malformations. *Physiol Genomics*. 2012;44:518–541. doi: 10.1152/physiolgenomics.00013.2012.
17. Browning BL Yu Z. Simultaneous genotype calling and haplotype phasing improves genotype accuracy and reduces false-positive associations for genome-wide association studies. *Am J Hum Genet*. 2009;85:847–861.
18. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. *Nat Methods*. 2011;9:179–181. doi: 10.1038/nmeth.1785.
19. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10:5–6. doi: 10.1038/nmeth.2307.
20. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet*. 2012;44:955–959. doi: 10.1038/ng.2354.
21. Liu JZ, Tozzi F, Waterworth DM, Pillai SG, Muglia P, Middleton L, et al; Wellcome Trust Case Control Consortium. Meta-analysis and imputation refines the association of 15q25 with smoking quantity. *Nat Genet*. 2010;42:436–440. doi: 10.1038/ng.572.
22. Willaredt MA, Gorgas K, Gardner HA, Tucker KL. Multiple essential roles for primary cilia in heart development. *Cilia*. 2012;1:23. doi: 10.1186/2046-2530-1-23.
23. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26:2336–2337. doi: 10.1093/bioinformatics/btq419.
24. Johnston HR, Chopra P, Wingo TS, Patel V, Epstein MP, Mulle JG, et al; International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. PEMapper and PECaller provide a simplified approach to whole-genome sequencing. *Proc Natl Acad Sci USA*. 2017;114:E1923–E1932. doi: 10.1073/pnas.1618065114.
25. Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*. 2014;30:1006–1007. doi: 10.1093/bioinformatics/btt730.

26. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev*. 1999;13:2465–2477.
27. Schoenfelder S, Clay I, Fraser P. The transcriptional interactome: gene expression in 3D. *Curr Opin Genet Dev*. 2010;20:127–133. doi: 10.1016/j.gde.2010.02.002.
28. Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*. 2010;467:430–435. doi: 10.1038/nature09380.
29. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518:331–336. doi: 10.1038/nature14222.
30. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. 2012;485:376–380. doi: 10.1038/nature11082.
31. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159:1665–1680. doi: 10.1016/j.cell.2014.11.021.
32. McMillan DR, Kayes-Wandover KM, Richardson JA, White PC. Very large G protein-coupled receptor-1, the largest known cell surface protein, is highly expressed in the developing central nervous system. *J Biol Chem*. 2002;277:785–792. doi: 10.1074/jbc.M108929200.
33. Kumamoto N, Gu Y, Wang J, Janoschka S, Takemaru K, Levine J, et al. A role for primary cilia in glutamatergic synaptic integration of adult-born neurons. *Nat Neurosci*. 2012;15:399–405, S1. doi: 10.1038/nn.3042.
34. Middendorp S, Küntziger T, Abraham Y, Holmes S, Bordes N, Paintrand M, et al. A role for centrin 3 in centrosome reproduction. *J Cell Biol*. 2000;148:405–416.
35. Verzi MP, McCulley DJ, De Val S, Dodou E, Black BL. The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev Biol*. 2005;287:134–145. doi: 10.1016/j.ydbio.2005.08.041.
36. Lin Q, Schwarz J, Bucana C, Olson EN. Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science*. 1997;276:1404–1407.
37. Sun JP, Li R, Ren HZ, Xu AT, Yu X, Xu ZG. The very large G protein coupled receptor (Vlgr1) in hair cells. *J Mol Neurosci*. 2013;50:204–214. doi: 10.1007/s12031-012-9911-5.
38. McMillan DR White PC. Studies on the very large G protein-coupled receptor: from initial discovery to determining its role in sensorineural deafness in higher animals. *Adv Exp Med Biol*. 2010;706:76–86.
39. Weston MD, Luijendijk MW, Humphrey KD, Möller C, Kimberling WJ. Mutations in the VLGR1 gene implicate G-protein signaling in the pathogenesis of Usher syndrome type II. *Am J Hum Genet*. 2004;74:357–366. doi: 10.1086/381685.
40. Laoukili J, Perret E, Middendorp S, Houcine O, Guennou C, Marano F, et al. Differential expression and cellular distribution of centrin isoforms during human ciliated cell differentiation in vitro. *J Cell Sci*. 2000;113 (pt 8):1355–1364.
41. Potthoff MJ, Olson EN. MEF2: a central regulator of diverse developmental programs. *Development*. 2007;134:4131–4140.
42. Barnes RM, Harris IS, Jaehnig EJ, Sauls K, Sinha T, Rojas A, et al. MEF2C regulates outflow tract alignment and transcriptional control of Tdgf1. *Development*. 2016;143:774–779. doi: 10.1242/dev.126383.
43. Pane LS, Zhang Z, Ferrentino R, Huynh T, Cutillo L, Baldini A. Tbx1 is a negative modulator of Mef2c. *Hum Mol Genet*. 2012;21:2485–2496. doi: 10.1093/hmg/dds063.
44. Takeuchi JK, Mileikovskaia M, Koshiba-Takeuchi K, Heidt AB, Mori AD, Arruda EP, et al. Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. *Development*. 2005;132:2463–2474. doi: 10.1242/dev.01827.
45. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. 2012;485:381–385. doi: 10.1038/nature11049.
46. Hou C, Li L, Qin ZS, Corces VG. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Mol Cell*. 2012;48:471–484. doi: 10.1016/j.molcel.2012.08.031.
47. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*. 2012;148:458–472. doi: 10.1016/j.cell.2012.01.010.
48. Choi SH, Ruggiero D, Sorice R, Song C, Nutile T, Vernon Smith A, et al. Six novel loci associated with circulating VEGF levels identified by a meta-analysis of genome-wide association studies. *PLoS Genet*. 2016;12:e1005874. doi: 10.1371/journal.pgen.1005874.
49. Stalmans I, Lambrechts D, De Smet F, Jansen S, Wang J, Maity S, et al. VEGF: a modifier of the del22q11 (DiGeorge) syndrome? *Nat Med*. 2003;9:173–182. doi: 10.1038/nm819.
50. Guo T, Chung JH, Wang T, McDonald-McGinn DM, Kates WR, Hawuła W, et al. Histone modifier genes alter conotruncal heart phenotypes in 22q11.2 deletion syndrome. *Am J Hum Genet*. 2015;97:869–877. doi: 10.1016/j.ajhg.2015.10.013.
51. Zweier M, Gregor A, Zweier C, Engels H, Sticht H, Wohlleber E, et al; Cornelia Kraus. Mutations in MEF2C from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish MECP2 and CDKL5 expression. *Hum Mutat*. 2010;31:722–733. doi: 10.1002/humu.21253.
52. Hotz A, Hellenbroich Y, Sperner J, Linder-Lucht M, Tacke U, Walter C, et al. Microdeletion 5q14.3 and anomalies of brain development. *Am J Med Genet A*. 2013;161A:2124–2133. doi: 10.1002/ajmg.a.36020.
53. Bienvenu T, Diebold B, Chelly J, Isidor B. Refining the phenotype associated with MEF2C point mutations. *Neurogenetics*. 2013;14:71–75. doi: 10.1007/s10048-012-0344-7.

CLINICAL PERSPECTIVE

The 22q11.2 deletion syndrome (DiGeorge syndrome/velocardiofacial syndrome) occurs in 1 of 4000 live births, and 60% to 70% of affected individuals have congenital heart disease, ranging from mild to severe. In our cohort of 1472 subjects with 22q11.2 deletion syndrome, a total of 62% (n=906) have congenital heart disease and 36% (n=326) of these have tetralogy of Fallot, comprising the largest subset of severe congenital heart disease in the cohort. One of the main questions clinicians are interested in understanding is the basis for variable phenotypic expression. To address this question, using 22q11.2 deletion syndrome as a model, we performed a genome-wide association study and found tetralogy of Fallot was significantly associated with a genetic interval on chromosome 5q14.3. The associated region is within an intron of GPR98. This intron may contain regulatory sequences that could effect expression of *GPR98* or other genes in the region. Among them, *MEF2C* (Myocyte-specific enhancer factor 2C) is the only gene that is known to affect heart development. On the basis of this work, common DNA variants on 5q14.3 may explain, in part, why phenotypic variation occurs despite the fact that they have the same deletion. Further work will need to be done to determine whether the same genetic variants can alter *MEF2C* expression and risk for congenital heart disease in the general population or alter clinical outcomes. In the future, understanding how genetic variations influence phenotype in both genetic syndromes and in more common disease will make it possible to improve genetic counseling, rehabilitation, clinical outcomes, and in the creation of novel therapeutics.