RESEARCH ARTICLE

Human Mutation

OFFICIAL JOURNAL HGVS HUMAN GENOME VARIATION SOCIETY www.hgvs.org

Whole-Genome Sequencing and Integrative Genomic Analysis Approach on Two 22q11.2 Deletion Syndrome Family Trios for Genotype to Phenotype Correlations

Jonathan H. Chung,[1†] Jinlu Cai,[2†] Barrie G. Suskin,[3] Zhengdong Zhang,[1] Karlene Coleman,[4] and Bernice E. Morrow[1*]

[1]*Department of Genetics, Albert Einstein College of Medicine, Bronx, New York;* [2]*Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York;* [3]*Department of Obstetrics & Gynecology and Women's Health, Montefiore Medical Center, Bronx, New York;* [4]*Children's Healthcare of Atlanta at Egleston, Atlanta, Georgia*

Communicated by Nancy B. Spinner

Received 15 January 2015; accepted revised manuscript 1 May 2015.
Published online 15 May 2015 in Wiley Online Library (www.wiley.com/humanmutation). DOI: 10.1002/humu.22814

ABSTRACT: The 22q11.2 deletion syndrome (22q11DS) affects 1:4,000 live births and presents with highly variable phenotype expressivity. In this study, we developed an analytical approach utilizing whole-genome sequencing (WGS) and integrative analysis to discover genetic modifiers. Our pipeline combined available tools in order to prioritize rare, predicted deleterious, coding and noncoding single-nucleotide variants (SNVs), and insertion/deletions from WGS. We sequenced two unrelated probands with 22q11DS, with contrasting clinical findings, and their unaffected parents. Proband P1 had cognitive impairment, psychotic episodes, anxiety, and tetralogy of Fallot (TOF), whereas proband P2 had juvenile rheumatoid arthritis but no other major clinical findings. In P1, we identified common variants in *COMT* and *PRODH* on 22q11.2 as well as rare potentially deleterious DNA variants in other behavioral/neurocognitive genes. We also identified a de novo SNV in *ADNP2* (NM_014913.3:c.2243G>C), encoding a neuroprotective protein that may be involved in behavioral disorders. In P2, we identified a novel nonsynonymous SNV in *ZFPM2* (NM_012082.3:c.1576C>T), a known causative gene for TOF, which may act as a protective variant downstream of *TBX1*, haploinsufficiency of which is responsible for congenital heart disease in individuals with 22q11DS.

Hum Mutat 36:797–807, 2015. © 2015 Wiley Periodicals, Inc.

KEY WORDS: 22q11.2 deletion syndrome; integrative analysis; schizophrenia psychosis; tetralogy of Fallot; juvenile rheumatoid arthritis

Introduction

The 22q11.2 deletion syndrome (22q11DS), also known as velocardio-facial syndrome or DiGeorge syndrome (MIMs #192430, 188400) occurs in 1:4,000 live births and is caused by a similar sized 3 Mb hemizygous deletion on chromosome 22q11.2 in the majority of patients [Morrow et al., 1995; Shaikh et al., 2000]. One of the most striking features of 22q11DS is the highly variable clinical presentation or phenotype expressivity of affected individuals. A potential source of the variable expressivity could be genetic variants. It has been hypothesized that the 22q11.2 deletion could act to unmask recessive mutations or sensitize the population to genetic modifiers with small effects [Stalmans et al., 2003; McDonald-McGinn et al., 2013]. A large proportion of individuals with 22q11DS have learning disabilities or psychiatric disorders. For example, approximately 25% of adults with 22q11DS have schizophrenia (SCZD; MIM #181500), which is 30-fold higher than what occurs in the general population [Schneider et al., 2014].

In addition to cognitive impairment and psychiatric disorders, over 70% of individuals with 22q11DS have congenital heart disease (CHD), primarily of the conotruncal type, including tetralogy of Fallot (TOF; MIM #187500). TOF occurs in approximately 35% of individuals with 22q11DS [Momma et al., 1996] as compared with five in 10,000 live births in the general population. Mouse genetic studies identified *Tbx1* (T-box 1; Entrez Gene ID: 21380), encoding a T-box transcription factor as one of the key genes in the 22q11.2 region as being responsible for CHD [Jerome and Papaioannou, 2001; Lindsay, 2001; Merscher et al., 2001]. However, mutation analysis of the remaining allele of *TBX1* (MIM #602054) in human subjects with varying cardiac defects failed to identify potential modifier alleles that could explain this variability [Guo et al., 2011]. This leaves open the possibility that genetic modifiers exist in other genes on 22q11.2 or elsewhere in the genome.

There are many additional clinically important features in 22q11DS that occur at a lower frequency [Cancrini et al., 2014]. Some anomalies of particular interest are immunodeficiency and autoimmune disease, including juvenile rheumatoid arthritis (JRA; MIM #604302) [Rasmussen et al., 1996; Gennery, 2013]. JRA is seen in approximately 2% of 22q11DS patients, but 0.12% in the general population [Cobb et al., 2014].

Personal genomics approaches to understand an individual's susceptibility to disease are now possible with next-generation sequencing methods. Whole-genome sequencing (WGS) provides the potential to directly identify causative coding and noncoding variants. However, interpretation of the vast amount of data can be a limiting factor. This task can be aided by utilizing integrative genomic analysis methods, which combine knowledge from multiple data sources to identify potentially causative variants [Wang et al., 2010; Khurana et al., 2013]. Integrative approaches have successfully been used to identify primary causative mutations in studies of

Additional Supporting Information may be found in the online version of this article.

†Jonathan H. Chung and Jinlu Cai made equal contribution to this article.

*Correspondence to: Bernice E. Morrow, Department of Genetics, Albert Einstein College of Medicine, Bronx, NY, USA. E-mail: Bernice.morrow@einstein.yu.edu

Contract grant sponsors: Department of Genetics, Human Genetics Program Endowment; American Heart Association (AHA) 14PRE199800006; NIH P01 HD070454 and R01 DC05186.

various complex disorders including cancer and cardiovascular disease [Khurana et al., 2013; Ware et al., 2013]. Most tools annotate coding variants with relatively few that are able to annotate noncoding genetic variations. However, recent efforts to characterize the noncoding genome, such as the ENCODE project and Epigenomics Roadmap Project have made significant progress toward identifying functional variants from the noncoding elements [Bromberg, 2013; Khurana et al., 2013]. These new tools and resources facilitate our ability to identify potential genetic modifiers in individuals with 22q11DS.

In this study, we performed WGS on DNA from two trios with 22q11DS in order to test whether integrative genomic methods are able to discover genetic modifiers that could explain variable phenotypes in these affected individuals. This work may lead the way to future larger studies to find genetic modifiers of this or other genomic disorders occurring with variable expressivity.

Materials and Methods

Subject Recruitment and Phenotype Information

Informed consent for genetic analyses was obtained from all individuals in the study. The protocol was approved by the Internal Review Board at Albert Einstein College of Medicine (IRB: 1999-201-047). The two probands were diagnosed with 22q11DS after fluorescence in situ hybridization (FISH) testing for the 22q11.2 deletion. Complete family and medical history was obtained for two probands with de novo 22q11.2 deletions and their unaffected parents (Fig. 1; Supp. Table S1).

Proband 1 (P1), now a 32-year-old male, was referred for genetics evaluation at age 13 years, with characteristic clinical features of 22q11DS. He was the product of an uneventful pregnancy of a nonconsanguineous couple. He had developmental delays in motor and speech milestones, learning disabilities, and attention-deficit hyperactivity disorder (MIM #143465). At age 27 years, P1 developed psychotic symptoms, including visual hallucinations, and was diagnosed with other specified SCZD spectrum and other psychotic disorder (DSM-5 298.8) after review of clinical charts. He is currently being treated with venlafaxine and olanzapine. A SCID (structured clinical interview for DSM disorders) was performed in November of 2014 and the proband was diagnosed with clinical anxiety but did not present with psychosis, indicating a good response to antipsychotic medication. A paternal great uncle had a ventricular septal defect and SCZD, whereas a paternal aunt died shortly after birth from pneumonia, and possibly a heart defect (Fig. 1A). On physical examination at age 13 years, the proband was obese and had the typical facies seen in individuals with 22q11DS. A high arch and bifid uvula were medically confirmed. Fingers were tapered and the hands had multiple fine palmar creases with fifth digit camptodactyly, as is typical with 22q11DS. He also had joint laxity. He had TOF, as diagnosed by echocardiogram, but no other cardiovascular malformations. Hypogonadism was also observed. FISH confirmed a de novo 22q11.2 deletion (data not shown). Detailed phenotype information can be found in Supp. Table S1.

Proband 2 (P2), now a 15-year-old female, was referred for a genetics evaluation at age 8 years from a craniofacial center and had a de novo 22q11.2 deletion as determined by FISH. She was the product of an uneventful pregnancy of a nonconsanguineous couple. Family history was unremarkable except for unknown mental illness and cancer in the paternal grandfather and grandmother, respectively (Fig. 1). She performed at an average level in elementary school. She had a history of hypernasal speech, velopharyngeal incompetence, occipitalization of the atlas, and spina bifida occulta as typical in 22q11DS. At age 2 years, she was diagnosed with JRA and treated with methotrexate. The JRA is now in remission. She had occasional emotional lability, but met all of her milestones appropriately and performed well in her classes. On physical examination, she was found to be in the 5th percentile for height and weight. Other than a slightly cylindrical appearance of her nose, she had normal facial features. There was no strong history of mental illness in P2. However, she is still within the risk age group for mental illness in 22q11DS. Detailed phenotype information can be found in Supp. Table S1.

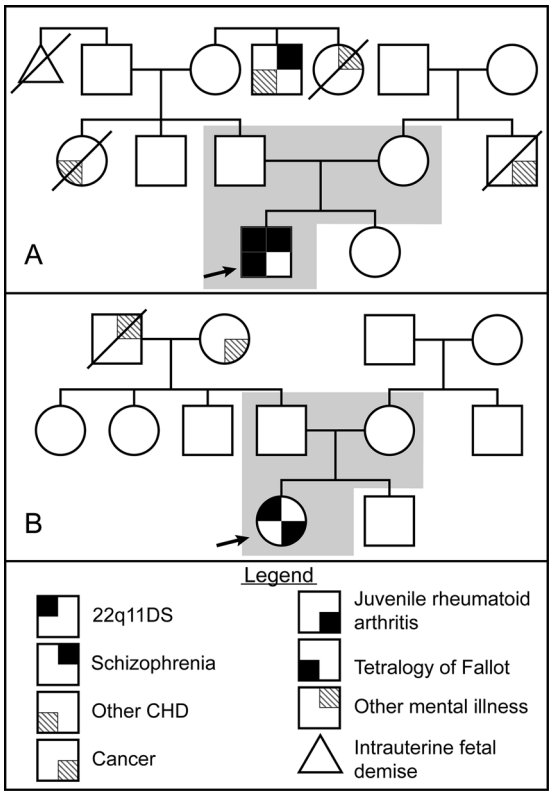

Figure 1. Pedigree and family history for the P1 and P2 families. Probands (arrows) have 22q11DS (upper left quadrant solid fill). We performed WGS on the probands and parents (gray blocks). **A**: Proband P1 presented with 22q11DS, TOF, and SCZD. The P1 family pedigree shows a history of CHD and mental illness in the paternal branch. **B**: Proband P2 presented with 22q11DS, JRA. The P2 family pedigree shows no history of CHD. The paternal grandfather died of Alzheimer's disease and the paternal grandmother was diagnosed with breast cancer.

DNA Preparation and Sequencing

Blood samples were obtained from all subjects and DNA was purified (Puregene system, Gentra Corp., Einstein Molecular Cytogenetics Core). All individuals were given internal identification codes, BM1452.001 (P1, proband 1), BM1452.100 (P1M, mother of proband 1), BM1452.200 (P1F, father of proband 1), BM1453.001 (P2, proband 2), BM1453.100 (P2M, mother of proband 2), BM1453.200 (P2F, father of proband 2). Genomic DNA was extracted from blood samples and then mechanically fragmented. After purification by electrophoresis, DNA fragments were ligated with adapter oligonucleotides to form paired-end libraries with a span size of 500 bp. Libraries were sent for WGS at Beijing Genome

Institute (BGI, Beijing, China) using Illumina the HiSeq2000 sequencing system (Illumina, San Diego, CA).

Variant Calling

For variant calling, GATK version v2.2 was used. In general, the sequencing reads that contained adapter sequence or high rate of low-quality bases were removed. The sequence reads were aligned to the reference genome hg19 using the Burrows–Wheeler Aligner (BWA version 0.6.1) [Li and Durbin, 2009]. Alignment files were converted from sequence alignment map (SAM) format to binary alignment map (BAM) format using SAMtools (version 0.1.18) [Li et al., 2009]. Duplicate reads were then marked using Picard tools (version 1.70, http://broadinstitute.github.io/picard/). The local realignment, recalibration of base quality values, and adjustment of per-base alignment qualities were performed by components from GATK pipeline.

The Unified Genotyper was used to call SNPs or insertion/deletions (INDELs) [DePristo et al., 2011]. The raw single-nucleotide variant (SNV) calls were filtered based on quality by depth (QD < 2), root-mean-square mapping quality (MQ < 40), strand bias (FS > 60), haplotype score (HaplotypeScore > 13), alternate versus reference read mapping quality (MQRankSum < -12.5), and alternate versus reference read position bias (ReadPosRankSum < –8). The filtering criteria for INDELs were QD < 2, ReadPosRankSum < –20, and FS > 200. To obtain phased genotypes and haplotype information, GATK was used to perform both transmission (PhaseByTransmission) and physical (ReadBackedPhasing) phasing.

In addition to the quality control measures used in GATK, we removed variants with genotype quality score ≤ 20, read depth ≤ 15, and alternate allele ratio (depending on genotype): homozygous reference ≥ 0.15, heterozygous alternate allele ratio ≤ 0.3 or ≥ 0.7, and homozygous alternate ≤ 0.85 [Patel et al., 2014]. Variant call format (VCF version 4.1) genotype calls were uploaded to the Database of Genotypes and Phenotypes (dbGaP, http://www.ncbi.nlm.nih.gov/gap) and BAM files were submitted to the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra, Study Accession: phs000837.v1.p1).

Copy-Number Detection for 22q11.2 Deleted Region

Deletion size was determined by copy-number variant (CNV) analysis of the 22q11.2 region. The chromosome 22 was extracted from aligned sequencing reads. Three tools (CNVnator version 0.2.2, BreakDancer version 1.1_2011_02_21, and EDRS version 1.06) were used with default settings to call structural variants with consensus between two out of three methods required to call a structural variant [Chen et al., 2009b; Abyzov et al., 2011; Zhu et al., 2012]. Bedtools (version 2.16.2) was used to identify CNVs with 50% reciprocal overlap among the three methods [Quinlan and Hall, 2010].

Variant Annotation and Filtering

The basic functional annotation of SNVs/INDELs was performed by ANNOVAR version 2014Jul14 [Wang et al., 2010]. We used the table annotation script to annotate variants with RefGene, dbSNP (version 141), Exome Sequencing Project (ESP6500, 2013/05 release), 1000 Genomes project (1000 Genomes, 2012/10 release), dbNSFP v2.3 (for PolyPhen, Sift, LRT, and MutationTaster annotations), and Segmental Duplications [Setti et al., 1990; Chun and Fay, 2009; Kumar et al., 2009; Adzhubei et al., 2010; Schwarz et al., 2010; Liu et al., 2011; Liu et al., 2013]. The DNA and amino acid changes were annotated using HGVS mutation nomenclature [den Dunnen and Antonarakis, 2000]. We excluded variants in segmental duplications because of the high rate of false-positive calls in these areas. Novel variants were defined as not being previously reported in the dbSNP 141, 1000 Genomes, and ESP6500. Rare variants were defined as those with a minor allele frequency (MAF) ≤ 0.05 in the 1000 Genomes Project using all populations.

For coding sequence analysis, we only considered variants with possible deleterious effects including nonsynonymous SNVs, frameshift INDELs, and SNVs/INDELs that affect stop codons and splice sites (2 bp from exon–intron boundary). The cutoffs for predicting pathogenic scores are SIFT ≤ 0.05, PolyPhen2 HDIV ≥ 0.453, LRT = deleterious, and MutationTaster = "disease causing" or "disease causing automatic." Nonsynonymous SNVs were considered deleterious if variants were predicted pathogenic by at least two out of the four algorithms; we created an aggregate deleterious prediction count (DPC), ranging from 0 to 4, indicating the number of deleterious predictions for a given variant.

We analyzed rare noncoding SNVs with MAF ≤ 0.05 in the 1000 Genomes Project database using the FunSeq (v0.1) standalone tool [Khurana et al., 2013]. FunSeq annotates noncoding variants with data from the ENCODE database, Gencode project, and protein–protein interaction information and regulatory networks. It generates a score ranging from 0 to 6, with 6 indicating maximum deleterious effect, and a score ≥4 indicating potential deleterious effects.

Genes with compound heterozygous mutations were identified by filtering based on strict genotype quality and inheritance criteria (Supp. Fig. S1). Rare variants, MAF ≤ 0.05, were selected in the 1000 Genomes Project dataset that had genotype phasing information. Variants were filtered based on predicted pathogenicity, selecting variants predicted to be pathogenic by two or more prediction methods. Genes were selected with two or more pathogenic alleles that were located on different chromosomes.

A pipeline was implemented using Snakemake (v2.5.2), a flexible bioinformatics workflow engine [Koster and Rahmann, 2012]. Scripts used for analysis are available on GitHub (https://github.com/jhchung/22q11DS_WGS_Trio_Analysis).

Candidate Gene Prioritization

The ToppGene suite (https://toppgene.cchmc.org/) was used to perform gene ontology (GO), pathway enrichment, and candidate gene prioritization for coding SNVs and INDELs [Chen et al., 2009a]. The ToppFun tool was applied, using the default parameters to annotate target genes for GO biological process and curated pathway terms. The ToppGene tool was used to prioritize candidate genes using known disease-causing genes as training gene sets. The following training parameters were used all with FDR correction, *P*-value cutoff = 0.05, and gene limits between 1 and 1,500: GO: molecular function, GO: biological process, GO: cellular component, human phenotype, mouse phenotype, pathway, PubMed, interaction, transcription factor binding site, and disease. Test parameters were set to the default values.

For training sets, disease genes in humans and mouse models were downloaded from the OMIM (http://www.ncbi.nlm.nih.gov/omim) and MGI Human-Mouse Disease Connection databases (http://www.informatics.jax.org/humanDisease.shtml), respectively. The training sets for TOF, SCZD, and JRA can be found in Supp. Table S2.

Sanger Sequencing

We validated SNVs and INDELs identified by WGS using Sanger sequencing of PCR-amplified products. PCR primers were designed using Primer3Plus Web-based tool (http://primer3plus.com/). Target product size ranges were set to 150–500 bp with default parameters for all other settings. Primers used for PCR amplification can be found in Supp. Table S3. The PCR assays were performed using FastStart High Fidelity PCR system (0478292001, Roche, Indanapolis, IN) Touchdown PCR with 70°C and 58°C starting and final T_m, respectively, was used for amplification [Korbie and Mattick, 2008]. Products were purified using a PCR Purification Kit (28104; Qiagen, Valencia, CA) and sequenced on 3730 DNA Analyzer (Applied Biosystems, Foster City, CA) at the Albert Einstein College of Medicine Genomics Core Facility. Chromatograms were visualized using SnapGene Viewer 2.7.1 (GSL Biotech, http://www.snapgene.com/products/snapgene_viewer/).

Results

Two Trio Families with De Novo 22q11.2 Deletion

Two unrelated 22q11DS probands with discordant phenotypes, P1 and P2, and their unaffected parents, were ascertained for genetic research and WGS was performed on all six individuals (Fig. 1A and B; gray-shaded area). Subject P1, a male, presented with intellectual disability, behavioral findings, TOF, and hypogonadism. Additionally, he was diagnosed at age 27 years with other specified SCZD spectrum disorder (DSM-V 298.8). Subject P2, a female, had only mild clinical findings including velo-pharyngeal insufficiency and resolved JRA (Supp. Table S1). The de novo 3-Mb deletion was determined in both probands by WGS (Fig. 2A). The deletions were on the maternal allele for P1 and on the paternal allele for P2.

WGS Results and Variant Prioritization Pipeline

We then analyzed the WGS of the two family trios to identify potential coding and noncoding SNVs and small INDELs. Each sample had a mean depth of 35.8×, covering 99.6% of the human reference genome with at least 20× coverage over 90.6% of the genome. Among all six samples, we detected 5,588,483 SNVs and 782,608 INDELs (Supp. Table S4). Each sample had on average 3,176,477 SNVs and 350,038 INDELs, which is comparable to previously reported WGS data from the 1000 Genomes Project [Genomes Project et al., 2012]. A total of 98.2% of SNVs and 86.92% of INDELs were previously reported in dbSNP version 141, the 1000 Genomes Project (2012/10 release), or the Exome Sequencing Project version 6500 (Supp. Table S4).

In order to prioritize the variants that might be pathogenic, we developed an analytical pipeline shown in Figure 3. The pipeline utilized several established tools to annotate coding and noncoding variants, filter the variants, and prioritize candidate variants using integrative genomic analysis methods. We used ANNOVAR to annotate variants with population frequency, phylogenetic conservation scores, gene regions, and exonic functions. Nonsynonymous variants were predicted to be deleterious if there was consensus between two out of four prediction algorithms. We next separated coding and noncoding variants for downstream analysis. We also filtered variants based on MAF to identify de novo, novel, and rare (MAF ≤ 0.05). For noncoding variants, we selected rare variants and used FunSeq to identify potentially functional noncoding variants. After filtering, we used ToppGene to prioritize the genes affected by the target variants based on functional similarity with established disease-causing genes.

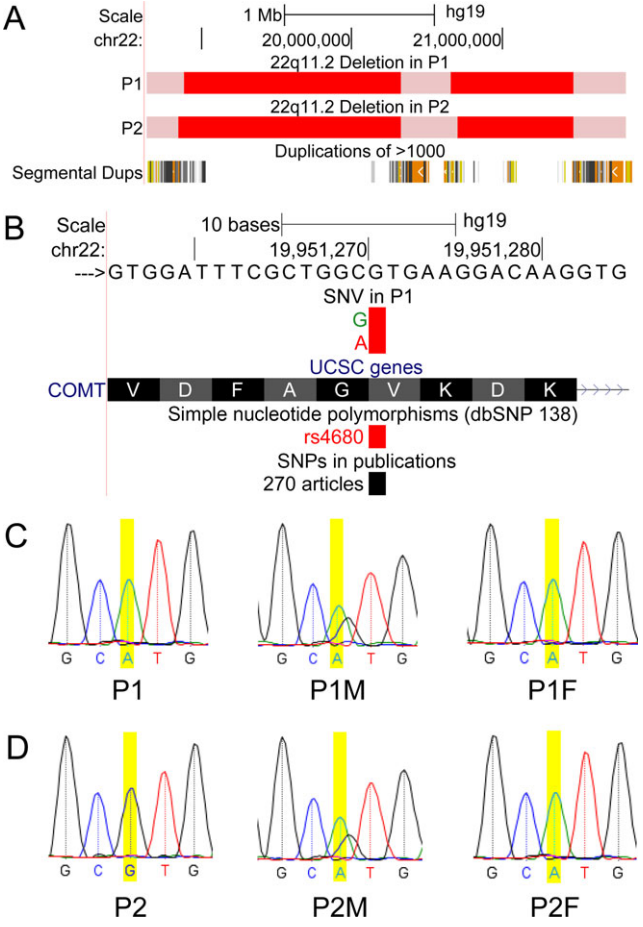

Figure 2. The 22q11.2 deletion and validation of common variants on remaining allele. **A**: De novo deletions identified from CNV calling on the 22q11.2 region using WGS data. Red sections indicate the high-confidence deletions; pink segments indicate LCR regions where there are low-quality alignments and inaccurate CNV endpoint predictions. **B**: A UCSC browser view showing rs4680 in COMT. P1 carries a single A allele. **C**: Sequencing of rs4680 in family 1 using Sanger sequencing. P1 and the father (P1F) are homozygous for the A allele, whereas the mother (P1M) is heterozygous G/A. The variant is highlighted in yellow. **D**: Sequencing of rs4680 in family 2 shows P2 was hemizygous for the G allele, the mother (P2M) was heterozygous G/A and the father (P2F) was homozygous A.

Analysis of the Remaining Allele of Chromosome 22q11.2

We first searched for SNVs or INDELs on the remaining allele of chromosome 22q11.2 since the 22q11.2 deletion can unmask deleterious alleles on the remaining chromosome. The 3-Mb region on chromosome 22q11.2 includes flanking low copy repeats (LCRs), also known as segmental duplications, (chr22:18656000–21792000; Fig. 2A) [Bittel et al., 2009]. The LCR sequences were not included in the WGS analysis. In total, we identified 1,068 variants in P1 and 970 variants in P2 on the remaining allele of 22q11.2, the majority of which were noncoding.

The first priority was to identify nonsynonymous, potentially deleterious, variants in the coding regions because hemizygous deletions could uncover damaging variants on the remaining

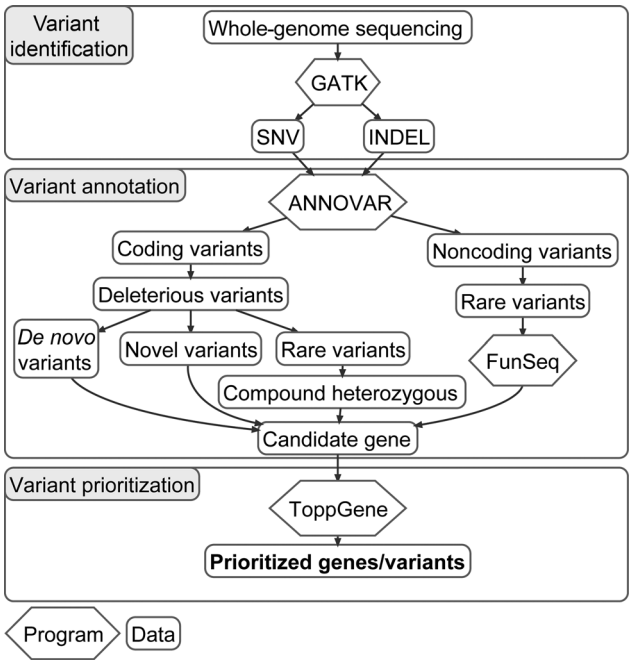

Figure 3. Variant prioritization pipeline. After variant calling, variant annotation was performed using ANNOVAR to identify coding and noncoding variants. ANNOVAR was also used to annotate MAF data from dbSNP, 1000 Genomes, and Exome Sequencing Project databases. Predicted deleterious variants were selected if they were predicted to be deleterious by two out of four prediction methods (MutationTaster, Polyphen2, LRT, and SIFT). For noncoding variants, FunSeq was used to identify potentially deleterious variants, followed by allele frequencies and DPCs to identify candidate genes for variant prioritization. ToppGene was used to prioritize the candidate "target genes" using known disease-specific genes as "training sets."

allele. On the remaining allele of 22q11.2, there were five and four nonsynonymous or frameshift variants in P1 and P2, respectively (Table 1). We found that P1 was hemizygous for the alternate alleles of common nonsynonymous SNVs rs4680 in catechol-O-methyltransferase (*COMT*; MIM #116790; Fig. 2B), encoding an enzyme that regulates catecholamine levels and rs450046 in proline dehydrogenase (oxidase) 1 (*PRODH*; MIM #606810), encoding an enzyme that converts proline to glutamine (Table 1), both connected to brain and behavioral findings in 22q11DS [Graf et al., 2001; Paterlini et al., 2005]. Proband P2 was hemizygous for the minor alleles of common nonsynonymous SNVs in armadillo repeat gene deleted in velocardiofacial syndrome (*ARVCF*; MIM #602269), purigenic receptor P2X, ligand-gated ion channel, 6 (*P2RX6*; MIM #608077), *DGCR2*, and an INDEL in clathrin, heavy polypeptide-like 1 (*CLTCL1*; MIM #601273) (Table 1). Both probands carried the same alleles in *DGCR2* and *CLTCL1*. The proband P2 did not carry any previously reported disease-associated variants on the remaining allele of 22q11.2. However, we identified a common SNV (rs2277838; NM_001159554.1:c.647G>A:p.R216H) that was predicted to be damaging (DPC = 2) in *P2RX6*, encoding an ATP-gated ion channel expressed in several tissues, of largely unknown function (Table 1). We did not identify any rare or novel-coding variants on the remaining allele of 22q11.2 in either P1 or P2.

To identify potentially causal noncoding variants on the remaining allele of 22q11.2, we used the FunSeq software tool, which integrates a variety of annotations to predict disease-causing potential [Khurana et al., 2013]. For P1, there were no SNVs on the intact chromosome that were predicted to be deleterious. In contrast, we identified a predicted functional variant, rs6269, in P2 (FunSeq score = 4), which was located in the promoter region of the soluble *COMT* isoform. In agreement with the software prediction, studies have found the G allele is significantly associated with promoter hypomethylation levels [Schreiner et al., 2011]. This SNV is also part of a common haplotype associated with higher COMT enzyme activity and mRNA stability [Halleland et al., 2009].

De Novo Variants Outside of the 22q11.2 Region

We next searched for de novo variants outside of the 22q11.2 deleted region that were predicted to be pathogenic to make genotype to phenotype correlations. We identified 48 and 32 de novo variants in P1 and P2, respectively. Of these, there was one nonsynonymous SNV in P1 with a DPC $\geq$ 2 but none in P2. P1 had a heterozygous de novo SNV in *ADNP2* (NM_014913.3:c.2243G>C; ADNP homeobox 2; Entrez Gene ID: 22850) resulting in a p.R748P amino acid change predicted to be deleterious by three different methods. The variant was confirmed by Sanger sequencing (Fig. 4A). The variant was located in the last exon of *ADNP2* in a zinc finger, C2H2-like domain, and is located at the beginning of a beta-sheet, as predicted by the SABLE Web server (http://sable.cchmc.org/) (Fig. 4C) [Adamczak et al., 2005]. There is a previously reported SNV (rs377757808) at the same nucleotide position but with a different substitution (NP_055728:c.2243G>A:p.R748Q) compared with the one we identified.

Since the majority of de novo variants were in noncoding regions, we used the FunSeq software tool to prioritize these noncoding variants [Khurana et al., 2013]. Only one de novo variant was predicted to be pathogenic in either proband. In P1, we identified a predicted

Table 1. Nonsynonymous SNVs and Frameshift INDELs on the Remaining Allele of Chromosome 22q11.2.

Subject	Chr	Position	Ref	Alt	rsID	Gene	DNA change	AA change	DPC
P1	chr22	18901004	C	T	rs450046	PRODH	NM_001195226:c.1238G>A	p.R413Q	0
P1	chr22	19026613	A	G	rs2072123	DGCR2	NM_001173533:c.1295T>C	p.V432A	0
P1	chr22	19951271	G	A	rs4680	COMT	NM_007310:c.322G>A	p.V108M	0
P1	chr22	21354970	C	G	rs426938	THAP7	NM_030573:c.343G>C	p.A115P	0
P1	chr22	19189003	–	C	rs11386977	CLTCL1	NM_001835:c.601dup3>G	p.V1201fs	frameshift insertion
P2	chr22	19026613	A	G	rs2072123	DGCR2	NM_001173533:c.1295T>C	p.V432A	0
P2	chr22	19959473	C	T	rs165815	ARVCF	NM_001670:c.2717G>A	p.R906Q	0
P2	chr22	21377650	G	A	rs2277838	P2RX6	NM_001159554:c.647G>A	p.R216H	2
P2	chr22	19189003	–	C	rs11386977	CLTCL1	NM_001835:c.601dup3>G	p.V1201fs	frameshift insertion

Estimated typically deleted region breakpoints of chr22:18,656,000–21,792,000. Deleterious Prediction Count (DPC) indicates the number of prediction algorithms that categorize a variant as pathogenic; score can be between 0–4. Framshift INDEL or splice altering variants are automatically categorized as deleterious.

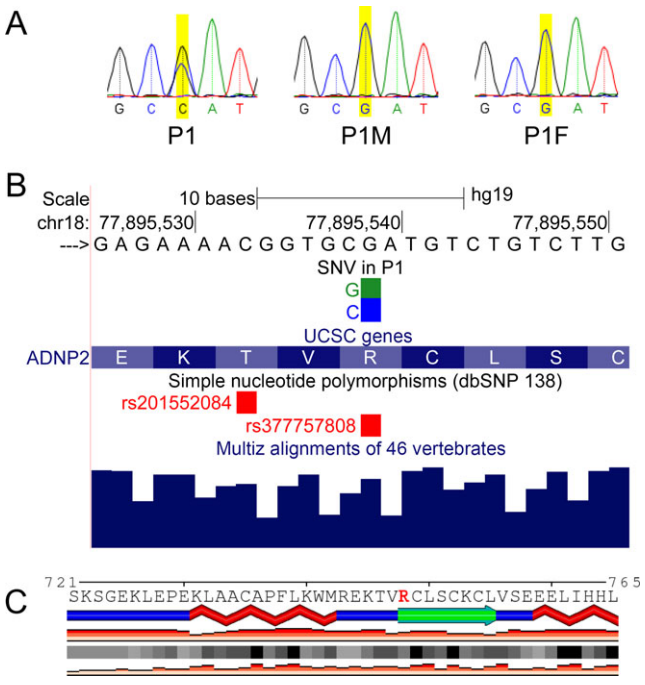

Figure 4. De novo SNV in ADNP2 found in P1. **A**: Sanger sequencing of the de novo variant (NM_014913:c.2243G>C, highlighted in yellow) confirmed that P1 is heterozygous G/C, whereas both parents are homozygous for G allele. **B**: UCSC genome browser views of the de novo SNV in ADNP2. The NHLBI Exome Sequencing Project identified another variant at this position, rs377757808; however, the previously reported alternate allele is A, whereas P1 had C allele. Sequence conservation analysis shows a high degree of conservation among the 46 vertebrate species. **C**: The G allele results in p.R748P amino acid change. SABLE prediction of protein secondary structure of *ADNP2* suggests the de novo mutation affects the arginine (red amino acid letter) at the start of a predicted ß-sheet (green arrow). The first row shows amino acid sequence, second row shows graphical representation of secondary structure (SS) prediction, third row show SS prediction confidence, fourth row shows relative solvent accessibility (RSA), and fifth row shows RSA prediction confidence. SS legend: blue line = coil; red zigzag = α and other helices; green arrow = β-strand or bridge.

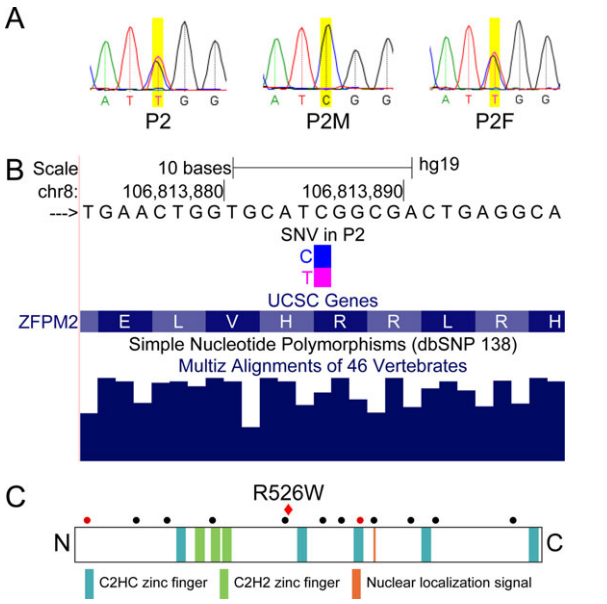

Figure 5. Novel inherited SNV in *ZFPM2* found in P2. **A**: A novel heterozygous SNV in *ZFPM2* (NM_012082:c.1576C>T, yellow) was confirmed by Sanger sequencing. The variant alternate T allele was inherited from the father. **B**: UCSC browser view shows that the SNV is in an evolutionarily conserved region of *ZFPM2*. **C**: Schematic of ZFPM2 protein and known previously reported CHD-related mutations [Pizzuti et al., 2003; De Luca et al., 2011; Tan et al., 2012; Huang et al., 2014]. Circles show previously published mutations; red circles show mutations that occur in two or more subjects. Red diamond indicates position of novel mutation.

functional de novo SNV (NM_144508.4:g.-9092G>T, FunSeq score = 4) in the promoter region of cancer susceptibility candidate 5 (*CASC5*; MIM #609173). We validated this variant by Sanger sequencing (Supp. Fig. S2). The *CASC5* gene encodes for a protein that acts as a scaffold in the formation of kinetochore-microtubule attachments during chromosome segregation. Mutations in *CASC5* have also been found in subjects with microcephaly; however, its role in development is not well understood [Genin et al., 2012]. P2 had no de novo noncoding variants prioritized by FunSeq.

Novel Predicted Deleterious SNV and INDELs

Novel inherited variants could act as second-hit modifiers altering risk for clinical findings in 22q11DS because the probands have the 22q11.2 deletions, which were not present in parental samples. P1 had 49 novel predicted deleterious variants (Supp. Table S5) and P2 had 44 novel predicted deleterious variants (Supp. Table S6). In P1, 19 variants were inherited from the mother (P1M) and 28 were inherited from the father (P1F) (Supp. Fig. S3A). In P2, 20 variants were inherited from the mother (P2M) and 21 variants were inherited from the father (P2F) (Supp. Fig. S3B). There were no genes with novel deleterious variants shared between probands. Interestingly, in P2, we found a novel predicted deleterious variant in zinc finger protein, multitype 2 (*ZFPM2*; MIM #603693, NM_012082.3:c.1576C>T), which results in a p.R526W amino acid change. We validated this variant by Sanger sequencing (Fig. 5A). *ZFPM2* is a known candidate gene for TOF and will be described in more detail below.

Rare Compound Heterozygous-Coding Mutations

We next searched for genes that had a compound heterozygous inheritance pattern for rare predicted deleterious variants; which might implicate loss or reduced function of both copies of the gene. We searched for rare variants with a MAF $\leq$ 0.05 in the 1000 Genomes Project database with a DPC $\geq$ 2. We identified five genes with compound heterozygous mutations in P1 (Supp. Table S5) and nine genes in P2 (Supp. Table S6). Among the genes affected in P1, the genes, FERM, RhoGEF (ARHGEF), and pleckstrin domain protein 1 (*FARP1*; MIM #602654) and Reelin (*RELN*; MIM #600514) are important for neuronal development [Cheadle and Biederer, 2012; Jossin, 2004]. Among the genes affected in P2, *DLG1* (discs large, Drosophila, homolog of, 1; aka *DLGH1* and *SAP97*; MIM #601014) is a candidate gene for autoimmune disorders [Zanin-Zhorov et al., 2012].

Noncoding Variant Prioritization

The majority of variants identified by WGS were noncoding SNVs and INDELs. We selected rare variants with a MAF $\leq$ 0.05 in the

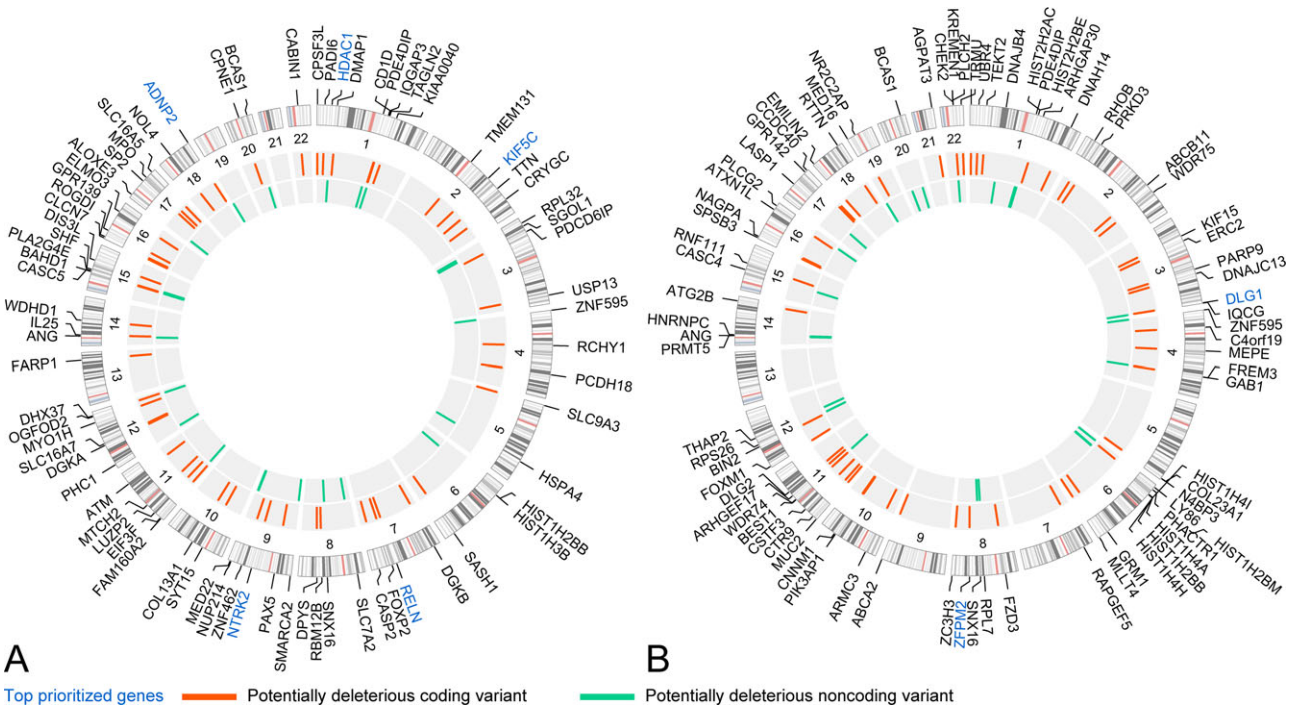

Figure 6. Summary of potentially deleterious target variants. **A**: Candidate functional DNA variants and genes in P1. **B**: Candidate variants and gene in P2. Genes in blue indicate top prioritized candidates from variant filtering and ToppGene analysis. Circular tracks from outside to inside: gene symbols, ideogram of chromosome with cytogenetic bands, chromosome name, potentially deleterious coding variants (red bars), potentially deleterious noncoding variants (green bars). Circular plots were created using the Circos software tool.

1000 Genomes Project database and used the FunSeq program to identify candidate genes with potentially damaging noncoding variants [Khurana et al., 2013]. A total of approximately 120,000 rare inherited noncoding SNVs were present in each subject. In P1, there were 48 SNVs with a predicted functional effect (FunSeq score ≥ 4; Supp. Table S5), which were predicted to affect 23 genes. In P2, there were 51 SNVs with possible functional effects (Supp. Table S5), which were predicted to affect 29 genes.

Target Genes for Gene Set Enrichment and Prioritization

We collected the potentially deleterious variants and their associated genes into "target gene" sets in P1 and P2. The target genes consisted of genes with (1) de novo and novel predicted deleterious variants, (2) rare compound heterozygous deleterious variants, and (3) rare noncoding predicted deleterious variants (Supp. Fig. S4; dark gray node). We excluded variants on the remaining allele of chromosome 22q11.2 because they did not meet our filtering criteria. There were 76 target genes in P1 (Fig. 6A) and 82 target genes in P2 (Fig. 6B). The majority of the affected genes were unique to each subject and only six genes with rare, noncoding variants were shared between P1 and P2 (Supp. Fig. S5). For each proband, we analyzed three sets of target genes: coding (Fig. 6, red bars), noncoding (Fig. 6, green bars), and combined coding and noncoding genes.

Gene Set Enrichment Analysis of Target Genes

Gene set enrichment analysis was performed on the target genes identified in P1 and P2 with GO biological process terms using the ToppFun software tool. Proband P1 had significant enrichment of chromatin organization (GO: 0006325, FDR q value = 4.304e-03) and the p53 signaling pathway (PW: 0000718, FDR q value = 1.199e-2) (Supp. Table S7). P2 also had significant enrichment for GO terms and pathways related to chromatin and nucleosome assembly; however, the associations in P2 were driven by noncoding variants affecting several histone genes (Supp. Table S8).

Candidate Gene Prioritization for Major Phenotypes

We next prioritized the target genes based upon sets of well-established human disease genes using the ToppGene tool that prioritizes and ranks a target gene set based on functional annotation similarity with training set genes made up of known disease genes [Chen et al., 2009a]. For the training gene sets, we used the OMIM and MGI Human-Mouse: Disease Connection databases for genes with high-confidence connections with the observed phenotypes (Supp. Table S2).

Cognitive and Psychiatric Candidate Gene Prioritization

There were 36 known genes that are associated with SCZD (we chose SCZD because P1 was diagnosed with SCZD spectrum disorder and it is the most common and severe behavioral defects in individuals with 22q11DS), in the OMIM and MGI databases, which we used as the training set (Supp. Table S2). Many of these genes are shared with other behavioral or neurocognitive deficits. We performed ToppGene prioritization on the 76 target genes in P1 and 82 target genes in P2. When testing all coding and noncoding target genes, 51.3% (39/76) of the target genes in P1 and 8.54% (7/82) target genes in P2 had significant functional similarity with SCZD training set genes (Fisher's exact test, P-value = 2.13e-9) (Table 2; Supp. Table S9). This enrichment seems to be driven by

Table 2. ToppGene Prioritization Results for Coding, Non-coding and all Target Gene Analysis.

Target gene set	Phenotype	P1 significant	P2 significant	P1 input	P2 input	*P*-value
All genes	SCZD	39	7	76	82	**2.13e-09**
	TOF	21	5	76	81	**4.17e-04**
	JRA	13	7	76	82	0.15
Coding	SCZD	17	3	53	53	**8.77e-04**
	TOF	8	8	53	52	1
	JRA	7	3	53	53	0.319
Non-coding	SCZD	1	0	23	29	0.442
	TOF	11	8	23	29	0.157
	JRA	3	8	23	29	0.308

Significant gene columns indicate the number of genes with significant functional similarity scores after multiple testing correction (FDR) for the number of input genes tested. *P*-values are Fisher's exact test *p*-values comparing the proportion of significant genes in P1 vs. P2.

variants that affect coding genes (17/53 in P1 vs. 3/53 in P2; Fisher's exact test, *P*-value = 8.77e-04) since there was no significant enrichment when testing only noncoding target genes (Fisher's exact test, *P*-value = 0.442) (Table 2). These results suggest P1 has a burden of deleterious coding variants in the behavioral/cognitive-related genes. The top ranked genes in P1 included neurotrophic tyrosine kinase, receptor, type 2 (*NTRK2*; MIM #600456), *RELN*, histone deacetylase 1 (*HDAC1*; MIM #601241), and kinesin family member 5C (*KIF5C*; MIM #604593) (Fig. 6A, blue gene labels on chromosomes 9, 7, 1, and 2, respectively). We validated these variants by Sanger sequencing (Supp. Fig. S2).

TOF Candidate Gene Prioritization

There are six well-established human disease genes listed in the OMIM and MGI databases responsible for TOF (Supp. Table S2). When we analyzed coding and noncoding target genes together, there was a significant enrichment of prioritized genes in P1 (21/76 target genes) as compared with P2 (5/81 target genes) (Fisher's exact test, *P*-value = 4.17e-04; Table 2). The top prioritized target gene in P1 was *HDAC1* (Fig. 6A, blue gene label on chromosome 1). This variant was validated by Sanger sequencing (Supp. Fig. S2). P1 was heterozygous G/C for a rare noncoding SNV (rs187821252; NC_000001.11:g.33234860G>C) in *HDAC1* that was predicted to affect its putative distal regulatory module (FunSeq score = 4; Supp. Table S10).

Interestingly, proband P2, who does not have any cardiac defects, has a novel heterozygous SNV in *ZFPM2*. Mutations in this gene are associated with TOF in OMIM (Fig. 5C). This gene was mentioned above when discussing novel, predicted deleterious SNVs and it was also identified by gene set enrichment. The SNV found in P2 is located proximal to the fifth zinc finger domain (Fig. 5C).

JRA Candidate Gene Prioritization

JRA is a relatively rare phenotype associated with 22q11DS [Keenan et al., 1997; Sullivan et al., 1997]. There were two genes linked to JRA in the OMIM and MGI databases (Supp. Table S2). There was no significant difference in the number of target genes prioritized by ToppGene between P1 and P2 when using coding, noncoding, or all target gene sets (Table 2). We focused on P2 because they presented with JRA. P2 had rare compound heterozygous mutations in *DLG1* (Fig. 6, blue gene label on chromosome 3; Supp. Table S11); rs34492126 and rs1802668 inherited from the father and mother, respectively. Both variants were validated using Sanger sequencing (Supp. Fig. S6). *DLG1* encodes a member of the membrane-associated guanylate kinase protein family that regulates T-cell activation and is dysregulated in individuals with rheumatoid arthritis [Knowlton et al., 2009; Humphries et al., 2012; Zanin-Zhorov et al., 2012].

Discussion

Our goal was to develop an efficient analytical WGS analysis pipeline to define the genetic variation of single individuals with 22q11DS. We specifically chose two, unrelated individuals with 22q11DS that have discordant phenotypes. In this way, they serve as a model for genotype to phenotype correlations. Our focus was on the main clinical findings observed in each affected individual, including the brain and heart. Using this analytical pipeline, we prioritized common variants on the nondeleted allele of 22q11.2 as well as novel and rare deleterious variants (Fig. 6), which can then be tested in much larger datasets for replication and model systems for functional effects in the future.

Enrichment of SCZD-Related Genes

Subject P1 suffered from cognitive as well as behavioral problems during adolescence and was diagnosed with psychotic disorder at 28 years of age. Family history of several family members on the paternal side were reported to be affected by mental illness, including SCZD, indicating that P1 could carry inherited risk factors. We found P1 carried the disease-associated variants in *COMT* and *PRODH* on the remaining allele of chromosome 22q11.2. These variants have been extensively studied as possible risk factors for cognitive and/or behavioral phenotypes [Prasad et al., 2008; Carmel et al., 2014; Radoeva et al., 2014]. A de novo heterozygous mutation was found in *ADNP2*, in P1, predicted to be pathogenic. This variant may disrupt protein secondary structure based on SABLE in silico predictions. Although there is little known about the function of *ADNP2*, it may have a potential role in brain development similar to its paralog activity-dependent neuroprotective protein (*ADNP*; MIM #611386) [Kushnir et al., 2008; Dresner et al., 2011]. Several of the top prioritized target genes in P1 have previously been associated with psychiatric disorders. For example, the top four prioritized coding target genes (*NTRK2*, *RELN*, *HDAC1*, and *KIF5C*) have been previously associated with cognitive disorders or psychiatric illness, including SCZD [Sharma et al., 2008; Otnaess et al., 2009; Ovadia and Shifman, 2011; Bergen et al., 2014]. As a whole, we identified an enrichment of SCZD candidate genes in P1 compared with P2,

which is in line with recent studies on SCZD, showing that affected individuals have increased burden of genetic variants [Ahn et al., 2014; Chan, et al. 2014; Kong et al., 2014]. These findings highlight the possible importance of these genes as modifiers of the mental illness phenotype seen in P1.

Identification of TOF Candidate Genetic Modifiers

TOF is a complex conotruncal heart defect that requires surgical repair for survival [Starr, 2010; Lee et al., 2014]. When we performed ToppGene analysis on all coding and noncoding target genes, there was an enrichment of TOF-prioritized genes in P1 with TOF as compared with P2, with a normal heart. The top prioritized gene for P1 was *HDAC1*. In mice, *Hdac1* regulates *Nkx2-5,* and important transcription factor for heart development, and it genetically interacts with *Gata4,* another important heart transcription factor to repress downstream genes [Liu et al., 2009; Stefanovic et al., 2014]. We also found that P1 target genes were enriched for the p53 pathway, with candidate deleterious variants in three genes (*HDAC1*, *RCHY1*, and *ATM*). Haploinsufficiency of *TBX1*, encoding a T-box transcription factor is one of the strongest candidate genes for TOF in humans with 22q11DS [Jerome and Papaioannou, 2001; Lindsay, 2001; Merscher et al., 2001]. Of interest, suppression of p53 was able to partially rescue CHD caused by the loss of *Tbx1* during mouse heart development [Caprio and Baldini, 2014].

Surprisingly, P2, who had no heart anomalies, carried a novel predicted deleterious SNV in a known human TOF causal gene, *ZFPM2.* The ZFPM2 protein acts as a cofactor for dosage-sensitive GATA transcription factors including GATA4, during embryonic heart development in mouse models [Svensson et al., 1999; Fossett et al., 2001; Garg et al., 2003; Pu et al., 2004]. Mutations in *ZFPM2* have been previously reported in patients with sporadic TOF (Fig. 5C) [Pizzuti et al., 2003; De Luca et al., 2011]. Gene profiling revealed that inactivation of *Tbx1* in mice results in ectopic expression of *Gata4* and *Zfpm2*, in the mesoderm of the pharyngeal apparatus referred to as the second heart field [Liao et al., 2008]. Taken together, it is possible that a loss-of-function variant on one allele of *ZFPM2* may normalize expression of *GATA4* or other downstream genes, suppressing the phenotype.

Challenges of Sequencing Annotation and Variant Prioritization

We aimed to test the utility of currently available tools for identifying candidate genetic modifiers in a personal genomics manner. Our pipeline incorporated computational tools and recommended workflows for the interpretation and prioritization of coding and noncoding variants to create a personal genomics view of individuals with 22q11DS. Using this pipeline, we were able to identify several candidate genes for the various phenotypes seen in these two probands with 22q11DS. However, it is likely that there were some false negative (Type II error) and false positive (Type I error) findings due to stringent filtering criteria, incomplete annotation, or inaccurate functional predictions. Functional validation is a critical step in establishing the role of a variant and its target gene in the etiology of a phenotype [Casanova et al., 2014; MacArthur et al., 2014].

Predicting the functional consequences of coding and noncoding variants remains a significant challenge. For example, well known variants in *COMT* (rs4680) and *PRODH* (rs450046) have been shown to affect protein function, but the majority of prediction algorithms categorize these variants as benign. The interpretation of noncoding variants is especially challenging; however, efforts by the ENCODE project and other similar projects have greatly advanced our knowledge of functional elements in the noncoding genome.

We have performed the first comprehensive analysis of coding and noncoding SNVs and INDELs in 22q11.2DS patients. Despite the limitations of this study, this work suggests that it may be possible to take a personal or individual genomics approach in identifying genetic modifiers for 22q11DS.

Acknowledgments

We would like to sincerely thank the two families participating in this study. We would like to thank Dr. Joe Cubells and Dr. Vandana Shashi for their helpful discussions regarding the psychiatric phenotype seen in P1. Genotyping services were provided through the RS&G Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the National Heart, Lung, and Blood Institute.

Disclosure statement: The authors have no conflict of interests to declare.

References

Abyzov A, Urban AE, Snyder M, Gerstein M. 2011. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res 21:974–984.

Adamczak R, Porollo A, Meller J. 2005. Combining prediction of secondary structure and solvent accessibility in proteins. Proteins 59:467–475.

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7:248–249.

Ahn K, An SS, Shugart YY, Rapoport JL. 2014. Common polygenic variation and risk for childhood-onset schizophrenia. Mol Psychiatry. DOI: 10.1038/mp.2014.158

Bergen SE, O'Dushlaine CT, Lee PH, Fanous AH, Ruderfer DM, Ripke S, International Schizophrenia Consortium SSC, Sullivan PF, Smoller JW, Purcell SM, Corvin A. 2014. Genetic modifiers and subtypes in schizophrenia: investigations of age at onset, severity, sex and family history. Schizophr Res 154:48–53.

Bittel DC, Yu S, Newkirk H, Kibiryeva N, Holt A 3rd, Butler MG, Cooley LD. 2009. Refining the 22q11.2 deletion breakpoints in DiGeorge syndrome by aCGH. Cytogenet Genome Res 124:113–120.

Bromberg Y. 2013. Building a genome analysis pipeline to predict disease risk and prevent disease. J Mol Biol 425:3993–4005.

Cancrini C, Puliafito P, Digilio MC, Soresina A, Martino S, Rondelli R, Consolini R, Ruga EM, Cardinale F, Finocchi A, Romiti ML, Martire B, et al. 2014. Clinical features and follow-up in patients with 22q11.2 deletion syndrome. J Pediatr 164:1475–1480 e2.

Caprio C, Baldini A. 2014. p53 suppression partially rescues the mutant phenotype in mouse models of DiGeorge syndrome. Proc Natl Acad Sci USA 111:13385–13390.

Carmel M, Zarchi O, Michaelovsky E, Frisch A, Patya M, Green T, Gothelf D, Weizman A. 2014. Association of COMT and PRODH gene variants with intelligence quotient (IQ) and executive functions in 22q11.2DS subjects. J Psychiatr Res 56:28–35.

Casanova JL, Conley ME, Seligman SJ, Abel L, Notarangelo LD. 2014. Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies. J Exp Med 211:2137–2149.

Chan Y, Lim ET, Sandholm N, Wang SR, McKnight AJ, Ripke S, Consortium D, Consortium G, Consortium G, Consortium I, Consortium PGC, Daly MJ, et al. 2014. An excess of risk-increasing low-frequency variants can be a signal of polygenic inheritance in complex diseases. Am J Hum Genet 94:437–452.

Cheadle L, Biederer T. 2012. The novel synaptogenic protein Farp1 links postsynaptic cytoskeletal dynamics and transsynaptic organization. J Cell Biol 199:985–1001.

Chen J, Bardes EE, Aronow BJ, Jegga AG. 2009a. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res 37(Web Server issue):W305–W311.

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, Shi X, Fulton RS, et al. 2009b. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 6:677–681.

Chun S, Fay JC. 2009. Identification of deleterious mutations within three human genomes. Genome Res 19:1553–1561.

Cobb JE, Hinks A, Thomson W. 2014. The genetics of juvenile idiopathic arthritis: current understanding and future prospects. Rheumatology (Oxford) 53:592–599.

De Luca A, Sarkozy A, Ferese R, Consoli F, Lepri F, Dentici ML, Vergara P, DeZorzi A, Versacci P, Digilio MC, Marino B, Dallapiccola B. 2011. New mutations in ZFPM2/FOG2 gene in tetralogy of Fallot and double outlet right ventricle. Clin Genet 80:184–190.

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7–12.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

Dresner E, Agam G, Gozes I. 2011. Activity-dependent neuroprotective protein (ADNP) expression level is correlated with the expression of the sister protein ADNP2: deregulation in schizophrenia. Eur Neuropsychopharmacol 21:355–361.

Fossett N, Tevosian SG, Gajewski K, Zhang Q, Orkin SH, Schulz RA. 2001. The Friend of GATA proteins U-shaped, FOG-1, and FOG-2 function as negative regulators of blood, heart, and eye development in Drosophila. Proc Natl Acad Sci USA 98:7342–7347.

Garg V, Kathiriya IS, Barnes R, Schluterman MK, King IN, Butler CA, Rothrock CR, Eapen RS, Hirayama-Yamada K, Joo K, Matsuoka R, Cohen JC, et al. 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 424:443–447.

Genin A, Desir J, Lambert N, Biervliet M, Van Der Aa N, Pierquin G, Killian A, Tosi M, Urbina M, Lefort A, Libert F, Pirson I, et al. 2012. Kinetochore KMN network gene CASC5 mutated in primary microcephaly. Hum Mol Genet 21:5306–5317.

Gennery AR. 2013. Immunological features of 22q11 deletion syndrome. Curr Opin Pediatr 25:730–735.

Graf WD, Unis AS, Yates CM, Sulzbacher S, Dinulos MB, Jack RM, Dugaw KA, Paddock MN, Parson WW. 2001. Catecholamines in patients with 22q11.2 deletion syndrome and the low-activity COMT polymorphism. Neurology 57: 410–416.

Guo T, McDonald-McGinn D, Blonska A, Shanske A, Bassett AS, Chow E, Bowser M, Sheridan M, Beemer F, Devriendt K, Swillen A, Breckpot J, 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.

Halleland H, Lundervold AJ, Halmoy A, Haavik J, Johansson S. 2009. Association between catechol O-methyltransferase (COMT) haplotypes and severity of hyperactivity symptoms in adults. Am J Med Genet B Neuropsychiatr Genet 150B:403–410.

Huang X, Niu W, Zhang Z, Zhou C, Xu Z, Liu J, Su Z, Ding W, Zhang H. 2014. Identification of novel significant variants of ZFPM2/FOG2 in non-syndromic tetralogy of Fallot and double outlet right ventricle in a Chinese Han population. Mol Biol Rep 41:2671–2677.

Humphries LA, Shaffer MH, Sacirbegovic F, Tomassian T, McMahon KA, Humbert PO, Silva O, Round JL, Takamiya K, Huganir RL, Burkhardt JK, Russell SM, et al. 2012. Characterization of in vivo Dlg1 deletion on T cell development and function. PLoS One 7:e45276.

Jerome LA, Papaioannou VE. 2001. DiGeorge syndrome phenotype in mice mutant for the T-box gene, Tbx1. Nat Genet 27:286–291.

Jossin Y. 2004. Neuronal migration and the role of reelin during early development of the cerebral cortex. Mol Neurobiol 30:225–251.

Keenan GF, Sullivan KE, McDonald-McGinn DM, Zackai EH. 1997. Arthritis associated with deletion of 22q11.2: more common than previously suspected. Am J Med Genet 71:488.

Khurana E, Fu Y, Colonna V, Mu XJ, Kang HM, Lappalainen T, Sboner A, Lochovsky L, Chen J, Harmanci A, Das J, Abyzov A, et al. 2013. Integrative annotation of variants from 1092 humans: application to cancer genomics. Science 342:1235587.

Knowlton N, Jiang K, Frank MB, Aggarwal A, Wallace C, McKee R, Chaser B, Tung C, Smith L, Chen Y, Osban J, O'Neil K, et al. 2009. The meaning of clinical remission in polyarticular juvenile idiopathic arthritis: gene expression profiling in peripheral blood mononuclear cells identifies distinct disease states. Arthritis Rheum 60:892–900.

Kong SW, Lee IH, Leshchiner I, Krier J, Kraft P, Rehm HL, Green RC, Kohane IS, MacRae CA. 2014. Summarizing polygenic risks for complex diseases in a clinical whole-genome report. Genet Med.[Epub ahead of print]. DOI: 10.1038/gim.2014.143

Korbie DJ, Mattick JS. 2008. Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc 3:1452–1456.

Koster J, Rahmann S. 2012. Snakemake—a scalable bioinformatics workflow engine. Bioinformatics 28:2520–2522.

Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4: 1073–1081.

Kushnir M, Dresner E, Mandel S, Gozes I. 2008. Silencing of the ADNP-family member, ADNP2, results in changes in cellular viability under oxidative stress. J Neurochem 105:537–545.

Lee CH, Kwak JG, Lee C. 2014. Primary repair of symptomatic neonates with tetralogy of Fallot with or without pulmonary atresia. Korean J Pediatr 57: 19–25.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079.

Liao J, Aggarwal VS, Nowotschin S, Bondarev A, Lipner S, Morrow BE. 2008. Identification of downstream genetic pathways of Tbx1 in the second heart field. Dev Biol 316:524–537.

Lindsay EA. 2001. Chromosomal microdeletions: dissecting del22q11 syndrome. Nat Rev Genet 2:858–868.

Liu X, Jian X, Boerwinkle E. 2011. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 32:894–899.

Liu X, Jian X, Boerwinkle E. 2013. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat 34:E2393–E2402.

Liu Z, Li T, Liu Y, Jia Z, Li Y, Zhang C, Chen P, Ma K, Affara N, Zhou C. 2009. WNT signaling promotes Nkx2.5 expression and early cardiomyogenesis via downregulation of Hdac1. Biochim Biophys Acta 1793:300–311.

MacArthur DG, Manolio TA, Dimmock DP, Rehm HL, Shendure J, Abecasis GR, Adams DR, Altman RB, Antonarakis SE, Ashley EA, Barrett JC, Biesecker LG, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. Nature 508:469–476.

McDonald-McGinn DM, Fahiminiya S, Revil T, Nowakowska BA, Suhl J, Bailey A, Mlynarski E, Lynch DR, Yan AC, Bilaniuk LT, Sullivan KE, Warren ST, et al. 2013. Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11.2DS. J Med Genet 50:80–90.

Merscher S, Funke B, Epstein JA, Heyer J, Puech A, Lu MM, Xavier RJ, Demay MB, Russell RG, Factor S, Tokooya K, Jore BS, et al. 2001. TBX1 is responsible for cardiovascular defects in velo-cardio-facial/DiGeorge syndrome. Cell 104:619–629.

Momma K, Kondo C, Matsuoka R. 1996. Tetralogy of Fallot with pulmonary atresia associated with chromosome 22q11 deletion. J Am Coll Cardiol 27:198–202.

Morrow B, Goldberg R, Carlson C, Das Gupta R, Sirotkin H, Collins J, Dunham I, O'Donnell H, Scambler P, Shprintzen R, et al. 1995. Molecular definition of the 22q11 deletions in velo-cardio-facial syndrome. Am J Hum Genet 56:1391–1403.

Otnaess MK, Djurovic S, Rimol LM, Kulle B, Kahler AK, Jonsson EG, Agartz I, Sundet K, Hall H, Timm S, Hansen T, Callicott JH, et al. 2009. Evidence for a possible association of neurotrophin receptor (NTRK-3) gene polymorphisms with hippocampal function and schizophrenia. Neurobiol Dis 34:518–524.

Ovadia G, Shifman S. 2011. The genetic variation of RELN expression in schizophrenia and bipolar disorder. PLoS One 6:e19955.

Patel ZH, Kottyan LC, Lazaro S, Williams MS, Ledbetter DH, Tromp H, Rupert A, Kohram M, Wagner M, Husami A, Qian Y, Valencia CA, et al. 2014. The struggle to find reliable results in exome sequencing data: filtering out Mendelian errors. Front Genet 5:16.

Paterlini M, Zakharenko SS, Lai WS, Qin J, Zhang H, Mukai J, Westphal KG, Olivier B, Sulzer D, Pavlidis P, Siegelbaum SA, Karayiorgou M, et al. 2005. Transcriptional and behavioral interaction between 22q11.2 orthologs modulates schizophrenia-related phenotypes in mice. Nat Neurosci 8:1586–1594.

Pizzuti A, Sarkozy A, Newton AL, Conti E, Flex E, Digilio MC, Amati F, Gianni D, Tandoi C, Marino B, Crossley M, Dallapiccola B. 2003. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. Hum Mutat 22:372–377.

Prasad SE, Howley S, Murphy KC. 2008. Candidate genes and the behavioral phenotype in 22q11.2 deletion syndrome. Dev Disabil Res Rev 14:26–34.

Pu WT, Ishiwata T, Juraszek AL, Ma Q, Izumo S. 2004. GATA4 is a dosage-sensitive regulator of cardiac morphogenesis. Dev Biol 275:235–244.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Radoeva PD, Coman IL, Salazar CA, Gentile KL, Higgins AM, Middleton FA, Antshel KM, Fremont W, Shprintzen RJ, Morrow BE, Kates WR. 2014. Association between autism spectrum disorder in individuals with velocardiofacial (22q11.2 deletion) syndrome and PRODH and COMT genotypes. Psychiatr Genet 24:269–272.

Rasmussen SA, Williams CA, Ayoub EM, Sleasman JW, Gray BA, Bent-Williams A, Stalker HJ, Zori RT. 1996. Juvenile rheumatoid arthritis in velo-cardio-facial syndrome: coincidence or unusual complication? Am J Med Genet 64: 546–550.

Schneider M, Debbane M, Bassett AS, Chow EW, Fung WL, van den Bree M, Owen M, Murphy KC, Niarchou M, Kates WR, Antshel KM, Fremont W, et al. 2014. Psychiatric disorders from childhood to adulthood in 22q11.2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11.2 Deletion Syndrome. Am J Psychiatry 171:627–639.

Schreiner F, El-Maarri O, Gohlke B, Stutte S, Nuesgen N, Mattheisen M, Fimmers R, Bartmann P, Oldenburg J, Woelfle J. 2011. Association of COMT genotypes with S-COMT promoter methylation in growth-discordant monozygotic twins and healthy adults. BMC Med Genet 12:115.

Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. 2010. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods 7:575–576.

Setti P, Percudani M, Tarasconi P, Delfrate R, Lampugnani R, Chiampo L. 1990. [Laparocele in the aged]. Acta Biomed Ateneo Parmense 61:207–212.

Shaikh TH, Kurahashi H, Saitta SC, O'Hare AM, Hu P, Roe BA, Driscoll DA, McDonald-McGinn DM, Zackai EH, Budarf ML, Emanuel BS. 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.

Sharma RP, Grayson DR, Gavin DP. 2008. Histone deactylase 1 expression is increased in the prefrontal cortex of schizophrenia subjects: analysis of the National Brain Databank microarray collection. Schizophr Res 98:1111–1117.

Stalmans I, Lambrechts D, De Smet F, Jansen S, Wang J, Maity S, Kneer P, vonder Ohe M, Swillen A, Maes C, Gewillig M, Molin DG, et al. 2003. VEGF: a modifier of the del22q11 (DiGeorge) syndrome? Nat Med 9:173–182.

Starr JP. 2010. Tetralogy of fallot: yesterday and today. World J Surg 34:658-668.

Stefanovic S, Barnett P, van Duijvenboden K, Weber D, Gessler M, Christoffels VM. 2014. GATA-dependent regulatory switches establish atrioventricular canal specificity during heart development. Nat Commun 5:3680.

Sullivan KE, McDonald-McGinn DM, Driscoll DA, Zmijewski CM, Ellabban AS, Reed L, Emanuel BS, Zackai EH, Athreya BH, Keenan G. 1997. Juvenile rheumatoid arthritis-like polyarthritis in chromosome 22q11.2 deletion syndrome (DiGeorge anomalad/velocardiofacial syndrome/conotruncal anomaly face syndrome). Arthritis Rheum 40:430–436.

Svensson EC, Tufts RL, Polk CE, Leiden JM. 1999. Molecular cloning of FOG-2: a modulator of transcription factor GATA-4 in cardiomyocytes. Proc Natl Acad Sci USA 96:956–961.

Tan ZP, Huang C, Xu ZB, Yang JF, Yang YF. 2012. Novel ZFPM2/FOG2 variants in patients with double outlet right ventricle. Clin Genet 82:466–471.

Wang K, Li M, Hakonarson H. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38: e164.

Ware JS, Petretto E, Cook SA. 2013. Integrative genomics in cardiovascular medicine. Cardiovasc Res 97:623–630.

Zanin-Zhorov A, Lin J, Scher J, Kumari S, Blair D, Hippen KL, Blazar BR, Abramson SB, Lafaille JJ, Dustin ML. 2012. Scaffold protein disc large homolog 1 is required for T-cell receptor-induced activation of regulatory T-cell function. Proc Natl Acad Sci USA 109:1625–1630.

Zhu M, Need AC, Han Y, Ge D, Maia JM, Zhu Q, Heinzen EL, Cirulli ET, Pelak K, He M, Ruzzo EK, Gumbs C, et al. 2012. Using ERDS to infer copy-number variants in high-coverage genomes. Am J Hum Genet 91:408–421.

1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56–65.