Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

_____ July 14th, 2016

Benjamin L. Rambo-Martin

Genomics of Down Syndrome Associated Congenital Heart Defects

By

Benjamin L Rambo-Martin
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

_____

Stephanie L. Sherman, Ph.D.
Advisor

_____

Michael E. Zwick, Ph.D.
Advisor

_____

David J. Cutler, Ph.D.
Committee Member

_____

Michael P. Epstein, Ph.D.
Committee Member

_____

Jennifer G. Mulle, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

Genomics of Down Syndrome Associated Congenital Heart Defects

By

Benjamin L Rambo-Martin
B.S., The University of Georgia, 2006
M.S., The University of Georgia, 2009

Advisor: Stephanie L. Sherman, Ph.D.
Advisor: Michael E. Zwick, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

2016

Abstract

Genomics of Down Syndrome Associated Congenital Heart Defects

By Benjamin L Rambo-Martin

Individuals with Down syndrome are 2,000 times more likely to be born with an atrioventricular septal defect than people in the euploid population. The genetic loci that incite this risk, causing AVSD in 1 in 5 individuals with DS, are poorly understood. Within this dissertation we test the hypothesis that genes interrupted by copy number variants on chromosome 21 provide protection from AVSD by reducing that locus to disomy. We followed stringent quality control methods in producing a high quality set of deletions and duplications across 188 case individuals with DS+AVSD and 211 control individuals with DS and a normal heart. We found that no individual CNV, or any individual gene intersected by a CNV associated with AVSD in DS. From burden analyses, we found that African American controls had more bases covered by rare deletions than did cases. Inversely, Caucasian cases had more genes intersected by rare duplications than did Caucasian controls. We performed gene set enrichment analysis in Caucasians and found a suggestion of cilia genes being more often intersected by deletions in controls and more often by duplications in cases. Pathway analyses of genes intersected only by deletions show an overrepresentation in genes involved in protein heterotrimerization and of genes intersected only by duplications being overrepresented in synaptic vesicle endocytosis. We also show that CNVs on chromosome 21 previously associated with DS+AVSD are likely false positives. We then apply whole genome and whole exome sequencing to a sub-cohort and test for epistatic SNP effects in cilia genes and find suggestive enrichment. Finally, we perform pathway analysis of genes harboring rare variants nominally associated to cases or controls, leading us to a novel candidate pathways perturbed in congenital heart defects. This research adds to the swell of evidence indicating that DS-associated AVSD is similarly heterogeneous as is AVSD in the euploid population.

Genomics of Down Syndrome Associated Congenital Heart Defects

By

Benjamin L Rambo-Martin
B.S., The University of Georgia, 2006
M.S., The University of Georgia, 2009

Advisor: Stephanie L. Sherman, Ph.D.
Advisor: Michael E. Zwick, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in
Graduate Division of Biological and Biomedical Science
Genetics and Molecular Biology

2016

My colleagues,
My friends,
My family,
Thank you.

Tables

Figures

I. Introduction

I.I Early Heart Development Overview

Two weeks into human gestation the anterior plate mesoderm of the embryo derives cells that assemble into a crescent destined to become the heart (Figure 1.1). This process is initiated by Wnt and BMP signaling that trigger a host of transcription factors including *NKX2.5* and *GATA-4/5/6* (Figure 1.2)(Lyons et al., 2005; De Calisto et al., 2005; Qian and Bodmer, 2009). A week later, these cells form a loose tube of interior endocardial cells and external myocardial cells, separated by an extracellular "cardiac jelly" layer to mediate requisite molecular cross talk. At the formation of the cardiac crescent there are already two distinct populations of cells, the first and secondary heart field (FHF and SHF). The FHF will go on to contribute significantly to the left ventricle, as well as other chambers, as a scaffold through which the SHF will migrate. The SHF will become the majority of the right ventricle and outflow tract. In concert with *NKX2.5,* the transcription factors *HAND1* and *HAND2* play a role in the differentiation of the two heart fields. The disruption of *HAND1* leads to left ventricle defects and errors in *HAND2* lead to right ventricle defects in mice (Yamagishi et al., 2001). The SHF will also lead to the diverse cell types of both myocardium and the interior smooth muscle, a process reliant on the transcription factor *Mef2*, which is necessary for all types of muscle differentiation (Gajewski et al., 1997). *TBX1* performs a key regulatory role in SHF differentiation at least through the establishment of the outflow track (Xu et al., 2004).

As the heart tube begins the looping that will result in the four-chambered heart, signaling pathways, particularly Wnt/β-catenin, trigger cardiac neural crest cell (cNCC) migration from the dorsal neural tube and into the caudal pharyngeal arches to contribute to the septum and the aortic arch (De Calisto et al., 2005). A fourth population of cells originates from the coelomic mesothelium around the liver bud and travels around the looping heart tube to form the epicardium. A portion of these cells will undergo an epithelial-mesenchymal transition (EMT) and provide connective tissue for the coronary vessels (Kodo and Yamagishi, 2011). At this time, cells from the SHF, now making up the primitive right ventricle and outflow tract, are also signaling back and forth with cells from the FHF, triggering the SHF cells to undergo EMT and form the ventricular floor, outflow tract, cardiac cushion and valves (Koefoed et al., 2014).

I.II Congenital Heart Defects (CHD)

Congenital heart defects (CHD) are the most common birth defect, presenting in 80 out of 1,000 live births and causing 25% of infant mortality due to birth defects (Reller et al. 2008; Yang et al. 2006; Hartman et al. 2011). When CHD are part of a syndrome, survival is sometimes further reduced. For example, for individuals with Down syndrome (DS), presence of CHD increases their risk of death nearly 5 fold between the neonatal period to 20 years of age (Kucik et al., 2013). CHDs are also expensive and account for over half of hospital costs related to birth

defects, estimated at nearly $1.4 billion in 2004 (Russo and Elixhauser, 2007).

CHDs encompass a wide variety of specific heart defects. They range in severity, affected tissue, and tissue specificity. Atrial septal defects (ASD) are a relatively benign defect resulting from failure of complete septation between the atria and occurring in 65 out 10,000 live births (Figures 1.3 and 1.4)(Mai et al., 2015). 99.6% of individuals born with an ASD can expect to live through 27 years (data is limited to 20 year follow-up; Verheugt et al., 2008). Increasing in severity is tetralogy of Fallot (TOF) (Figure 1.5). Individuals with TOF have four classic features: 1. pulmonary valve narrowing, 2. right ventricular wall thickening, 3. overriding aorta, and 4. ventricular septal defect. They occur in 4 out of 10,000 live births and despite the complexity in defects, have good rates of survival of over 93% in 20 year follow-up studies, though as many as 10% of patients will have to have multiple surgeries to rescue failed initial repairs (Mai et al., 2015; Verheugt et al., 2008). Other common defects are listed in Table 1.1.

Atrioventricular septal defects (AVSD) are a serious, albeit rare CHD resulting from the failure of endocardial cushion and subsequent mitral and tricuspid valve formation, leading to improper mixing of oxygenated and deoxygenated blood (Figure 1.6). Only 13-30% percent of babies survive to term after a prenatally diagnosed AVSD (Rasiah et al., 2008). Only 47-66% of infants with AVSD survive to age 25 (Miller et al., 2010).

While the gross structural defect is typically repaired during the first year of life, patients with AVSD face increased risk of sequelae including arrhythmias, endocarditis, stroke, congestive heart failure, pulmonary hypertension, and

continued heart valve problems (Le Gloan et al. 2011). AVSD diagnoses encompasses a continuum of phenotypes from partial defects of atrial (23.4% of cases) or ventricular shunting with tricuspid and mitral valve clefts (10% of cases), to complete AVSD, which accounts for 66.6% of cases (Loffredo et al., 2001).

I.II.i *AVSD epidemiology*

Loffredo et al. (2001) published a comprehensive epidemiological report of AVSD using data from the Baltimore-Washington Infant Study (1981-1989) finding AVSDs in 3.5 per 10,000 live births and 7.3% of 4,390 infants born with CHDs. 13% of these AVSDs had complex phenotypes associated with heart tube looping and conotruncal defects. Within the 213 complete AVSDs, 12.2% were not associated with any other anomaly, 81.6% were associated with chromosomal abnormalities (79.3% of those have DS), 2.3% were associated with a known Mendelian defect (Smith-Lemli-Opitz, Schinzel-Giedon, Holt-Oram, and van der Woude syndromes) and 3.7% with unknown etiology. Reller et al. (2008), using data from the Metropolitan Atlanta Birth Defects Program (1998-2005), found a similar rate of live births with AVSD at 2.2 per 10,000, noting an increased prevalence in females (60% of complete AVSDs) that was statistically significant ($p < 0.05$). They also found that AVSDs were associated with lower gestational age, lower birth weight and increased maternal age ($p < 0.05$), which are all confounding variables and associated with increased risk for DS births that were not excluded here. The most recent estimates come from Mai et al. (2015), who parsed data from the National

Birth Defects Prevention Network (2008-2012), finding an overall mean rate of AVSD across participating states of 4.2 per 10,000 live births. While this rate is higher than previously found, they do not distinguish between complete and partial AVSDs or rate of DS comorbidity. They also report differences between ethnic groups, with non-Hispanic blacks having the highest rate at 5.4/10,000, followed by non-Hispanic whites (4.2/10,000) and Hispanics (3.4/10,000).

I.II.ii *AVSD heritability*

An early estimate in a small cohort of 52 offspring from 52 patients (36 affected mothers and 16 affected fathers) with atrioventricular septal defects found 10% of the children or 14% of female children to also have AVSD (Emanuel et al., 1983). A later small study of 103 pedigrees with a non-syndromic proband diagnosed with AVSD found 11.7% of families to have at least one 1st or 2nd degree relative affected by a CHD (Digilio et al., 1993). The phenomenal tracking of healthcare records in Denmark has allowed for a much more precise estimate of recurrence risk of a variety of CHDs for the Danish population (Oyen et al., 2009). Looking at family data from 18,207 children born with a CHD between 1977 and 2005, the relative risk for a same-sex twin was 12.5 (95% CI: 10.9-14.3) and that for a 1st degree relative was 3.2 (95% CI: 3-3.5). First-degree relatives of children born with an AVSD had a relative risk of 24.3 (95% CI: 12.2-48.7).

I.II.iii *Genetics of CHDs with a focus on AVSD*

The majority of CHDs are not clearly familial and most of the histories of causal gene discovery arise from studying CHD-associated syndromes. Holt-Oram syndrome is very rare, occurring in 1 in 100,000 live births, and has a number of developmental defects including heart anomalies. Following the mapping of Holt-Oram syndrome to chromosome 12q24 and the creation of yeast artificial chromosomes combined with exon trapping methods, the causal gene was discovered to be a master transcription-factor regulator, *TBX5* (Li et al., 1997). *TBX5* and other transcription factors active in cardiac progenitor cells are now known to also associate with AVSD (Raemon-Buettner and Borlak, 2005).

The alternative to studying heart defects in non-related individuals with the same diagnosed syndrome is to utilize large pedigrees that appear to have a rare variant of large effect segregating through their family. Within a large pedigree segregating nonsyndromic AVSD, a locus at chromosome 1p21-1p31 (named *AVSD1*) was discovered using linkage analysis of 13 cases and 14 controls, but a gene was never implicated (Sheffield et al., 1997). Around that time, a second locus (called *AVSD2*) was discovered by the breakpoint mapping of 3p deletion syndrome and its subsequent association to AVSD (Green et al., 2000). The causal gene at this locus was found to be *CRELD1* (Rupp et al., 2002). A small study of 52 patients with nonsyndromic AVSD found that 6% harbored a missense mutation in *CRELD1* (Robinson et al. 2003). These mutations were not fully penetrant, as they were present in unaffected parents. *CRELD1* is expressed in many developing tissues, including the heart and is a cysteine-rich cell adhesion molecule in a class of

epidermal growth factors (Rupp et al., 2002).

Transcription factors (TF) represent the largest class of genes identified carrying mutations associating with AVSD. Missense mutations in the TF *GATA4* were found segregating in familial CHDs that included AVSD (Garg et al. 2003). *GATA4* interacts with the T-box protein *TBX5*, in which variants at its DNA-binding site were implicated in a range of CHDs (Li et al. 1997). Familial atrial septal defects have been found to be caused by mutation in *GATA4* and in *NKX2.5*, a necessary cofactor for *TBX5*'s DNA binding (Hiroi et al., 2001; Hirayama-Yamada et al., 2005; Chen et al., 2010). Missense variants in the related TF *GATA6* were also found in two out of 319 patients with diverse CHDs; one of the two had an AVSD (Maitra et al., 2010). Another TF *HAND1* was discovered with missense somatic mutations in cardiac heart tissue collected during the repair of AVSD (Reamon-Buettner et al. 2009). These mutations were not found in any control samples, nor in matched blood tissue samples, highlighting the limitations of finding *de novo* causal variants in peripheral tissue if the causal mutation arises somatically in the affected tissue.

While these family-based linkage studies were designed to uncover highly penetrant rare variants segregating in a Mendelian fashion, the majority of CHDs occur sporadically. With the explosion of microarray technology following the completion of the Human Genome Project in 2003 a new disease model could be tested without the difficulties of recruiting large disease-burdened families. It was hypothesized that common diseases were influenced by common variants of moderate effects, which could be tested by recruiting moderately sized cohorts of unrelated affected and unaffected individuals from the population (Chakravarti,

1999). These studies became known as genome-wide association studies (GWAS).

GWAS have also been applied to the search for CHD loci. In 957 Han Chinese individuals with atrial septal defects (ASD), ventricular septal defects (VSD) and both ASD/VSDs were compared to 1,308 ethnically-matched controls without CHD (Hu et al., 2012). Eight nominally associated SNPs were then genotyped in 1,575 CHD individuals and 2,297 controls. Two of these SNPs remained statistically significant and in the associated direction in both the first replication as well as another replication of 583 CHD subjects and 1,565 controls. By fine mapping the associated SNPs with variants in linkage disequilibrium, they discovered two genes highly expressed in cardiac tissue. The first was *TBX15*, a transcription factor in the same family as the previously described *TBX5*. The other gene was *MAML3*, which is part of a class of genes required for the Notch signaling pathway. A second GWAS by Cordell et al. (2012) in the same Nature Genetics publication reported different associated locus. They compared the genotypes of 1,819 Caucasian individuals with diverse CHDs to 5,159 separately genotype population controls. They took the 10 and 21 most associated SNPs in the ASD and VSD subgroups, respectively, and performed a replication in 417 people with ASD and 209 with VSD. The VSD SNPs failed to replicate, but within the ASD group, significant associations were seen for SNPs at 4p16 near *MSX1* and *STX18. MSX1* is a transcription factor expressed in the atrial septum of chickens and mice. *STX8* on the other hand, is involved in protein shuttling between the Golgi and the endoplasmic reticulum and its relationship to heart development is unknown. They also tested these SNPs for association in a cohort of 73 patients with AVSD, but did not see an effect.

In an attempt to overcome the increasingly apparent heterogeneity of CHD, Kamp et al. (2010) performed a forward genetic mouse screen followed by a GWAS. By mutagenizing isogenic lines of mice and selecting for perinatal lethal mutations leading to CHD, they identified six loci of interest for AVSD. None of the loci were completely penetrant and many showed variable expressivity in crossing experiments.

With the advent of cost-effective high-throughput sequencing in the past decade, researchers have been able to measure rare variation within populations and attempt to associate those with phenotypes. Zaidi et al. (2013) performed whole exome sequencing on 362 case-parent trios of severe-CHD (excluding the relatively benign ventricular septal defects, atrial septal defects, patent ductus arteriosus or pulmonic stenosis). Filtering for rare, *de novo* variants, they found an enrichment of mutations in genes including *MLL2, WDR5, KDM5A,* H3K4 demethylases, and *CHD7,* all related to the global epigenetic mechanisms involved in adding and removing histone 3 lysine 4 methylation. While canonical H3K4 methylation represents open chromatin and thus active gene expression, they also observed mutations in *SMAD2*, which regulates H3K27 methylation, a mark of closed chromatin and transcription suppression. Thirteen of their cases also had *de novo* mutations in a list of preselected CHD candidate genes, which was significantly more than expected by chance (Monte Carlo simulated p-value = 0.0007). Together, they estimated that 10% of CHD cases were caused by rare *de novo* mutations.

In a small cohort of 13 non-syndromic AVSD cases, Al Turki et al. (2014) followed a similar paradigm by sequencing the exomes of these parent-offspring

trios. After identifying *de novo* missense mutations in nine genes and collecting exome sequences on another 112 unrelated AVSD cases and comparing them to 5,194 population-matched controls, they found *NR2F2* to be significantly enriched in rare missense variants. *NR2F2* is a transcriptional regulator with both activating and repressive functions when binding to known sequence motifs. They performed a luciferase assay with their found variants to discover that two of the six variants increased luciferase expression, two decreased expression, and two had no effect over controls.

Most recently, Priest et al. (2016) performed exome sequencing on 59 trios with an AVSD probands and unaffected parents. They applied a rare-disease inheritance model, assuming causal variants would be rare and have a large effect, by filtering for those that were *de novo*, homozygous, or compound heterozygous. This produced 710 variants across 399 genes. They then prioritized affected genes whose orthologs were expressed in mouse feta heart tissue and found mutated the previously AVSD-associated *GATA4, GATA6, NKX2-5,* and *CRELD1*. They examined mutations in genes that are co-expressed along with these genes and discovered two individuals with *de novo* missense mutations in *KCNJ3* and *NR1D2*. While the *KCNJ3* variant is seen at low frequencies within external control populations, the *NR1D2* mutation had only been seen once in 61,468 putatively health controls, providing strong evidence for its role in AVSD.

These studies show the progress made toward the identification of deleterious mutations and how they may perturb heart development. However, the fact that many of the mutations are not fully penetrant and that they associate with

different forms of CHDs highlights the heterogeneity of aberrant heart development.

I.III Down Syndrome (DS) as a Model to Identify Genetic Variation Contributing to CHD

CHDs in general associate with chromosomal abnormalities: 12.3% of 4,430 infants with CHDs identified through the Metropolitan Atlanta Birth Defect Program had aberrant cytogenetic results via fluorescence in situ hybridization (Hartman et al., 2011). Down syndrome (DS) was the most commonly associated chromosome abnormality, accounting for 53% of those with chromosome abnormalities. The combination of a CHD with another birth defect increases the rate of associating chromosomal anomalies to 25%, demonstrating how large chromosomal aberrations can have systemic developmental effects (Richards et al., 2008). If those with chromosome abnormalities are excluded, only one in 10,000 infants are born with AVSD (Parker et al., 2010; Hartman et al., 2011).

The risk of AVSD increases 2,000 fold in the Down syndrome population, leading to a rate of 1 in 5 (Freeman et al., 2008).

Therein lies the basis of this dissertation—that DS is a relatively common syndrome and provides a sensitized genetic background (trisomy 21) on which we can uncover AVSD risk variants. We hypothesize that the effect size of AVSD risk

variants are amplified and statistically detectable in cohorts that can be easily obtainable. That the rate of DS live births is 1 in 732, a large cohort has been feasible to assemble, albeit with great determination (Freeman et al., 2007). Based on the data from the National Down Syndrome Project cohort (collected between 2000-2004), 44% of the 1,469 eligible infants with DS had a CHD. Of those with CHD, 39% had an AVSD, 42% had a secundum atrial septal defect, 43% had an ASD and 6% had tetrology of Fallot. Significant to our genetic study is that they also observed a difference in AVSD rates between ethnicities. Compared to non-Hispanic Caucasians, African Americans were twice as likely to have AVSD (OR: 2.06; 95% CI: 1.32-3.21) and Hispanics were half as likely to have AVSD (OR: 0.48; 95% CI: 0.30-0.77). Consistent with a genetic etiology, among infants with DS reporting as African Americans, those with AVSD had significantly more Sub-Saharan African alleles than those without CHDs. Sex also plays a role in DS+AVSD prevalence with females having nearly twice the risk for AVSD compared with males (OR: 1.93; 95% CI: 1.40-2.67).

I.III.i *CHD gene discovery in Down syndrome*

Significant effort was spent on traditional genetic mapping techniques on chromosome 21 in DS to identify genes associated with CHD. The extracellular matrix functioning collagen genes *COL6A1* and *COL6A2* were identified early on as a CHD candidates (Davies et ak.,1995). Segmental or partial trisomies, whereby variable number of segments of chromosome 21 are present due to translocations

or deletions, have been useful in fine mapping CHD causal genes in DS. For example, Korbel et al. (2009) genotyped chromosome 21 in 30 individuals with partial trisomy 21 with and without CHDs. While they found segmental triplication of the distal locus containing *DSCAM* in all individuals with AVSD, it was not fully penetrant. That is, they found this region triplicated in individuals with partial trisomy without CHD. The lack of a consistent pattern is striking.

Traditional gene mapping has also found AVSD-risk genes outside of chromosome 21 in DS. *CRELD1* mutations were the first to be identified in a DS cohort. Two heterozygous missense mutations in *CRELD1* were discovered in two out of 39 individuals with DS (Maslen et al. 2006). Neither mutation was found in euploid controls, nor in a sample of people with DS and no CHD; however, one of the identified mutations was maternally inherited. Applying modern GWAS techniques, our group tested the hypothesis that common variants could increase the odds of AVSD by >2-fold by comparing the genotypes of 210 case individuals with DS and a complete AVSD to 242 controls with DS and no heart defect (Ramachandran et al., 2015). No genome-wide significant SNPs were found, but a number of loci were nominally significant. Two regions were on chromosome 21, one within *PDXK* and the other in *KCNJ6,* which is expressed in fetal heart. The other four suggestive loci indicated *NPHP4* at 1p36.3, mutations of which have been associated with cardiac laterality defects, *MED10* at 5p15.31, which cause cardiac cushion defects in zebrafish mutants, *FZD6* at 8q22.3 which is expressed in the heart is a receptor in Wnt signaling, and 17q22, which is 5 kb from a regulatory region that is a binding site for the transcription factors required for heart development,

GATA1/2/3 and *NR2F2*.

More recently, in a sequencing screen of nine candidate genes, a missense *ALK2* was found as a likely causal mutation in DS-associated AVSD (Joziasse et al., 2011). *ALK2* functions at an early stage in cardiac development as a type I BMP receptor. In a luciferase assay, the found mutation resulted in significantly reduced expression due to its impaired binding by its activator *BMP6*. In a zebrafish *ALK2* knockout model, expressing the variant *ALK2* was significantly poorer than the human wildtype allele at rescuing the mild dorsalization defect of the knockout.

Ackerman et al., 2012, sequenced 26 candidate genes shown to lead to AVSD phenotypes when disrupted. They compared 141 individuals with DS and a complete AVSD to 141 individuals with DS and no CHD and found an excess of predicted deleterious variants in their cases with AVSD. The most significant genes housing recurrent variants were *COL6A1, COL6A2, CRELD1, FBLN2, FRZB and GATA5*. All of these genes play a role in the vascular endothelial growth factor (VEGF) pathway. The mitogen VEGF-A is necessary for proper atrioventricular endocardial cushion and valve formation and too low or too high expression can lead to CHDs (Lambrechts and Carmeliet 2004).

I.III.ii *Mouse models of CHD in Down syndrome*

Mouse models are useful to explore molecular mechanisms of development on a controlled genetic background. Over-expressing the human chromosome 21 genes *DSCAM* and *COL6A2* in developing mouse hearts resulted in 42% neonatal

mortality and enlarged left ventricles and interventricular septum (Grossman et al., 2011). The overexpression of either gene independently however, did not result in noticeable differences from wildtype.

A large portion of human chromosome 21 (Hsa21) is syntenic with mouse chromosome 16 (Mmu16), which contains at least 243 Hsa21 orthologs. The remainder of Hsa21 syntenic loci lie on Mmu10 and Mmu17. The first DS-like mouse generated was Ts65Dn that has large portions of Mmu16 triplicated (Davisson et al., 1990). While these mice exhibit a diverse presentation of heart defect phenotypes, they also carry increased copy of other MMu16 genes not trisomic in DS (Williams et al., 2008). The Tc1 DS mouse model has a freely segregating copy of Hsa21 (O'Doherty et al., 2005). It also displays heart defects but the mice do not carry the extra Hsa21 in all cells, complicating the assessment of trisomy effects (Reeves, 2006).

More recently, chromosome-engineering technology has allowed for mouse models to be constructed with more precise trisomic regions. Lana-Elola et al. (2016) built seven mouse strains with progressively smaller triplications of Hsa21 syntenic loci from 23 Mb to 1.5 Mb. Mice bred with the largest trisomic region successfully modeled similar rates of CHDs as in humans with DS at 61.5% though the distribution of CHD types varied. Genetic modifiers cannot easily explain this incomplete penetrance and variable expressivity, as all mice were inbred and isogenic. Their mice with smaller Hsa21 syntenic trisomies within the largest 23 Mb region failed to produce similar rates of CHD, though most did show significantly increased rates. That two broods of mice, each with different Hsa21 syntenic

trisomic regions, can both increase risk of CHDs again highlights the disease's heterogeneity.

I.IV Copy Number Variants Contribute to the Susceptibility of CHD

I.IV.i *CNV origins*

Copy number variants are stretches of DNA that are duplicated or deleted as compared to a reference genome. They can range in size from a single base or two in the case of insertion-deletions, up to entire chromosomes in the case of aneuploidy. Canonically, they have been described as being from approximately 1 kilobase to over 100 megabases, though this definition is mostly bound by the resolution of available detection technology (Sharp et al., 2005). They arise within the genome from a variety of mechanisms, the most recognizable being non-allelic homologous recombination (NAHR). As the name implies, during recombination in meiosis I, paralogous sequences can recombine, resulting in one daughter chromatid with a copy number gain of the intervening locus and a copy number loss in its homologous chromatid. The frequency of NAHR-mediated CNV creation is positively correlated with the length of flanking low-copy repeats and negatively correlated with the repeats distance apart (Liu et al., 2011). NAHR is responsible for the majority of recurrent CNVs in the human population (Bailey et al, 2002).

Other proposed mechanisms leading to CNV formation include non-

homologous end joining (NHEJ), fork stalling and template switching (FoSTeS), and microhomology-mediated break-induced replication (MMBIR). NHEJ results from the aberrant repair of chromosomal breaks and likely leads to the majority of unbalanced translocations (Weckselblatt et al., 2015). FoSTeS and MMBIR can both produce complex rearrangements ranging up to megabases in size and have been proposed as evolutionary mechanisms of gene duplication and exon shuffling (Zhang et al., 2009). Each of these mechanisms has the potential to produce deleterious or benign CNVs.

The most comprehensive analysis of structural variants in the human population was recently published by The 1000 Genomes Project, cataloging SVs in 2,504 control individuals across 26 ethnic groups from five continents (Sudmant et al., 2015). By performing whole genome sequencing and applying numerous copy number calling algorithms, they identified 68,818 SVs. Ranging in size from a few bases to over 1 Mb, 61% were deletions and 8.7% were duplications, 4.2% were multi allelic (presenting as both deletions and duplications), 1.1% were inversions, 0.2% were mitochondrial insertions; and the remaining 24% were mobile insertion elements like *Alu*s and L1s. The majority of variants were rare with 65% found at frequencies < 0.2%. Variation followed expected evolutionary lineages with ~25% of rare CNVs being isolated to Africans. They estimated the mutation rate to be 0.226 deletions per human genome. Interestingly, 240 genes were observed to be nonessential, as they were found to be homozygously deleted in at least one individual. Phenomenally, CNVs represent the overwhelming majority of variation at the base level in a single human genome, with median coverage of 8.9 Mbp

compared to 3.6 Mbps by single nucleotide polymorphisms.

I.IV.ii *CNVs and disease*

CNVs have the potential to remove or multiply protein-coding sequences or disrupt regulatory motifs and thus have an enormous risk to be deleterious and disease causing. Of the 2,785 disease causing genes in the OMIM Morbid Map, 95% are overlapped by CNVs registered in The Database of Genomic Variants (MacDonald et al., 2014). Dozens of syndromes are caused by CNVs such as Williams-Beuren (Perez Juardo et al., 1996), Potocki-Lupski (Potocki et al., 2000), DiGeorge (Carey et al., 1992) and DiGeorge's reciprocal duplication 22q11.2 (Ensenauer et al., 2003).

CNVs also increase the risk for common, complex diseases. Sebat et al. (2007) looked in a small cohort and found large *de novo* CNVs in 10% of 118 sporadic cases of autism spectrum disorder (ASD) and only 1% in 196 controls, which was statistically significant. While these CNVs covered a wide range of different genomic loci, Moreno-De Luca et al. (2010) found a rare, recurrent 1.4 Mb deletion at 17q12 in 2 out of 1,182 individuals with ASD, in 4 out of 6,340 people with schizophrenia and never in 47,929 controls. Similarly, Mulle et al. (2010) found a recurrent deletion at 3q29 six times in 7,545 individuals with schizophrenia and once in 39,748 controls, thus increasing the odds of schizophrenia by 17 fold.

Other studies have discovered an increased overall burden of CNVs associated with disease, however simple burden tests can be biased toward large genes as they have a greater likelihood of being intersected by a variant.

Raychaudhuri et al. (2010) overcame this bias through the creation of a cnv-enrichment-test that adjusts for the size of tested CNVs against a pre-determined gene set and assigns a p-value based on permutation. With this test, they found nominally increased burden of rare CNVs in gene sets of neuronal-activity and learning in patients with schizophrenia versus controls.

I.IV.iii *CNVs and CHDs*

Due to the requisite fine-tuning of gene expression for proper heart development, it is not surprising that copy number variants have also been shown to increase risk of CHDs. DS can be considered on the upper extreme of copy number burden with an entire chromosome triplicated. At least one-third of genes on chromosome 21 have shown significantly aberrant expression compared to euploid expression with an enrichment of dysregulated domains associating with the early-replicating loci of nuclear lamina-domains (Vilardell et al., 2011; Letourneau et al., 2014). These trisomy 21 dosage effects do not occur in a vacuum and trans-effects genome wide can be seen in the differential expression of 324 genes not on chromosome 21.

The severe CHD Tetralogy of Fallot has been associated with private, *de novo*, and recurrent CNVs that overlap genes of unknown function as well as those previously associated with CHDs like *TBX1, NOTCH1, and JAG1* (Greenway et al. 2009; Silversides et al. 2012). Other studies have looked for CNVs associated with CHD in cohorts with diverse CHD diagnoses, such as left ventricular obstruction,

conotruncal defects, heterotaxy, hypoplastic left heart syndrome and isolated septal defects. Soemedi et al. (2012) found an increased burden of rare CNVs intersecting genes involved in Wnt signaling. Glessner et al. (2014) found an increased burden of rare CNVs in CHD trios versus control trios, as well as *de novo* CNVs over genes interacting with *NKX2* and *GATA4*.

Tomita-Mitchell et al. (2012) looked for CNVs spanning a prioritized list of 100 candidate genes with evident roles in heart development. Excluding case samples with chromosomal abnormalities, they found an increased burden of large (≥100 kb loss, ≥ 200 kb gain), rare CNVs in their cases versus an internal control cohort. They attributed specific CNVs to be disease causing in 4.3% of CHD cases. A range of CHD types were analyzed in their cohort, and while some were too few to be analyzed statistically, a significant increase in copy number gains was seen in individuals with AVSD + TOF. Including cases of complete AVSD, these 55 cases contained 40 individuals with DS, and while they were not analyzed separately, they did see an enrichment of duplications over the transcription factor *RUNX1* on chromosome 21 in cases. While their study replicated and provided further support for previous findings, detecting new CNVs of modest effect size was unlikely given their sample size and heterogeneity of CHD phenotypes.

Overcoming this heterogeneity limitation, we have previously published results on the association of large CNVs genome-wide on a well-phenotyped cohort with 210 case individuals with DS and a complete AVSD and 242 control individuals with DS and no CHD (Ramachandran et al. 2014). We showed a statistically significant increase in large, rare deletions in cases that also impacted more deleted

genes than in controls. Gene set enrichment tests suggested an enrichment of large deletions intersecting ciliome genes, although this finding was not statistically significant. Most importantly, the scale of this study showed that even in the sensitized DS population, there are no large, common CNVs with a major effect on AVSD that could account for the 2000-fold increased risk in DS.

Sailani et al. (2013) measured smaller CNVs on chromosome 21 in a similarly defined case-control DS cohort. Their methodology allowed for CNVs as small as 100 bp to be called and they found three CNVs associated with AVSD. CNV1 at chr21: 42,066,443–42,071,313 (hg18) was not seen in their 53 controls with DS while being present in DS+AVSD cases as deletions and duplications at 18% and 7% frequencies, respectively. No deletions were found in 62 healthy euploid controls, but 9% showed a duplication by qPCR. They also attempted to validate CNV1 with eight Nanostring nCounter probes in 49 DS+AVSD cases and 45 DS-without CHD controls. Two of the eight probes showed significantly fewer copies in cases versus controls and two other probes showed marginally fewer copies by Mann-Whitney U-test. The other four probes did not show any difference in copy number.

CNV2 at chr21:42,284,480–42,286,300 intersects *ZBTB21*. 11% of their controls contained a duplication but none had deletions. In cases, duplications and deletions were seen at 14% and 24% frequencies, respectively, which is a statistically significant difference in CNV rates with risk ratio (95% CI) of 1.85 (1.33-2.56). 14% of euploid controls contained duplications and none contained deletions measured with qPCR. Six out of seven tested Nanostring probes indicated significantly fewer copies in cases versus controls in their DS replication cohort.

CNV3 at chr21:45,541,600–45,555,054 was never seen in DS controls while being present in DS+AVSD cases as a deletion and duplication at 12% and 14% frequencies, respectively. qPCR showed that 16% of euploid controls had duplications at CNV3; however, none of 11 tested Nanostring probes indicated a difference in copy number counts between cases and controls in their DS replication cohort.

The result found by Sailani and colleagues differs from other genetic studies of AVSD, as it provides examples of two common variants with a large effect on disease in individuals with DS. Other studies have shown increased burden of CNVs in disease or highly penetrant rare variants. The uniqueness of Sailani's surprising result necessitates an independent replication study. If these CNVs continue to validate in other DS cohorts they would create exciting research opportunities in elucidating the mechanism in which they perturb endocardial cushion and AV-valve development. For example, CNV2 overlaps a transcriptional repressor *ZBTB21* that has not previously been associated with CHDs, but has a role in WNT-signaling (Wang et al. 2005; Glatter et al. 2009). As CNV1 is intergenic, it would be important to understand how non-transcribed DNA regulates tissue development and could help define future targets for CHD research that is mostly unexplored.

I.V The Ciliome: A Candidate Pathway Involved in AVSD

I.V.i *Overview of cilia and heart development*

In the beginning, there are cilia. Motile and sensory cilia determine the left-right axis at the embryonic node during gastrulation. Defects in cilia are well known to cause laterality defects like heterotaxy (Norris 2012). The distribution of primary cilia in the developing heart is not well understood. At embryonic day 9.5 of mouse embryos (~E28 in humans), primary cilia line the primordial atrial epithelia and the endothelial lining of the developing cardiac cushion during heart tube looping (Slough et al., 2008). Days later, at E12.5 in mice, primary cilia are present on the endocardial cushion's epicardium and mesenchymal cells, while simultaneous being depleted from the cushion's endothelium through the pressure of mechanical shearing (Slough et al., 2008; Iomini et al., 2004).

Cilia are integral in multiple intercellular signaling pathways, defects in which cause a multitude of heart defects (Figure 1.7). The Ellis van Creveld syndrome (EVC) proteins interact with smoothened, the activator of hedgehog signaling which relies on cilia compartmentalization, and are expressed in the ciliated outflow tract, atrial septa mesenchyme, and atrioventricular cushions (Sund et al., 2009). TGFβ/BMP signaling also relies on cilia and is requisite in formation of the outflow tract and AV canal (Arthur and Bamforth, 2011). Outflow tract and ventricular trabeculation also rely on Notch signaling, which in turn requires the cilia's extension to physically interact with another cell (High and Epstein, 2008).

I.V.ii *Cilia defects and AVSD*

To reveal genetic pathways deregulated in DS-associated AVSD, Ripoll et al. (2012) cultured lymphoblastic cell lines (LCLs) from seven DS individuals with AVSD, eight DS+ASD, six DS+VSD and 22 people with DS without a CHD (DS+NH). They measured gene expression with microarrays containing 48,701 probes and found 9,758 genes expressed across the cohort. Principal component analysis (PCA) revealed a clustering of DS+ASD and DS+VSD groups away from the DS+AVSD group. This finding was expected as the etiology of AVSD is thought to be different from the common pathologies associated with ASDs and VSDs. Based on PCA, they combined the ASD and VSD groups and performed burden analyses of the 889 genes differentially expressed between DS+AVSD cells and DS+NH and the 1,766 genes differentially expressed between DS+(ASD/VSD) and DS+NH. They tested for enrichment in the Notch, Wnt, Jak-Stat, Hedgehog, Epithelial-Mesenchymal Transition, Angiogenesis/Cardiogenesis, and Ciliome gene sets. In the ASD/VSD group, they found a significant enrichment of differentially expressed genes in the Jak-Stat ($p = 0.008$), Hedgehog ($p = 0.012$) and Ciliome ($p = 0.00027$) gene lists. In AVSD they observed a significant enrichment only in the Ciliome gene set ($p = 0.00017$). Over 11% of the Ciliome genes expressed in DS+NH LCLs were differentially expressed in DS+AVSD LCLs (134 out of 1,134).

Using the same cilia gene list as Ripoll et al., we tested for an enrichment of copy number variants in individuals with DS+AVSDs compared to DS+NH (Ramanchadran et al. 2015). We used gene-set enrichment analysis and found suggestive evidence that rare deletions in DS+AVSD cases are more enriched in cilia genes as compared to DS+NH controls, with 5.2% of case deletions intersecting cilia

genes versus 2.9% in controls (permuted p-value = 0.1).

In a monumental forward genetic mouse screen, Li et al. (2015) performed ultrasounds on 87,355 mutagenized C57BL/6J fetal mice and recovered 218 with CHD. 30% of these had complex heterotaxy and died *in utero*. They also observed high frequencies of double outlet right ventricle (DORV) and AVSD (5%), with and without laterality defects. Mutations were then uncovered in 113 of the mutant lines via exome sequencing, 91 mutations across 61 genes were named pathogenic, as they were homozygous across lines with the same phenotype. 34 of these genes are cilia-related. 22 of these are required in the primary cilia, half of which caused laterality defects and half did not. 15 of the 26 non-cilia genes related to the heart-critical cell signaling pathways SHH, Wnt and TGF-β/BMP, all of which interact with the cilium.

Deleting the Shh receptor in the anterior heart field of transgenic mice caused partial-AVSDs 65% of 23 mice (Briggs et al., 2015). Histological examination of these mice revealed an inhibition of the dorsal mesenchymal protrusion due to reduced numbers of proliferating cells from the secondary heart field. They went on to assess the cross-regulation of Shh and Wnt signaling and found a reduction in downstream Wnt targets and remarkably, a 50% reduction in AVSDs by pharmacologically over-activating Wnt signaling with LiCl.

I.V References

Ackerman, C., Locke, A. E., Feingold, E., Reshey, B., Espana, K., Thusberg, J. et al. (2012). An Excess of Deleterious Variants in VEGF-A Pathway Genes in Down-Syndrome-Associated Atrioventricular Septal Defects. *American journal of human genetics*, *91*(4), 646-659.

Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M.-P., Lindsay, S. et al. (2014). Rare Variants in NR2F2 Cause Congenital Heart Defects in Humans. *American journal of human genetics*, *94*(4), 574-585.

Arthur, H. M., & Bamforth, S. D. (2011). TGFŒ≤ signaling and congenital heart disease: Insights from mouse studies. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *91*(6), 423-434.

Bailey, J. A., Gu, Z., Clark, R. A., Reinert, K., Samonte, R. V., Schwartz, S. et al. (2002). Recent Segmental Duplications in the Human Genome. *Science*, *297*(5583), 1003-1007.

Briggs, L. E., Burns, T. A., Lockhart, M. M., Phelps, A. L., Van den Hoff, M. J. B., & Wessels, A. (2015). Wnt/β-catenin and sonic hedgehog pathways interact in the regulation of the development of the dorsal mesenchymal protrusion. *Developmental Dynamics*, 103-113.

Carey, A. H., Kelly, D., Halford, S., Wadey, R., Wilson, D., Goodship, J. et al. (1992). Molecular genetic study of the frequency of monosomy 22q11 in DiGeorge syndrome. *Am J Hum Genet*
American Journal of Human Genetics, *51*(5), 964-970.

Chakravarti, A. (1999). Population genetics -- making sense out of sequence. *Nat Genet*, *21*, 56-60.

Chen, Y., Mao, J., Sun, Y., Zhang, Q., Cheng, H.-B., Yan, W.-H. et al. (2010). A novel mutation of GATA4 in a familial atrial septal defect. *Clinica chimica acta; international journal of clinical chemistry*, *411*(21-22), 1741-1745.

Cordell, H. J., Bentham, J., Topf, A., Zelenika, D., Heath, S., Mamasoula, C. et al. (2013). Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nature genetics*, *45*(7), 822-824.

DAVIES, G. E., HOWARD, C. M., FARRER, M. J., COLEMAN, M. M., BENNETT, L. B., CULLEN, L. M. et al. (1995). Genetic variation in the COL6A1 region is associated with congenital heart defects in trisomy 21 (Down's syndrome). *Annals of Human Genetics*, *59*(3), 253-269.

Davisson, M. T., Schmidt, C., & Akeson, E. C. (1990). Segmental trisomy of murine chromosome 16: a new model system for studying Down syndrome
. *Prog Clin Biol Res.*, *360*, 263-280.

De Calisto, J., Araya, C., Marchant, L., Riaz, C. F., & Mayor, R. (2005). Essential role of non-canonical Wnt signalling in neural crest migration. *Development*, *132*(11), 2587-2597.

Digilio, M. C., Marino, B., Cicini, M. P., Giannotti, A., Formigari, R., & Dallapiccola, B. (1993). Risk of congenital heart defects in relatives of patients with atrioventricular canal. *American journal of diseases of children (1911)*, *147*, 1295-1297.

Emanuel, R., Somerville, J., Inns, A., & Withers, R. (1983). Evidence of congenital heart disease in the offspring of parents with atrioventricular defects. *British heart journal*, *49*(November 1982), 144-147.

Ensenauer, R. E., Adeyinka, A., Flynn, H. C., Michels, V. V., Lindor, N. M., Dawson, D. B. et al. (2003). Microduplication 22q11.2, an Emerging Syndrome: Clinical, Cytogenetic, and Molecular Analysis of Thirteen Patients. *The American Journal of Human Genetics*, *73*(5), 1027-1040.

Freeman, S. B., Allen, E. G., Oxford-Wright, C. L., Tinker, S. W., Druschel, C., Hobbs, C. A. et al. (2007). The National Down Syndrome Project: Design and Implementation. *Public Health Rep*
Public Health Reports, *122*(1), 62-72.

Freeman, S. B., Bean, L. H., Allen, E. G., Tinker, S. W., Locke, A. E., Druschel, C. et al. (2008). Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. *Genetics in medicine : official journal of the American College of Medical Genetics*, *10*(3), 173-180.

Gajewski, K., Kim, Y., Lee, Y. M., Olson, E. N., & Schulz, R. A. (1997). D-mef2 is a target for Tinman activation during Drosophila heart development. *The EMBO Journal*, *16*(3), 515-522.

Garg, V., Kathiriya, I. S., Barnes, R., Schluterman, M. K., King, I. N., Butler, C. A. et al. (2003). GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature*, *424*, 443-447.

Glatter, T., Wepf, A., Aebersold, R., & Gstaiger, M. (2009). An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Molecular Systems Biology*, *5*(1).

Glessner, J., Bick, A. G., Ito, K., Homsy, J., Rodriguez-Murillo, L., Fromer, M. et al. (2014). Increased Frequency of De Novo Copy Number Variations in Congenital Heart Disease by Integrative Analysis of SNP Array and Exome Sequence Data. *Circulation research*.

Green, E. K., Priestley, M. D., Waters, J., Maliszewska, C., Latif, F., & Maher, E. R. (2000). Detailed mapping of a congenital heart disease gene in chromosome 3p25.

Journal of medical genetics, *37*(8), 581-587.

Greenway, S. C., Pereira, A. C., Lin, J. C., DePalma, S. R., Israel, S. J., Mesquita, S. M. et al. (2009). De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nature genetics*, *41*(8), 931-935.

Grossman, T. R., Gamliel, A., Wessells, R. J., Taghli-Lamallem, O., Jepsen, K., Ocorr, K. et al. (2011). Over-Expression of DSCAM and COL6A2 Cooperatively Generates Congenital Heart Defects. *PLoS genetics*, *7*(11), e1002344.

Hartman, R. J., Rasmussen, S. A., Botto, L. D., Riehle-Colarusso, T., Martin, C. L., Cragan, J. D. et al. (2011). The contribution of chromosomal abnormalities to congenital heart defects: a population-based study. *Pediatric cardiology*, *32*(8), 1147-1157.

High, F. A., & Epstein, J. A. (2008). The multifaceted role of Notch in cardiac development and disease. *Nat Rev Genet*, *9*(1), 49-61.

Hirayama-Yamada, K., Kamisago, M., Akimoto, K., Aotsuka, H., Nakamura, Y., Tomita, H. et al. (2005). Phenotypes with GATA4 or NKX2.5 mutations in familial atrial septal defect. *American journal of medical genetics. Part A*, *135*(1), 47-52.

Hiroi, Y., Kudoh, S., Monzen, K., Ikeda, Y., Yazaki, Y., Nagai, R. et al. (2001). Tbx5 associates with Nkx2-5 and synergistically promotes cardiomyocyte differentiation. *Nat Genet*, *28*(3), 276-280.

Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., Lin, Y. et al. (2013). A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nature genetics*, *45*(7), 818-821.

Iomini, C., Tejada, K., Mo, W., Vaananen, H., & Piperno, G. (2004). Primary Cilia of Human Endothelial Cells Disassemble under Laminar Shear Stress. *The Journal of Cell Biology*, *164*(6), 811-817.

Joziasse, I. C., Smith, K. A., Chocron, S., van Dinther, M., Guryev, V., van de Smagt, J. J. et al. (2011). ALK2 mutation in a patient with Down's syndrome and a congenital heart defect. *Eur J Hum Genet*, *19*(4), 389-393.

Kamp, A., Peterson, M. A., Svenson, K. L., Bjork, B. C., Hentges, K. E., Rajapaksha, T. W. et al. (2010). Genome-wide identification of mouse congenital heart disease loci. *Human Molecular Genetics*, *19*(16), 3105-3113.

Kodo, K., & Yamagishi, H. (2011). A Decade of Advances in the Molecular Embryology and Genetics Underlying Congenital Heart Defects. *Circulation Journal*, *75*(10), 2296-2304.

Koefoed, K., Veland, I. R., Pedersen, L. B., Larsen, L. A., & Christensen, S. T. (2014). Cilia and coordination of signaling networks during heart development. *Organogenesis*, *10*(1), 108-125.

Korbel, J. O., Tirosh-Wagner, T., Urban, A. E., Chen, X.-N., Kasowski, M., Dai, L. et al. (2009). The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(29), 12031-12036.

Kucik, J. E., Shin, M., Siffel, C., Marengo, L., & Correa, A. (2013). Trends in survival among children with Down syndrome in 10 regions of the United States. *Pediatrics*, *131*(1), e27-36.

Lage, K., Møllgård, K., Greenway, S., Wakimoto, H., Gorham, J. M., Workman, C. T. et al. (2010). Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular Systems Biology*, *6*(1).

Lambrechts, D., & Carmeliet, P. (2004). Sculpting heart valves with NFATc and VEGF. *Cell*, *118*, 532-534.

Lana-elola, E., Watson-scales, S., Slender, A., Gibbins, D., Martineau, A., Douglas, C. et al. (2016). Genetic dissection of Down syndrome- associated congenital heart defects using a new mouse mapping panel. *1880*, 1-20.

Le Gloan, L., Mercier, L.-A., Dore, A., Marcotte, F., Ibrahim, R., Mongeon, F.-P. et al. (2011). Recent Advances in Adult Congenital Heart Disease. *Circulation Journal*, *75*(10), 2287-2295.

Letourneau, A., Santoni, F. A., Bonilla, X., Sailani, M. R., Gonzalez, D., Kind, J. et al. (2014). Domains of genome-wide gene expression dysregulation in Down's syndrome. *Nature*, *508*(7496), 345-350.

Li, Q. Y., Newbury-Ecob, R. A., Terrett, J. A., Wilson, D. I., Curtis, A. R., Yi, C. H. et al. (1997). Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature genetics*, *15*, 21-29.

Li, Y., Klena, N. T., Gabriel, G. C., Liu, X., Kim, A. J., Lemke, K. et al. (2015). Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature*, *521*, 520-524.

Liu, P., Lacaria, M., Zhang, F., Withers, M., Hastings, P. J., & Lupski, J. (2011). Frequency of Nonallelic Homologous Recombination Is Correlated with Length of Homology: Evidence that Ectopic Synapsis Precedes Ectopic Crossing-Over. *The American Journal of Human Genetics*, *89*(4), 580-588.

Loffredo, C. A. (2000). Epidemiology of cardiovascular malformations: prevalence

and risk factors. *American journal of medical genetics*, *97*(4), 319-325.

Lyons, I., Parsons, L. M., Hartley, L., Li, R., Andrews, J. E., Robb, L. et al. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes & Development*, *9*(13), 1654-1666.

MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, *42*(D1), D986-D992.

Mai, C. T., Isenburg, J., Langlois, P. H., Alverson, C. J., Gilboa, S. M., Rickard, R. et al. (2015). Population-based birth defects data in the United States, 2008 to 2012: Presentation of state-specific data and descriptive brief on variability of prevalence. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *103*(11), 972-993.

Maitra, M., Koenig, S. N., Srivastava, D., & Garg, V. (2010). Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatric research*, *68*(4), 281-285.

Maslen, C. L., Babcock, D., Robinson, S. W., Bean, L. J. H., Dooley, K. J., Willour, V. L. et al. (2006). CRELD1 mutations contribute to the occurrence of cardiac atrioventricular septal defects in Down syndrome. *American journal of medical genetics. Part A*, *140A*, 2501-2505.

Miller, A., Siffel, C., Lu, C., Riehle-Colarusso, T., Fr\'\ias, J. L., & Correa, A. (2010). Long-term survival of infants with atrioventricular septal defects. *The Journal of pediatrics*, *156*(6), 994-1000.

Moreno-De-Luca, D., Mulle, J. G., Kaminsky, E. B., Sanders, S. J., Myers, S. M., Adam, M. P. et al. (2010). Deletion 17q12 is a recurrent copy number variant that confers high risk of autism and schizophrenia. *American journal of human genetics*, *87*(5), 618-630.

Mulle, J. G., Dodd, A. F., McGrath, J. A., Wolyniec, P. S., Mitchell, A. A., Shetty, A. C. et al. (2010). Microdeletions of 3q29 confer high risk for schizophrenia. *American journal of human genetics*, *87*(2), 229-236.

Norris, D. P. (2012). Cilia, calcium and the basis of left-right asymmetry. *BMC Biology*, *10*(1), 1-8.

O'Doherty, A., Ruf, S., Mulligan, C., Hildreth, V., Errington, M. L., Cooke, S. et al. (2005). An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science (New York, N.Y.)*, *309*(5743), 2033-2037.

Parker, S. E., Mai, C. T., Canfield, M. A., Rickard, Russel, Wang, Y. et al. (2010). Updated National Birth Prevalence Estimates for Selected Birth Defects in the

United States, 2004-2006. *BIRTH DEFECTS RESEARCH PART A-CLINICAL AND MOLECULAR TERATOLOGY*
Birth Defects Res. Part A-Clin. Mol. Teratol., *88*(12, SI), 1008-1016.

Perez Jurado, L. A., Peoples, R., Kaplan, P., Hamel, B. C., & Francke, U. (1996). Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. *Am J Hum Genet*
American Journal of Human Genetics, *59*(4), 781-792.

Potocki, L., Chen, K.-S., Park, S.-S., Osterholm, D. E., Withers, M. A., Kimonis, V. et al. (2000). Molecular mechanism for duplication 17p11.2[mdash] the homologous recombination reciprocal of the Smith-Magenis microdeletion. *Nat Genet*, *24*(1), 84-87.

Priest, J. R., Osoegawa, K., Mohammed, N., Nanda, V., Kundu, R., Schultz, K. et al. (2016). <italic>De Novo</italic> and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. *PLoS Genet*, *12*(4), e1005963.

Qian, L., & Bodmer, R. (2009). Partial loss of GATA factor Pannier impairs adult heart function in Drosophila. *Human Molecular Genetics*, *18*(17), 3153-3163.

Ramachandran, D., Mulle, J. G., Locke, A. E., Bean, L. J. H., Rosser, T. C., Bose, P. et al. (2014). Contribution of copy-number variation to Down syndrome-associated atrioventricular septal defects. *Genetics in medicine : official journal of the American College of Medical Genetics*, October), 1-7.

Ramachandran, D., Zeng, Z., Locke, A. E., Mulle, J. G., Bean, L. J. H., Rosser, T. C. et al. (2015). Genome-Wide Association Study of Down Syndrome-Associated Atrioventricular Septal Defects. *G3: Genes$_{Genomes}$Genetics*, *5*(10), 1961-1971.

Rasiah, S. V., Ewer, A. K., Miller, P., Wright, J. G., Tonks, A., & Kilby, M. D. (2008). Outcome following prenatal diagnosis of complete atrioventricular septal defect. *Prenatal diagnosisdiagnosis*, *28*, 95-101.

Raychaudhuri, S., Korn, J. M., McCarroll, S. A., Altshuler, D., Sklar, P., Purcell, S. et al. (2010). Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS genetics*, *6*(9).

Reamon-Buettner, S. M., & Borlak, J. (2006). HEY2 mutations in malformed hearts. *Human Mutation 27*(1), 118.

Reamon-Buettner, S. M., Ciribilli, Y., Traverso, I., Kuhls, B., Inga, A., & Borlak, J. (2009). A functional genetic study identifies HAND1 mutations in septation defects of the human heart. *Human Molecular Genetics*, *18*(19), 3567-3578.

Reller, M. D., Strickland, M. J., Riehle-Colarusso, T., Mahle, W. T., & Correa, A. (2008). Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005. *The Journal of pediatrics*, *153*(6), 807-813.

Richards, A. A., Santos, L. J., Nichols, H. A., Crider, B. P., Elder, F. F., Hauser, N. S. et al. (2008). Cryptic Chromosomal Abnormalities Identified in Children With Congenital Heart Disease. *Pediatr Res*, *64*(4), 358-363.

Ripoll, C., Rivals, I., Ait Yahya-Graison, E., Dauphinot, L., Paly, E., Mircher, C. et al. (2012). Molecular signatures of cardiac defects in down syndrome lymphoblastoid cell lines suggest altered ciliome and hedgehog pathways. *PloS one*, *7*(8), e41616.

Rupp, P. A., Fouad, G. T., Egelston, C. A., Reifsteck, C. A., Olson, S. B., Knosp, W. M. et al. (2002). Identification, genomic organization and mRNA expression of CRELD1, the founding member of a unique family of matricellular proteins. *Gene*, *293*(1,Äì2), 47-57.

Russo, A., & Elixhauser, A. (2007). STATISTICAL BRIEF # 24 Hospitalizations for Birth Defects, 2004. *International Classification*, 1-9.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T. et al. (2007). Strong association of de novo copy number mutations with autism. *Science (New York, N.Y.)*, *316*(5823), 445-449.

Sharp, A. J., Locke, D. P., McGrath, S. D., Cheng, Z., Bailey, J. A., Vallente, R. U. et al. (2005). Segmental Duplications and Copy-Number Variation in the Human Genome. *The American Journal of Human Genetics*, *77*(1), 78-88.

Sheffield, V. C., Pierpont, M. E., Nishimura, D., Beck, J. S., Burns, T. L., Berg, M. A. et al. (1997). Identification of a Complex Congenital Heart Defect Susceptibility Locus by Using DNA Pooling and Shared Segment Analysis. *Human Molecular Genetics*, *6*(1), 117-121.

Silversides, C. K., Lionel, A. C., Costain, G., Merico, D., Migita, O., Liu, B. et al. (2012). Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS genetics*, *8*(8), e1002843.

Slough, J., Cooney, L., & Brueckner, M. (2008). Monocilia in the embryonic mouse heart suggest a direct role for cilia in cardiac morphogenesis. *Developmental Dynamics*
Dev. Dyn., *237*(9), 2304-2314.

Soemedi, R., Wilson, I. J., Bentham, J., Darlay, R., T\"opf, A., Zelenika, D. et al. (2012). Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *American journal of human genetics*, *91*(3), 489-501.

Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J. et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature*, *526*(7571), 75-81.

Sund, K. L., Roelker, S., Ramachandran, V., Durbin, L., & Benson, D. W. (2009). Analysis of Ellis van Creveld syndrome gene products: implications for cardiovascular development and disease. *Human Molecular Genetics*, *18*(10), 1813-1824.

Tomita-Mitchell, A., Mahnke, D. K., Struble, C. A., Tuffnell, M. E., Stamm, K. D., Hidestrand, M. et al. (2012). Human gene copy number spectra analysis in congenital heart malformations. *Physiological Genomics*, *44*, 518-541.

Vilardell, M., Rasche, A., Thormann, A., Maschke-Dutz, E., P√©rez-Jurado, L. A., Lehrach, H. et al. (2011). Meta-analysis of heterogeneous Down Syndrome data reveals consistent genome-wide dosage effects related to neurological processes. *BMC Genomics*, *12*(1), 1-16.

Wang, J., Kudoh, J., Takayanagi, A., & Shimizu, N. (2005). Novel human BTB/POZ domain-containing zinc finger protein ZNF295 is directly associated with ZFP161. *Biochemical and Biophysical Research Communications*, *327*(2), 615-627.

Weckselblatt, B., Hermetz, K. E., & Rudd, M. K. (2015). Unbalanced translocations arise from diverse mutational mechanisms including chromothripsis. *Genome Research*, *25*(7), 937-947.

Williams, A. D., Mjaatvedt, C. H., & Moore, C. S. (2008). Characterization of the cardiac phenotype in neonatal Ts65Dn mice. *Developmental Dynamics Dev. Dyn.*, *237*(2), 426-435.

Xu, H., Morishima, M., Wylie, J. N., Schwartz, R. J., Bruneau, B. G., Lindsay, E. A. et al. (2004). Tbx1 has a dual role in the morphogenesis of the cardiac outflow tract. *Development*, *131*(13), 3217-3227.

Yamagishi, H., Yamagishi, C., Nakagawa, O., Harvey, R. P., Olson, E. N., & Srivastava, D. (2001). The Combinatorial Activities of Nkx2.5 and dHAND Are Essential for Cardiac Ventricle Formation. *Developmental Biology*, *239*(2), 190-203.

Yang, Q., Rasmussen, S. A., & Friedman, J. (2002). Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. *The Lancet*, *359*(9311), 1019-1025.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D. et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, *498*(7453), 220-223.

Zhang, F., Khajavi, M., Connolly, A. M., Towne, C. F., Batish, S. D., & Lupski, J. R. (2009). The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet*, *41*(7), 849-853.

Øyen, N., Poulsen, G., Boyd, H. A., Wohlfahrt, J., Jensen, P. K. A., & Melbye, M. (2009). Recurrence of congenital heart defects in families. *Circulation*, *120*(4), 295-301.

Congenital heart defect	Abbreviation	Characteristics
Aortic stenosis	AoS	Obstruction of blood flow between left ventricle and the aorta. This may be caused by abnormalities of the aortic valve (aortic valve stenosis, AVS), muscular obstruction or narrowing of the aorta immediately above the valve.
Atrial septal defect	ASD	Incomplete septation of the atria.
Atrioventricular septal defect	AVSD	Developmental defects that arise from developmental defects of the endocardial cushions. Such defects affect may affect the lower part of the atrial septum, the ventricular septum and the mitral and tricuspid valves.
Coarctation of the aorta	CoA	A narrowing of the aorta.
Hypoplastic left heart syndrome	HLHS	All structures of the left side of the heart, including the left ventricle, mitral and aortic valves, are severely underdeveloped.
Patent ductus arteriosus	PDA	Failure of closure of the ductus arteriosus (DA) at birth. DA is a blood vessel which allows passage between the pulmonary artery and the aorta. This passage allows bypass of the lungs in fetal circulation.
Persistent left superior vena cava	PLSVC	Failure of obliteration of the left superior vein.
Transposition of the great arteries	TGA	The aorta and pulmonary artery are reversed, so that the aorta arises from the right ventricle and the pulmonary artery arises from the left ventricle. The result is that there is no connection between systemic and pulmonary circulation.
Tetralogy of Fallot	TOF	Involves four anatomical abnormalities in the heart: 1) Ventricular septal defect (hole between ventricles) 2) Pulmonary stenosis (pulmonary artery is narrow) 3) Overriding aorta (the aorta is positioned between the two ventricles) 4) Hypertrophic (thickening of) right ventricle.
Ventricular septal defect	VSD	Incomplete septation of the ventricles.

Table 1.1: Common congenital heart defects and their characteristic developmental anomalies. Sources: KOEFOED K., VELAND I.R., PEDERSEN L.B., LARSEN L.A. & CHRISTENSEN S.T. 2014. Cilia and coordination of signaling networks during heart development. *Organogenesis* 10: 108-125.

Cincinnati Childrens Hospital. Heart Institute Encyclopedia. http://www.cincinnatichildrens.org

American Heart Association. About congenital heart diseases. www.heart.org

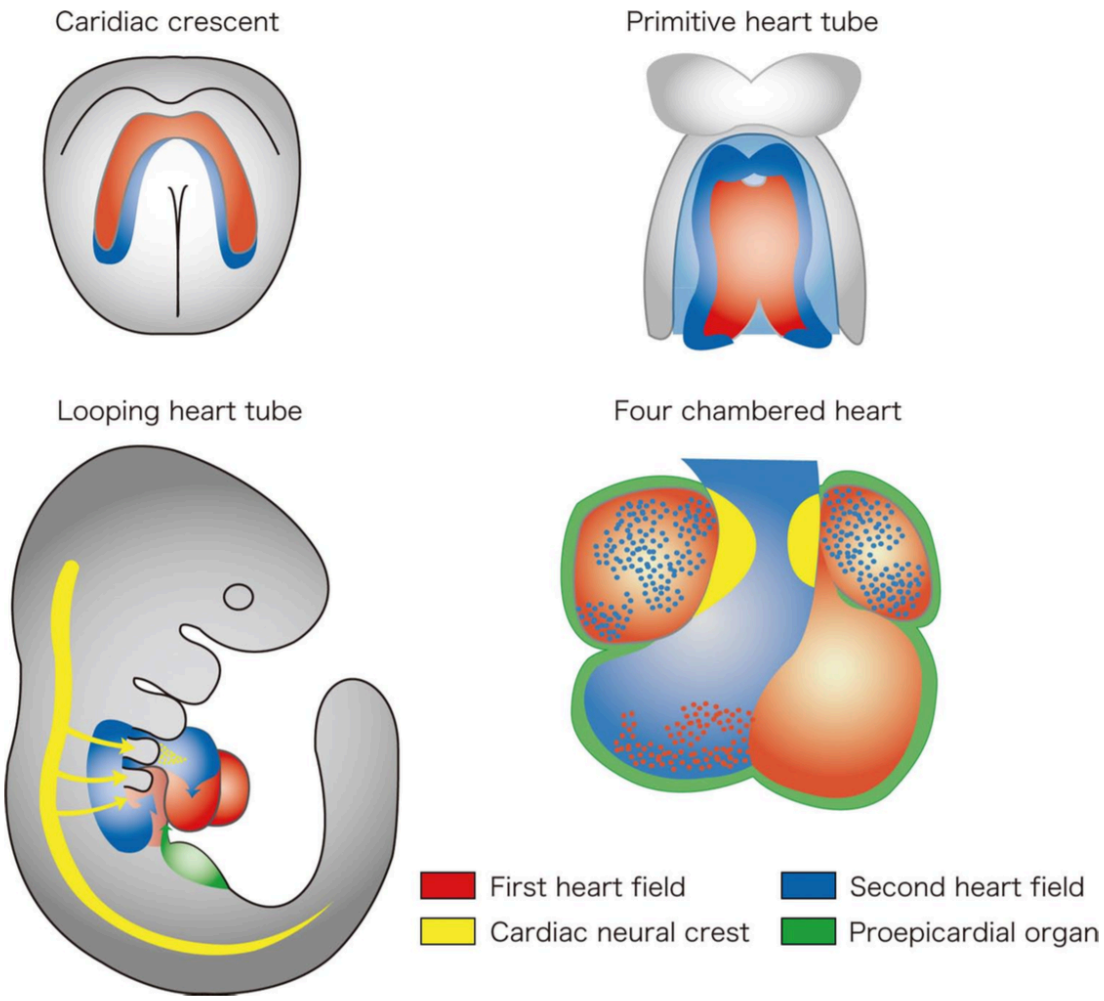


Figure 1.1: General schematic of the major differentiating cell populations in cardiac development.

Source: KODO K. & YAMAGISHI H. 2011. A Decade of Advances in the Molecular Embryology and Genetics Underlying Congenital Heart Defects. *Circulation Journal* 75: 2296-2304.

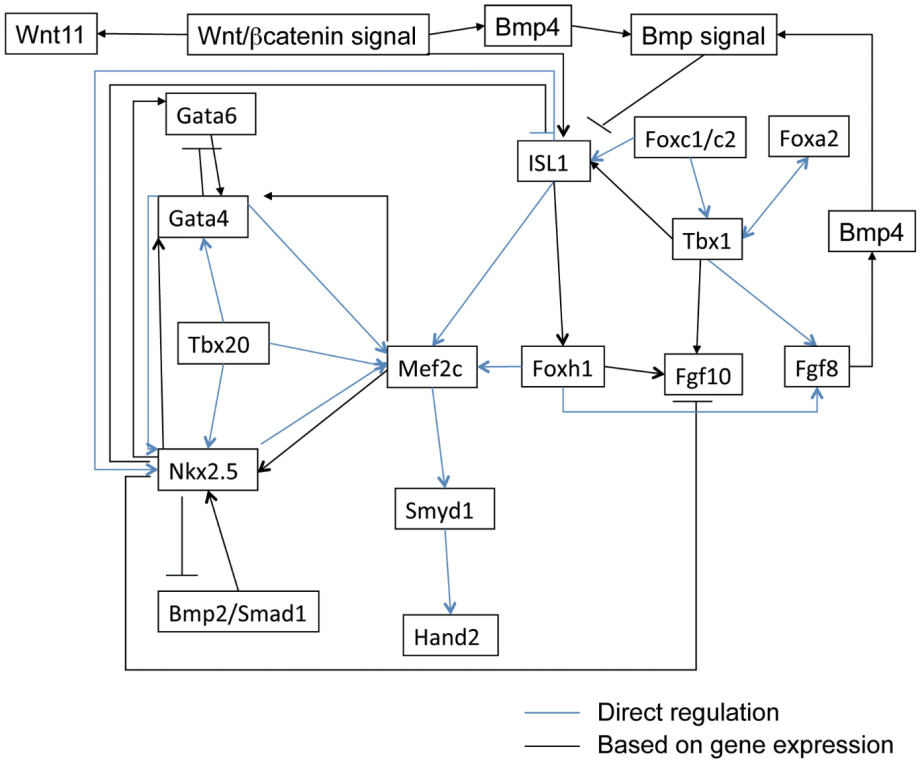


Figure 1.2: Diagram of transcription factor regulation in the secondary heart field. Blue lines indicate direct *in vivo* regulation while black lines indicate a relationship established by genetic expression data.

Source: KODO K. & YAMAGISHI H. 2011. A Decade of Advances in the Molecular Embryology and Genetics Underlying Congenital Heart Defects. *Circulation Journal* 75: 2296-2304.

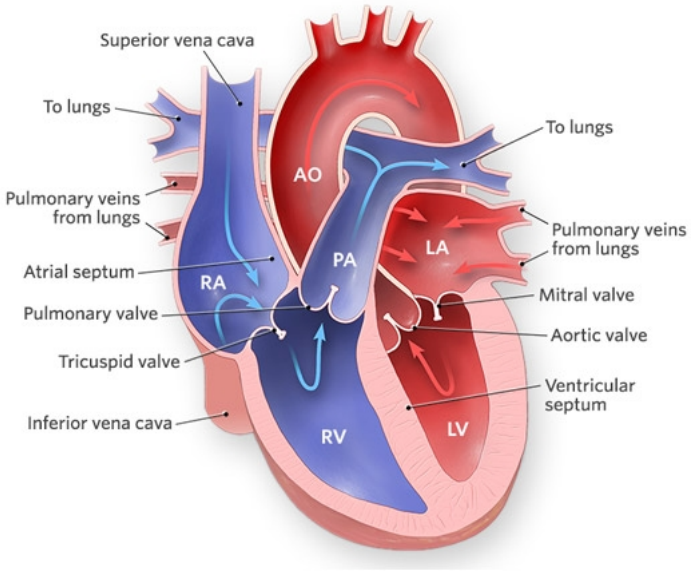


Figure 1.3: Schematic of a normally developed heart.

Source: The Children's Hospital of Philadelphia: www.chop.edu

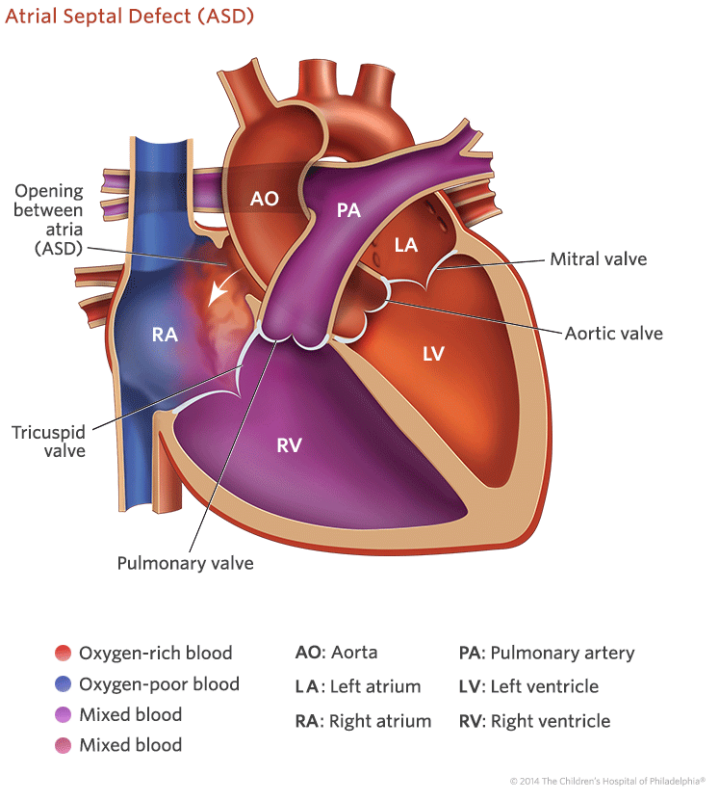


Figure 1.4: Schematic of the common atrial septal defect (ASD).

Source: The Children's Hospital of Philadelphia: www.chop.edu

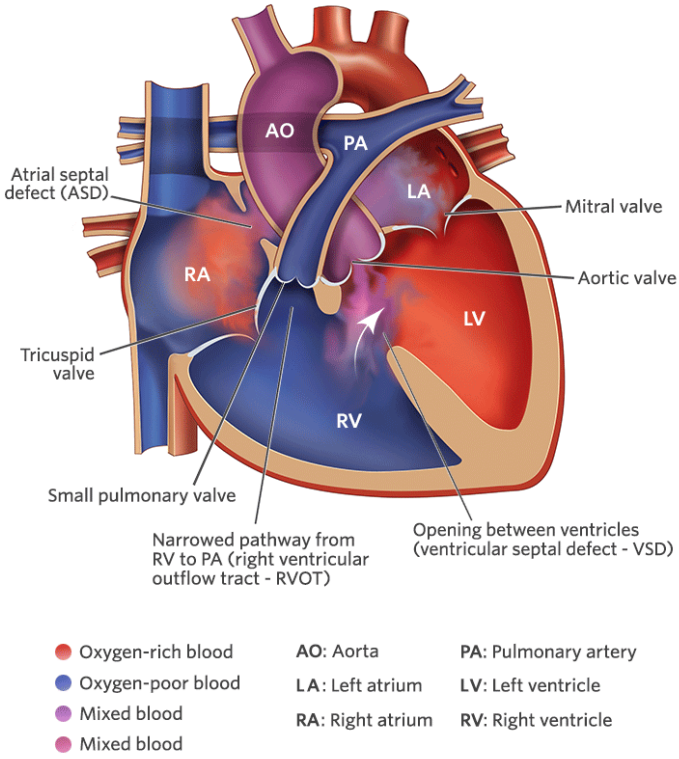


Figure 1.5: Schematic of tetralogy of Fallot (TOF).

Source: The Children's Hospital of Philadelphia: www.chop.edu

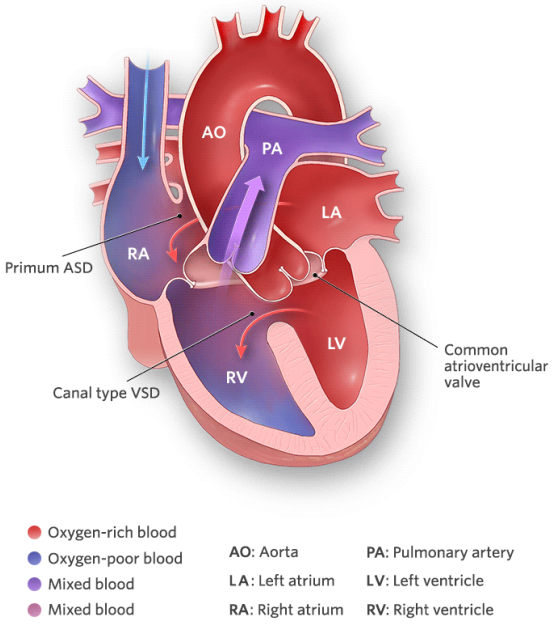


Figure 1.6: Schematic of atrioventricular septal defect (AVSD).

Source: The Children's Hospital of Philadelphia: www.chop.edu

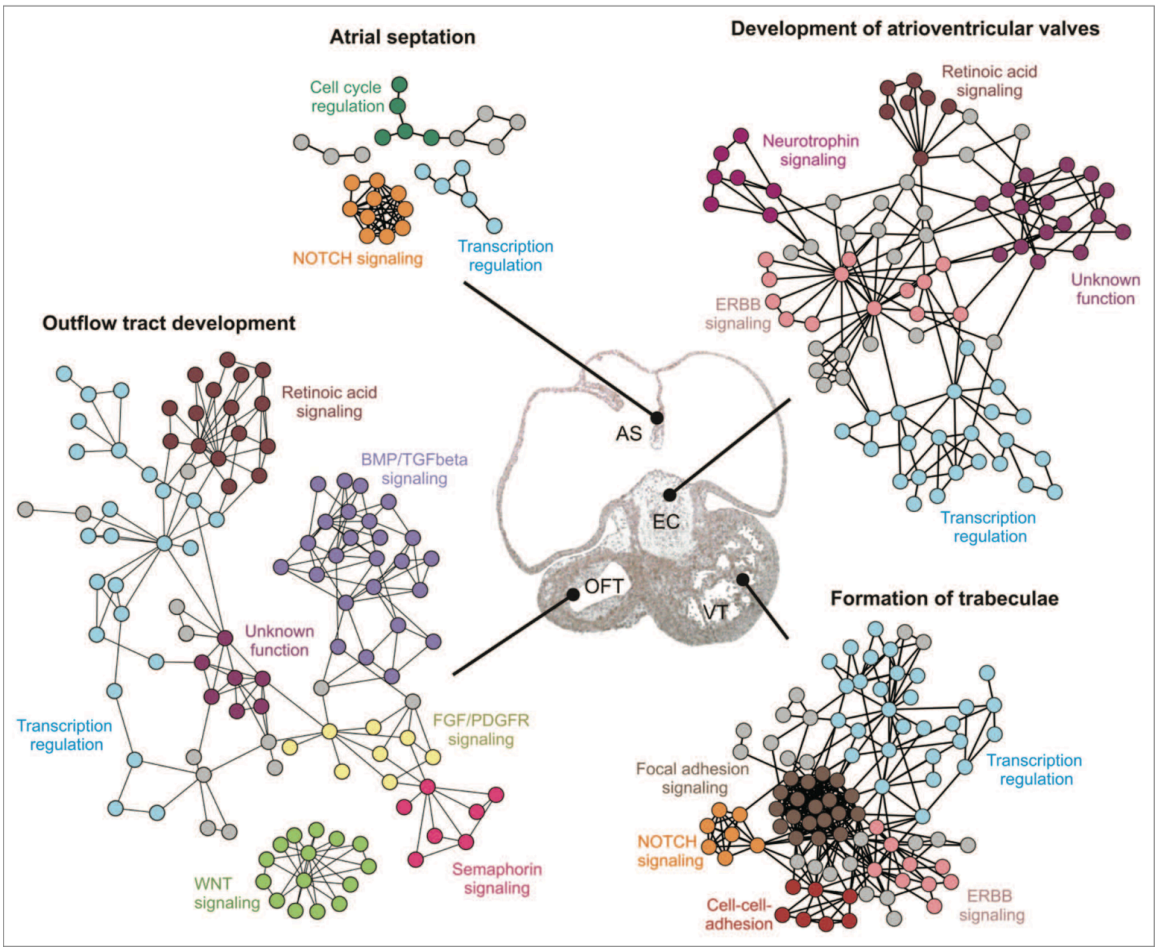


Figure 1.7: Gene ontology networks active during heart development, grouped by heart tissues AS = Atrial septation; EC = endocardial cushion; VT = ventricle; OFT = outflow tract.

Sources: KOEFOED K., VELAND I.R., PEDERSEN L.B., LARSEN L.A. & CHRISTENSEN S.T. 2014. Cilia and coordination of signaling networks during heart development. *Organogenesis* 10: 108-125.

LAGE K., MØLLGÅRD K., GREENWAY S., WAKIMOTO H., GORHAM J.M., WORKMAN C.T., BENDSEN E., HANSEN N.T., RIGINA O., ROQUE F.S., WIESE C., CHRISTOFFELS V.M., ROBERTS A.E., SMOOT L.B., PU W.T., DONAHOE P.K., TOMMERUP N., BRUNAK S., SEIDMAN C.E., SEIDMAN J.G. & LARSEN L.A. 2010. Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Molecular Systems Biology* 6:

Analysis of Copy Number Variants on Chromosome 21 in Down Syndrome Associated Congenital Heart Defects

Benjamin L. Rambo-Martin, Jennifer G. Mulle, David J. Cutler, Lora J.H. Bean, Tracie C. Rosser, Kenneth J. Dooley, Clifton Cua, George Capone, Cheryl L. Maslen, Roger H. Reeves, Stephanie L. Sherman, Michael E. Zwick

ABSTRACT

One in 5 individuals with Down syndrome (DS) are born with an atrioventricular septal defect (AVSD), an incidence 2,000 times higher than in the euploid population. The genetic loci that contribute to this risk are poorly understood, though dosage changes of chromosome 21 may be a factor. Within this study we test two hypotheses: 1) individuals with DS carrying chromosome 21 CNVs that interrupt exons may be protected from AVSD, because these CNVs return AVSD susceptibility loci back to disomy; and 2) individuals with DS carrying chromosome 21 genes spanned by microduplications are at increased risk for AVSD, because these microduplications boost the dosage of AVSD susceptibility loci over a tolerable threshold. We tested 198 case individuals with DS+AVSD and 211 control individuals with DS and a normal heart, using a custom microarray with dense probes tiled on chromosome 21 for array CGH. We followed stringent quality control methods to produce a high confidence set of microdeletions and

microduplications across chromosome 21. We found that no individual CNV, or any individual gene intersected by a CNV, is associated with AVSD in DS. From burden analyses, we found that African American controls had more bases covered by rare deletions than did African American cases. Inversely, we found that Caucasian cases had more genes intersected by rare duplications than did Caucasian controls. We focused on CNVs that disrupted exons and found no difference between cases and controls. We performed gene set enrichment analysis in Caucasians and found suggestive evidence that cilia genes are more likely to contain deletions in controls and more likely to contain duplications in cases, though no difference in cilia genes with exons disrupted by CNVs. Pathway analyses of genes with exons intersected by CNVs show an overrepresentation in controls of genes involved in protein heterotrimerization. In cases, genes that were completely duplicated and thus at a ploidy level of 4x or greater, were enriched in histone methylating proteins. Finally, we show that previously DS+AVSD-associated common CNVs on chromosome 21 are likely false positives. This research adds to the swell of evidence indicating that DS-associated AVSD is similarly heterogeneous as is AVSD in the euploid population.

INTRODUCTION

Understanding the rules by which variation impacting genome dosage impacts phenotypes remains one of the central challenges of human genetics (Zarrei et al., 2015). Down syndrome (DS), caused by trisomy 21, provides an extreme example of a dosage change that impacts multiple aspects of an individual's phenotype. Congenital heart

defects (CHD) are among the most common and significant birth defects found in individual with DS. In the disomic population, CHD are the most common birth defect, presenting in 80 out of 1,000 live births and causing 25% of infant mortality (Reller et al. 2008; Yang et al. 2006; Hartman et al. 2011; Mai et al., 2015). For children with trisomy 21, CHD incidence is substantially increased: nearly 450 out of 1,000 live births have a CHD (Loffredo et al., 2001; Freeman et al., 2008).

Atrioventricular septal defects (AVSD) are a serious CHD resulting from the failure of endocardial cushion and subsequent mitral and tricuspid valve formation. In the presence of an AVSD, there is improper mixing of oxygenated and deoxygenated blood. While the heart is typically repaired during the first year of life, patients with AVSD face increased risk of sequelae including arrhythmias, endocarditis, stroke, congestive heart failure, pulmonary hypertension, and continued heart valve problems (Le Gloan et al. 2011). Trisomy 21 is the single greatest risk factor for AVSD. While 1 in 10,000 people in the general population present with AVSD, among infants with DS, the rate is 1 in 5 (Freeman et al., 2008). This 2000-fold increased risk suggests the hypothesis that the DS population may represent a sensitized population in which genetic variation contributing to the risk of AVSD may have a larger effect size than in the general population. Using this population to identify AVSD risk loci may therefore yield high statistical power, even with a small sample size (Zwick et al., 1999).

In our prior study, the largest genetic study of its kind to date, we characterized genome-wide copy number variants (CNVs) in a well-phenotyped cohort with 210 case individuals with DS and a complete AVSD (DS+AVSD) and 242 control individuals with DS and structurally normal hearts (DS+NH) (Ramachandran et al., 2015). We

showed a statistically significant increase in large, rare deletions in DS+AVSD cases that also impacted more genes than those in DS+NH controls. Gene set enrichment tests suggested an enrichment of large deletions intersecting ciliome genes. Most importantly, the scale of this study showed that even in the sensitized DS population that there are no large, common CNVs with a major effect on AVSD that could account for the 2000-fold increased risk in DS. We have also shown that common SNPs cannot account for the increased risk of CHD in this same cohort (Ramachandran et al., 2015).

In the current study, we focus specifically on CNVs on chromosome 21, and test two primary hypotheses: 1) individuals with DS carrying chromosome 21 deletions may be protected from AVSD, because these deletions return AVSD susceptibility loci back to disomy; or 2) individuals with DS carrying chromosome 21 duplications are at increased risk for AVSD, because these duplications boost the dosage of AVSD susceptibility loci over a tolerable threshold. Furthermore, Sailani et al. (2013) screened for CNVs on chromosome 21 in a similarly defined DS cohort of 55 DS+AVSD cases and 53 DS+NH controls and reported two common CNVs to be significantly associated with AVSD. In addition to testing our primary hypotheses, our larger cohort of 188 DS+AVSD cases and 211 DS+NH controls allow us to attempt to replicate this finding.

RESULTS

We used rigorous quality control (see METHODS) to identify deletions and duplications on the trisomic chromosomes 21 in 409 DS individuals, including 355 Caucasians (174 DS+AVSD cases and 181 DS+NH controls) and 54 African Americans (24 DS+AVSD

cases and DS+NH 30 controls). This analysis revealed a high quality set of 215 individual deletions and 59 individual duplications (Table 2.1). For Caucasians and African Americans respectively, 91% and 100% of these deletions had 50% reciprocal overlap with deletions in the Database of Genomic Variants (MacDonald et al., 2014, http://dgv.tcag.ca). For duplications, 82% and 60% of these variants were reported in the DGV.

No single CNV of large effect is associated with AVSD in DS

We performed association testing of single deletion and duplication regions along chromosome 21 as well as single genes intersected by deletions or duplications, controlling for possible population stratification (see METHODS). Though in Caucasians we had 80% power to detect risk variants of 5% allele frequency with an odds ratio of 2.2 or greater (alpha level of 0.05), no single CNV region was associated with AVSD (Figure 2.1). We also tested for association of single genes with any intersection by CNVs and found no suggestive association. With our small African American cohort, we had 80% power to detect a risk CNV with an odds ratio of 6.3 at an allele frequency of 0.05 and an alpha level of 0.05. Again, we found no single CNV, or any CNV-intersected gene on chromosome 21 associated with AVSD in our Down syndrome population.

Burden of chromosome 21 deletions

We tested our first hypothesis that individuals with DS carrying chromosome 21 deletions may be protected from AVSD, because these deletions return AVSD susceptibility loci back to disomy. To do this, we compared the increased "burden" of chromosome 21 deletions among DS+NH controls. We determined whether there was an increased average number of deletions per person, as an increased number of bases covered by deletions, and as an increased average number of genes intersected by deletions on chromosome 21 using PLINK v1.07 (Purcell et al., 2007). We tested all deletions, filtered by allele frequency (common ≥ 0.01 or rare < 0.01), and whether they were reported in the DGV. Our analyses did not reveal an effect of deletions providing a protective effect against AVSDs in Caucasians (Table 2.2). In contrast, African American DS+NH controls were significantly more likely to have more bases covered by deletions within the full deletion set (average total bases covered by deletions: 33.45 kb in DS+NH controls vs. 13.06 kb in DS+AVSD cases; empirical p-value = 0.04; Table 2.2). When we filtered deletions by frequency, we found that this effect in African Americans was driven by rare variants less than 1% frequency in our study sample: African American DS+NH controls with rare deletions have on average 45.63 kb covered by rare deletions versus 12.8 kb in DS+AVSD cases (empirical p-value = 0.02, Table 2.2).

Burden of chromosome 21 duplications

To test our second hypothesis that chromosome 21 duplications increase the risk for AVSD, we compared the burden of chromosome 21 duplications among DS+AVSD cases compared with DS+NH controls. In Caucasians, a number of findings were

consistent with this hypothesis (Table 2.3). We observed that duplications, on average, affect more bases in cases (83.53 kb) than in controls (40.49 kb) (empirical p-value = 0.09). Caucasian cases also had twice the rate of genes duplicated than did controls (0.22 in cases versus 0.1 in controls; empirical p-value = 0.07). Rare CNVs in Caucasians drive these effects. For example, cases had a higher rate of rare duplications than controls (0.09 in cases versus 0.03 in controls; empirical p-value = 0.06). More specifically, cases have five times the rate of genes intersected by rare duplications (0.16) compared with controls (0.03) (empirical p-value = 0.04). These effects remain by filtering for variants not in the DGV, as they are all rare variants. Given the low number of duplications in the African American samples, we did not observe an increased burden of duplications among DS+AVSD cases (Table 2.3).

Burden of CNV-interrupted genes

We hypothesized that genes on chromosome 21 interrupted by CNVs provide protection from AVSD in individuals with DS by reducing that risk-locus to disomy. We built a set of "reduced to disomy" CNVs by including deletions that intersected an exon, as well as duplications that intersected an exon but did not envelope an entire gene. In African Americans, this reduced the set to only two CNVs, a deletion and duplication in two controls. This is not significant (1-sided Fisher's exact p-value = 0.32). In Caucasians, this produced a set of 41 CNVs in DS+AVSD cases and 41 CNVs in DS+NH controls that reduce a gene back to disomy. We found no indication that individuals with DS

without heart defects are protected by CNVs that reduce a gene back to disomy (Table 2.4).

Burden of chromosome 21 duplicated genes

Genes on chromosome 21 may be tolerated at three copies and incur AVSDs when found at four or more copies. We filtered duplications for those that contained a full gene and found six in cases and one in a control. This is nearly significant (1-sided Fisher's exact p-value = 0.1, Odds Ratio = 5.3 and 95% C.I. = 0.75-Infinity)(Figures 2.2.a-f). Two of these duplications reside in the same case individual.

Gene Set Enrichment and Gene Ontology Analyses

Previous reports suggest that genetic variation in cilia genes play a role in AVSD in DS (Ripoll et al., 2012; Ramachandran et al., 2015; Burnicka-Turek et al., 2016). There are 19 genes on chromosome 21 implicated as part of the ciliome. We performed Gene Set Enrichment Analysis (GSEA) with PLINK v1.07 (Purcell et al., 2007) to test two hypotheses: 1) chromosome 21 deletions are more likely to intersect cilia genes than other genes in DS+NH controls and 2) chromosome 21 duplications are more likely to intersect cilia genes than other genes in DS+AVSD cases. Within Caucasians, two putative cilia genes, *DYRK1A* and *PDXK,* are intersected by deletions in controls; no cilia genes are intersected by deletions in cases (Table 2.5). While the counts are low, this finding is suggestive (p-value = 0.1) by GSEA. Similarly, for duplications, we observed

the inverse: duplications intersect with two cilia genes, *USP25* and *ITSN1*, in cases and none in controls (p-value = 0.2). When we combine deletions and duplications that disrupt exons and thus reduce that gene to disomy, we observe 1 case and 2 controls with a CNV disrupting a cilia gene exon, which is not significant (p-value = 0.25).

To uncover novel pathways disrupted by CNVs in DS-associated AVSD, we performed a Gene Ontology analysis with the ClueGO v2.2.5 (Bindea et al., 2009) plugin in Cytoscape v3.3.0 (Shannon et al., 2003), providing lists of genes that were intersected by deletions and duplications only in cases or only in controls (Tables 2.6, 2.7 and 2.8). As we assume the same pathways leading to AVSD will be perturbed in all humans, we combined lists of genes from African Americans and Caucasians, creating eight gene lists: 1. Genes intersected by deletions only in cases, or 2. Only in controls, 3. Genes intersected by duplications only in cases, or 4. Only in controls, 5. Genes with an exon intersected by a deletion or non-gene-enveloping duplication only in cases, or 6. Only in controls, and 7. Genes completely duplicated only in cases or 8. Only in controls. In the deletion gene lists, only those within DS+NH controls clustered into a pathway. Deleted genes in controls were overrepresented in protein heterotrimerization (GO:0070208, p-value = 0.0002). In the duplication gene lists, significant pathway enrichment was only found in DS+AVSD cases. These duplication-intersected genes were significantly enriched for synaptic vesicle endocytosis (GO:0048488, p-value = 0.0001). While genes with exons disrupted by CNVs in cases showed no enrichment in biological pathways, those in controls were enriched in protein heterotrimerization (GO:0070208, p-value = 0.0047). The was only one gene and one a non-coding RNA completely duplicated in

controls, while there were 19 genes duplicated in cases and they were enriched in the process of histone methylation (GO:0016571, p-value = 0.0017).

Replication of Previous Findings

We next sought to replicate two loci previously reported to be statistically significantly associated with AVSD in a collection of individuals with DS (Sailani et al., 2013). CNV1 at chr21:43,193,374-43,198,244 (hg19) was observed as a deletion in 18% of the 55 cases with DS+AVSD and 0% of the 53 DS+NH controls and as a duplication in 7% of cases and 0% in controls. CNV2 at chr21:43,411,411-43,413,231 was found as a deletion in 24% of controls versus 0% in cases and as a duplication in 14% of cases versus 11% of controls (Table 2.9). In an internal replication study based on 49 DS+AVSD cases and 45 DS+NH controls, Sailani et al. used NanoString nCounter technology and found significant differences in copy number ratios in probes targeting these loci.

Our aCGH experiments had 19 probes within CNV1 and did not detect any CNV, though our sample size is four times larger than that of Sailani et al. (Table 2.9). Our custom array only had three probes in the CNV2 locus, and thus we were not able to detect it with our stringent criteria that required 6 or more probes to call a CNV. We performed a Nanostring experiment on a subset of our sample (49 cases and 45 controls) using the same codeset probes that Sailani et al. found to be significantly associated with DS+AVSD. We followed the same methodology as Sailani et al. to analyze the Nanostring data. For each probeset, a ratio of the probe's copy number count from a test individual over that of a reference DS sample was computed. A Mann-Whitney U-test

was then applied at each probe, testing for a difference in the mean copy number count ratios between cases and controls. In CNV1, one of the three probes found significant in the Sailani et al. sample showed a significant difference in nCounter copy number ratio between our cases and controls (p-value = 0.007) (Table 2.10). At CNV2, two of the six probes found significant in Sailani et al. were marginally significant in our dataset (p-values = 0.054 and 0.056).

The mixed results within Nanostring and aCGH experiments led us to assess the validity of these findings with a third technology (Figure 2.3). Two TaqMan probesets were selected within each CNV and tested in 46 DS+AVSD cases and 46 DS+NH controls, including the same cases and controls analyzed with Nanostring. No deletions were identified by either probeset in CNV1 or CNV2 (Table 2.11). At CNV1, a duplication was detected in one individual (A DS+NH control) by one probeset; the other did not detect a copy number change. At CNV2, a duplication was detected in one individual (a DS+NH control) by both probesets. No DS+AVSD cases had copy number calls at either CNV.

DISCUSSION

Our cohort of individuals with DS with complete AVSD and those with structurally normal hearts represents the largest study of its kind to date. In addition, our CNV dataset was built applying conservative quality control metrics on probe, array, and sample inclusion, providing robust conclusions after analysis. The composite set of CNVs, which required concordance between two well-established CNV calling algorithms, generated a

dataset with a low likelihood of false positive findings as indicated by their high representation in the DGV (94% of deletions and 80% of duplications).

Given our large sample size for this relatively rare condition, we had 80% power to detect a single CNV at 5% population frequency with an odds ratio of 2.2 or greater in Caucasians and 6.3 in African Americans (alpha = 0.05). We did not detect any single CNV with an effect size of this magnitude. Our data suggest that it is unlikely for a single common variant on chromosome 21 to explain the 2,000-fold increased risk for AVSD on a trisomy 21 background. This is consistent with our previous findings (Ramachandran et al., 2015; Ramachandran et al., 2015).

We found a suggestive association of perturbations in genes required for proper cilia functioning, consistent with previous findings. Multiple lines of evidence now provide support for the hypothesis that mutations in ciliome genes increase the risk of CHD. In a forward-genetic mouse screen, 87,355 fetuses from mutagenized mice resulted in 218 mice with CHDs (Li et al., 2015). Exome sequencing of 113 of these mice revealed 91 recessive mutations in 61 genes, of which 34 were involved in cilia functioning. Disruption of cilia-related genes has been shown to be associated with AVSD in lymphoblastoid cell lines (LCLs) from individuals with Down syndrome (Ripoll et al., 2012). Comparing gene expression profiles of LCLs from individuals with DS without CHDs to those with atrioseptal defects (ASD), ventricular septal defects (VSD), or AVSD, they found significant deregulation of cilia genes within the AVSD group. Furthermore, principal component analysis of the expression profiles separate individuals with AVSDs from those with ASDs or VSDs, pointing to a different etiology of disease progression and substantiating the need to study phenotypically distinct CHDs

independently. Most recently, Burnicka-Turek et al. characterized independent mouse lines with induced deleterious non-synonymous mutations induced in the cilia genes *Dnah11* and *Mks1* that caused AVSD (Burnicka-Turek et al., 2016).

We were unable to replicate two previously reported common copy number variants on chromosome 21 associated with DS+AVSD in a smaller cohort by Sailani et al. (2013). Although our custom array did not have enough probes to reliably detect CNV2, we were powered to detect CNV1 and did not find this CNV in either cases or controls in our larger population. Sailani et al. replicated their finding in an independent sample by reporting differences in means of Nanostring nCounter® probe ratios between cases and controls. At CNV1, we tested three of these significant probesets with available coordinates and found only one to be significant. At CNV2, we tested the six probes previously found significant and found two of the six to be marginally significant. Sailani et al. calculated ratios of probe counts of the test sample over that of the reference sample and tested for ratio differences between cases and controls using a 1-sided Mann-Whitney U-test. They did not report actual frequencies of deletions and duplications called by Nanostring. We applied their techniques to our larger cohort and found inconclusive results to support their findings. As a final validation of these proposed DS+AVSD-associated CNVs, we performed TaqMan™ Copy Number assays with two probe sets for each CNV. As detailed in RESULTS, no deletions were detected at CNV1 or CNV2. For a single control, duplications were detected at both probes in CNV2 and at one of two probes in CNV1. Thus, in our cohort that was four-times larger than that of Sailani et al., we failed to replicate their reported findings and conclude that the associations to DS+AVSD of CNV1 and CNV2 are likely false positives.

Our study stands in agreement with the consensus of other studies reporting complex heterogeneity of atrioventricular septum and valve development in both the disomic population and in individuals with trisomy 21 (Robinson et al., 2003; Ackerman et al., 2012; Al Turki et al., 2014; Ramachandran et al., 2015; Ramachandran et al., 2015; Priest et al., 2016). Our data support the nuanced hypotheses that deletions on chromosome 21 on a trisomic background reduce the risk for AVSD and duplications on chromosome 21 further increase risk of AVSD in DS. These effects were enriched when considering rare (< 0.01) deletions and duplications. Rare deletions have been previously implicated within our DS cohort, where DS+AVSD cases were found to have a greater genome-wide burden of rare, large (> 100kb) deletions (Ramachandran et al., 2015; Ramachandran et al., 2015)

Moving forward, genetic studies of CHD in DS, as well as nonsyndromic CHDs, should be designed with this considerable genetic heterogeneity in mind. It is clear that, while trisomy 21 alone increases the risk for AVSD 2,000 fold, its probable mode of action is through epistatic interactions among many genes, at least some of which are necessary for the structure and function of cilia. Untangling these complex risk factors will require a larger cohort of individuals with DS with and without CHDs to find susceptibility loci of measurable effect. As these cohorts continue to grow, efforts should focus on exome and whole genome sequencing approaches that identify rare variants, whose effects can be tested for burdening candidate genetic pathways of cardiogenesis. Finally, environmental factors need greater consideration and resources should be prioritized to gather broad epidemiological data and link them to genomic resources.

METHODS

DNA samples

Participant samples were collected as described previously (Freeman et al., 1998; Freeman et al., 2008; Locke et al., 2010; Ramachandran et al., 2015; Ramachandran et al., 2015). Individuals diagnosed with full or translocation trisomy 21, documented by karyotype, were recruited from centers across the United States. Institutional review boards at each enrolling institution approved protocols and informed consent was obtained from a custodial parent for each participant. A single cardiologist (K. Dooley) identified cases from medical records as individuals with a complete, balanced AVSD diagnosed by echocardiogram or surgical reports (DS+AVSD). Controls were classified as individuals with a structurally normal heart, patent foramen ovale, or patent ductus arteriosus (DS+NH).

Genomic DNA was extracted from lymphoblastoid cell lines with the Puregene DNA purification kit by manufacturer's protocol (Qiagen, Valencia, CA). DNA quantity and quality were checked on a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE) and assessed for integrity on 0.8% agarose gels stained with ethidium bromide.

Microarray Design and Processing

All analyses used the human genome reference hg19 build. We designed a custom 8x60k Agilent (Agilent Technologies, Santa Clara, CA) CGH array using eArray (https://earray.chem.agilent.com accessed April, 2011). The array consisted of 52,944 60-mer DNA probes targeting human chromosome 21, providing a mean spacing of 673 bp and a median spacing of 448 bp, as well as a genomic backbone of probes and Agilent's control probes (design file: ADM2Chr21_60k_final_033839_D_F_20120731.xml).

Array hybridization was processed according to Agilent's protocol and scanned on an Agilent Surescan High-Resolution Microarray Scanner at Emory University. A single female (SLS ID: 7210100.00) with trisomy 21 and no CHD was used as the reference sample for all test individuals. This individual had a known deletion at chr21:45,555,257-45,615,042, which would be detected as a duplication in all test samples.

Sample Quality Control

We performed aCGH on a total of 551 DS samples. We preformed three stages of sample/array quality control (QC). We first performed Agilent's recommended QC. Their recommended QC cutoff for arrays is a Derivative of Log_2 Ratio (DLR) < 0.3. DLR is a measure of probe-to-probe noise and is the standard deviation of adjacent probe's log_2 differences. Twenty-six samples failed to meet this threshold and were excluded (Figure 2.4).

Second, while the remaining 525 microarrays met Agilent's basic QC parameter of DLR < 0.3, visual inspection of log_2 plots revealed a number of arrays with an

increased probe variance. To quantitatively assess and account for this effect, we calculated the variances of intra-array probe $\log_2$ ratio to develop a conservative array inclusion criterion. We excluded 74 arrays with variance ≥ 1 standard deviation (SD) over the mean from any further analysis (Figure 2.5).

Third, to avoid biasing an individual microarray toward over or under calling gains or losses, it is important that the mean $\log_2$ ratio across the array is near the expected value of zero. The means of the intra-array probe $\log_2$ were calculated on the 451 remaining arrays (grand mean = -0.00045) and 25 arrays with individual means outside of 2 SD from the group mean were removed (Figure 2.6). After CNV detection (described below), we removed clear outlier samples that had the number of CNVs (deletions or duplications) called > 5 SD over the mean. This removed five samples.

To avoid spurious association results based on population stratification we performed principle component analysis on the majority of our samples that had genome-wide SNP data available from our previously published study (Ramachandran et al., 2015). Four samples without genotyping data were removed from further CNV analyses. In PLINK (version 1.9; Chang et al., 2015) SNPs were removed that had > 10% missingness or failed the Hardy-Weinberg equilibrium exact test with a p-value < $1x10^{-6}$. Common SNPs with minor allele frequency > 0.05 were pruned by PLINK's "--indep-pairwise" command within 50kb windows, a 5 SNP step, and an r^2 threshold of 0.2 leaving 552,943 SNPs. The first five eigenvectors were calculated using the R package SNPRelate (Zheng et al., 2012) and plotted (Figures 2.7a-e). PCA round 1 clearly separates self-identified African Americans from Caucasians. Round 2 was performed separately on African Americans and Caucasians. Six African Americans were visibly

clear outliers (Round 2 PC1 ≤ -0.127) and were removed from further analyses. Five Caucasians were clear outliers (Round 2 PC1 ≤ -0.1) and were also removed. The final cohort contained 198 cases and 211 controls (Table 2.12).

CNV Calling

We also evaluated the quality of data at the probe level. Because custom CGH arrays contain probes with unpredictable binding characteristics, the variances of normalized probe fluorescent signals were calculated and 2,193 probes with inter-array variance ≥ 1 SD above the mean were removed (Figure 2.8). These calculations were done on the full set of arrays passing the above DLR criteria and before the above intra- and inter-array probe log2 variance calculations and filtering.

We used two algorithms, ADM2 and GADA to identify putative CNVs (Pique-Regi, R et al., 2010). We required that CNVs be called by both algorithms to be included in the analysis. Parameters for Agilent's ADM2 algorithm were set within their Genomic Workbench software (version 7.0.4.0) as follows: ≥ 6 probes, average $\log_2$ shift of ± 0.2, use of the diploid peak centralization, 2 kb window GC correction, intra-array replicates combined, and Fuzzy Zero applied. GADA adjustable parameters are the minimum probe number for a CNV to be called (MinSegLen) and a threshold T_m referring to the minimum t-statistic that a predicted breakpoint must reach during its Backward elimination procedure. We empirically optimized the GADA T_m variable across a range of 4.5 to 20.5, by half steps, and evaluated performance based on two criteria: 1. Did the algorithm detect duplications in at least 80% of our test samples at our known reference

deletion, and 2. Did the algorithm detect common deletions found in 1000 Genomes' Phase 3 release of structural variants at a similar population frequency.

Compressed .vcf data and accompanying .tbi file for chromosome 21 produced from whole genome sequencing by the 1000 Genomes Consortium was downloaded from http://hgdownload.cse.ucsc.edu/gbdb/hg19/1000Genomes/phase3/ on Feb. 27th, 2016. Tabix commands created unzipped .vcf files covering chr21:13000000-47000000 and variants denoted 'SVTYPE' were filtered with shell commands and variants denoted as deletion, duplication, or both (multi-allelic) were analyzed. The lower limit of CGH detection was set at ≥ 1798 bp and variants less than 1798 bp were removed from the 1000 Genomes' comparison set.

Of seven common copy number variants on chromosome 21, our CGH array had at least six probes in two of these variants: esv3646598 and esv3646663. esv3646598 has a frequency of 0.064 in European Ancestry individuals (0.004 in African Ancestry). esv3646663 has a frequency of 0.227 in African Ancestry individuals (0.001 in Europeans). We do not call absolute copy number from CGH data and a deletion can represent zero, one, or two copies of three expected. Thus, an upper frequency bound was set as Freq = (3(d)+0(N-d)) / 3(N), while the lower bound was set as Freq = (1(d)+0(N-d)) / 3(N), where d equals the number of times the deletion was called and N equals the total number of chromosomes. For both variants, in their respective ancestral population, T_m=8 detects common structural variants within the expected range and also maximizes the detection of our reference deletion (Figures 2.9.a and 2.9.b). GADA was then launched using a custom R script applying the following parameters: estim.sigma2=TRUE, MinSegLen=6 and T_m=0.8

CNVs ≥ 1 Mb were removed (14 deletions; 7 duplications) after visually checking $\log_2$ plots to confirm these were likely false positives. Variants with breakpoints inside our reference deletion (chr21:45555257-45615042) were removed (0 deletions; 354 duplications). The p-arm and pericentromeric region of chromosome 21 are poorly mapped and variants with breakpoints inside chr21:0-15400000 were removed (2 deletions; 1 duplication). Clear outliers containing large number of deletions or duplications were removed. We used a threshold of >5 SD over the mean of 0.73 deletions and 0.15 duplications calculated among the 426 arrays. Five SDs over the mean corresponded to more than five deletions or two duplications within one array. These five samples contained 81 deletions and 6 duplications. The final dataset includes 215 deletions and 59 duplications, of which 92% of deletions and 73% of duplications have 50% reciprocal overlap with variants in the Database of Genomic Variants (http://dgv.tcag.ca), indicating a low rate of false positives (Table 2.13).

Replication of Findings in Sailani et al. (2013)

Two common CNVs were found to be associated with DS+AVSD in the study by Sailani et al. (2013). To try to replicate that finding, we used identical Nanostring probes in 48 cases and 48 controls from our DS cohort. We included probes that showed significant copy number differences between their cases and controls totaling three of the seven probes for CNV1 and six of the seven probes for CNV2 (Table 2.14). Samples were processed by the Gene Expression Analysis Laboratory at The University of Tennessee. Additionally, two TaqMan® (Applied Biosystems, Grand Island, NY) assays targeting

each locus were selected for CNV1 and CNV2 (Table 2.14). These assays were performed by the Emory Integrated Genomics Core on the same 96 samples tested by Nanostring. Copy number calls were made by TaqMan's CopyCaller software, and calls with a confidence probability less than 0.8 were dropped.

CNV Association and Burden Analyses

We used PLINK v.1.07 to carry out association and burden analyses separately for deletions and duplications. To explicitly test the hypothesis that a gene reduced to two functional copies provides protection to AVSD in DS, we combined deletions that intersect an exon (refGene-hg19 updated July 3, 2016) with duplications that have predicted breakpoints within an exon to form a "reduced to disomy" set of CNVs. We also explicitly tested the inverse hypothesis, that genes entirely duplicated increase the risk for AVSD in DS. Three testing paradigms were performed: 1) burden analyses using the --cnv-indiv-perm and --cnv-count commands, 2) associations with individual CNV regions using --cnv-count, and 3) associations with individual genes overlapped by a CNV --cnv-intersect and --cnv-test-region. Empirical p-values of significance were determined by performing one million permutations for each test. These p-values are 1-sided and we tested for excess burden of duplications in cases and for deletions in controls. These three testing paradigms were applied to the full dataset, as well as subsets of CNVs filtered by overlap in the Database of Genomic Variants (downloaded January, 2016), and by CNV frequency of greater than or less than 1%. Burden analysis in PLINK tests for differences between cases and controls using three different approaches: 1) Is

there a difference between the average number of CNVs per person (RATE)? 2) Is there a difference in the average number of bases covered by all CNVs (KBTOT)? and 3) Is there a difference in the average number of genes intersected by CNVs per person (GRATE)? We performed burden tests across deletions and duplications on chromosome 21 as entire sets and filtered by the allele frequency of the CNV (common or rare < 0.01) and by their existence in the Database of Genomic Variants or not.

Gene Set Enrichment and Gene Ontology Term Analyses

We used PLINK v1.07 to perform Gene Set Enrichment Analysis (GSEA) (Raychaudhuri et al., 2010) on a set of cilia-related genes (Table 2.15) compiled by Inglis et al., 2006 and previously implicated in Down syndrome associated AVSD (Ripoll et al., 2012; Ramachandran et al., 2014; Li et al., 2015). GSEA tests the hypothesis that cilia genes are enriched for CNVs compared to all chromosome 21 genic CNVs and assigns a 1-sided empirical p-value by one million permutations indicating positive enrichment in cases for duplications or in controls for deletions.

We performed a Gene Ontology (GO) term analysis using the ClueGO Extension in Cytoscape v.3.3.0 (Bindea et al., 2009). Gene lists were created based on those intersected by deletions or duplications only in cases or only in controls. Each gene set was analyzed with the following settings: Ontologies=Go-Biological Process downloaded May 5th, 2016, Evidence=All, Pathway Significance=0.05, GO Tree Interval=3 minimum level and 13 maximum level, 2 gene minimum or 1% of pathway genes, and Bonferroni

step down p-value correction with mid-P-values. Other parameters were left at default settings.

References

Ackerman, C., Locke, A. E., Feingold, E., Reshey, B., Espana, K., Thusberg, J. et al. (2012). An Excess of Deleterious Variants in VEGF-A Pathway Genes in Down-Syndrome-Associated Atrioventricular Septal Defects. *American journal of human genetics*, *91*(4), 646-659.

Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M.-P., Lindsay, S. et al. (2014). Rare Variants in NR2F2 Cause Congenital Heart Defects in Humans. *American journal of human genetics*, *94*(4), 574-585.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A. et al. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091-1093.

Burnicka-Turek, O., Steimle, J. D., Huang, W., Felker, L., Kamp, A., Kweon, J. et al. (2016). Cilia gene mutations cause atrioventricular septal defects by multiple mechanisms. *Human Molecular Genetics*.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1-16.

Freeman, S. B., Bean, L. H., Allen, E. G., Tinker, S. W., Locke, A. E., Druschel, C. et al. (2008). Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. *Genetics in medicine : official journal of the American College of Medical Genetics*, *10*(3), 173-180.

Freeman, S. B., Taft, L. F., Dooley, K. J., Allran, K., Sherman, S. L., Hassold, T. J. et al. (1998). Population-based study of congenital heart defects in Down syndrome
. *Am J Med Genet*, *80*(3), 213-217.

Inglis, P. N., Boroevich, K. A., & Leroux, M. R. (2006). Piecing together a ciliome. *Trends in Genetics*, *22*(9), 491-500.

Kucik, J. E., Shin, M., Siffel, C., Marengo, L., & Correa, A. (2013). Trends in survival among children with Down syndrome in 10 regions of the United States. *Pediatrics*, *131*(1), e27-36.

Le Gloan, L., Mercier, L.-A., Dore, A., Marcotte, F., Ibrahim, R., Mongeon, F.-P. et al. (2011). Recent Advances in Adult Congenital Heart Disease. *Circulation Journal*, *75*(10), 2287-2295.

Li, Y., Klena, N. T., Gabriel, G. C., Liu, X., Kim, A. J., Lemke, K. et al. (2015). Global genetic analysis in mice unveils central role for cilia in congenital heart disease. *Nature*, *521*, 520-524.

Locke, A. E., Dooley, K. J., Tinker, S. W., Cheong, S. Y., Feingold, E., Allen, E. G. et al. (2010). Variation in folate pathway genes contributes to risk of congenital heart defects among individuals with Down syndrome. *Genetic epidemiology*, *34*(6), 613-623.

MacDonald, J. R., Ziman, R., Yuen, R. K. C., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Research*, *42*(D1), D986-D992.

Pique-Regi, R., Caceres, A., & Gonzalez, J. R. (2010). R-Gada: a fast and flexible pipeline for copy number analysis in association studies. *BMC Bioinformatics*, *11*(1), 1-12.

Priest, J. R., Girirajan, S., Vu, T. H., Olson, A., Eichler, E. E., & Portman, M. A. (2012). Rare copy number variants in isolated sporadic and syndromic atrioventricular septal defects. *American journal of medical genetics. Part A*, *158A*(6), 1279-1284.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D. et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.

Ramachandran, D., Mulle, J. G., Locke, A. E., Bean, L. J. H., Rosser, T. C., Bose, P. et al. (2014). Contribution of copy-number variation to Down syndrome-associated atrioventricular septal defects. *Genetics in medicine : official journal of the American College of Medical Genetics*, October), 1-7.

Ramachandran, D., Zeng, Z., Locke, A. E., Mulle, J. G., Bean, L. J. H., Rosser, T. C. et al. (2015). Genome-Wide Association Study of Down Syndrome-Associated Atrioventricular Septal Defects. *G3: Genes$_{Genomes}$Genetics*, *5*(10), 1961-1971.

Raychaudhuri, S., Korn, J. M., McCarroll, S. A., Altshuler, D., Sklar, P., Purcell, S. et al. (2010). Accurately assessing the risk of schizophrenia conferred by rare copy-number variation affecting genes with brain function. *PLoS genetics*, *6*(9).

Reller, M. D., Strickland, M. J., Riehle-Colarusso, T., Mahle, W. T., & Correa, A. (2008). Prevalence of congenital heart defects in metropolitan Atlanta, 1998-2005. *The Journal of pediatrics*, *153*(6), 807-813.

Ripoll, C., Rivals, I., Ait Yahya-Graison, E., Dauphinot, L., Paly, E., Mircher, C. et al. (2012). Molecular signatures of cardiac defects in down syndrome lymphoblastoid cell lines suggest altered ciliome and hedgehog pathways. *PloS one*, *7*(8), e41616.

Robinson, S. W., Morris, C. D., Goldmuntz, E., Reller, M. D., Jones, M. A., Steiner, R. D. et al. (2003). Missense mutations in CRELD1 are associated with cardiac atrioventricular septal defects. *American journal of human genetics*, *72*(4), 1047-1052.

Sailani, M. R., Makrythanasis, P., Valsesia, A., Santoni, F. A., Deutsch, S., Popadin, K. et al. (2013). The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome. *Genome research*.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D. et al. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498-2504.

Yang, Q., Rasmussen, S. A., & Friedman, J. (2002). Mortality associated with Down's syndrome in the USA from 1983 to 1997: a population-based study. *The Lancet*, *359*(9311), 1019-1025.

Zarrei, M., MacDonald, J. R., Merico, D., & Scherer, S. W. (2015). A copy number variation map of the human genome. *Nat Rev Genet*, *16*(3), 172-183.

		# Subjects (Male/ Female)	Type	# Variants	# Variants Disrupting Exons	Average per Person	Size Range (kb)	Median Size (kb)
Caucasians	**Cases DS+AVSD**	77 / 97	Deletions	66	19	0.38	1.8 - 159.2	10.7
			Duplications	31	22	0.18	11.6 - 395.1	23.1
	Controls DS+NH	104 / 77	Deletions	71	21	0.39	1.8 - 239.1	10.7
			Duplications	23	20	0.13	10.4 - 200.0	18.0
African Americans	**Cases DS+AVSD**	6 / 18	Deletions	36	0	1.50	2.1 - 44.1	4.4
			Duplications	1	0	0.04	491.9	491.9
	Controls DS+NH	18 / 12	Deletions	42	1	1.40	2.1 - 260.3	7.2
			Duplications	4	1	0.13	4.5 - 14.4	10.9

Table 2.1: CNV summary statistics for cases and controls stratified by race/ethnicity. Cases (DS+AVSD) are defined as those with Down syndrome and complete atrioventricular septal defect. Controls (DS+NH) are individuals with Down syndrome without a congenital heart defect

		Caucasian			African American			
Filtering	**Burden Test**	**DS+AVSD Cases (n=174)**	**DS+NH Controls (n=181)**	**Empirical permuted p-value**	**DS+AVSD Cases (n=24)**	**DS+NH Controls (n=30)**	**Empirical permuted p-value**	**Combined p-value**
All	**Number of deletions**	66	71	na	36	42	na	
	Average number of deletions	0.38	0.39	0.46	1.50	1.40	0.67	0.67
	Average kb covered by deletions	31.24	37.95	0.23	13.06	33.45	**0.04**	0.05
	Avg. number of chr21 genes intersected by deletions	0.13	0.14	0.44	0.13	0.30	0.42	0.49
Common (≥0.01)	**Number of deletions**	45	46	na	19	18	na	
	Average number of deletions	0.26	0.25	0.58	0.79	0.60	0.88	0.86
	Average kb covered by deletions	29.52	28.04	0.60	5.57	6.94	0.09	0.22
	Avg. number of chr21 genes intersected by deletions	0.09	0.07	0.76	0.00	0.00	1.00	0.97
Rare (<0.01)	**Number of deletions**	21	25	na	17	24	na	
	Average number of deletions	0.12	0.14	0.39	0.71	0.80	0.42	0.46
	Average kb covered by deletions	26.01	42.41	0.21	12.80	45.63	**0.02**	0.03
	Avg. number of chr21 genes intersected by deletions	0.04	0.07	0.23	0.13	0.30	0.42	0.32
Registered in DGV	**Number of deletions**	61	62	na	36	40	na	
	Average number of deletions	0.35	0.34	0.59	1.50	1.33	0.75	0.80
	Average kb covered by deletions	26.82	26.92	0.49	13.06	21.41	0.09	0.18
	Avg. number of chr21 genes intersected by deletions	0.12	0.10	0.78	0.13	0.10	0.77	0.91
Not	**Number of deletions**	5	9	na	0	2	na	

registered in DGV	**Average number of deletions**	0.03	0.05	0.26	0.00	0.07	0.30	0.28
	Average kb covered by deletions	69.16	95.09	0.32	0.00	132.50	0.56	0.49
	Avg. number of chr21 genes intersected by deletions	0.01	0.04	0.08	0.00	0.20	0.56	0.19

Table 2.2: Results of deletion burden tests in PLINK. In Caucasians, no increased burden of average deletion number, average number of bases covered by deletions, or percentage of chromosome 21 genes intersected by deletions was seen. In African Americans, there is a significant increase in bases covered by deletions in DS+NH controls and this effect is driven primarily by rare deletions.

		Caucasian			**African American**			
Filtering	**Burden Test**	**DS+AVSD Cases (n=174)**	**DS+NH Controls (n=181)**	**Empirical permuted p-value**	**DS+AVSD Cases (n=24)**	**DS+NH Controls (n=30)**	**Empirical permuted p-value**	**Combined p-value**
Chr21	**Number of duplications**	31	23		1	4		
	Average number of duplications	0.18	0.13	0.15	0.04	0.13	0.91	0.40
	Average kb covered by duplications	83.53	40.49	0.09	491.90	13.56	0.08	0.04
	Avg. number of chr21 genes intersected by duplications	0.22	0.10	0.07	0.33	0.03	0.44	0.13
Common	**Number of duplications**	16	17		0	0		
	Average number of duplications	0.09	0.09	0.60	0.00	0.00	1.00	0.90
	Average kb covered by duplications	17.67	17.41	0.43	0.00	0.00	1.00	0.79
	Avg. number of chr21 genes intersected by duplications	0.06	0.07	0.71	0.00	0.00	1.00	0.95
Rare	**Number of duplications**	15	6		1	4		

	Average number of duplications	0.09	0.03	0.06	0.04	0.13	0.91	0.20
	Average kb covered by duplications	179.30	92.39	0.13	491.90	13.56	0.08	0.06
	Avg. number of chr21 genes intersected by duplications	0.16	0.03	**0.04**	0.33	0.03	0.45	0.09
DGV	**Number of duplications**	20	20		0	2		
	Average number of duplications	0.11	0.11	0.51	0.00	0.07	1.00	0.86
	Average kb covered by duplications	28.50	24.63	0.34	0.00	23.62	1.00	0.70
	Avg. number of chr21 genes intersected by duplications	0.10	0.09	0.47	0.00	0.03	1.00	0.82
NoDGV	**Number of duplications**	11	3		1	2		
	Average number of duplications	0.06	0.02	**0.04**	0.04	0.07	0.84	0.15
	Average kb covered by duplications	187.20	119.20	0.26	491.90	8.54	0.14	0.16
	Avg. number of chr21 genes intersected by duplications	0.11	0.01	**0.03**	0.33	0.00	0.44	0.08

Table 2.3: Results of duplications burden tests in PLINK. In Caucasians, DS+AVSD cases had significantly more genes intersected by rare duplications than did DS+NH controls and DS+AVSD cases had a greater number of rare duplications on average.

		Caucasian		
Filtering	Burden Test	DS+AVSD Cases (n=174)	DS+NH Controls (n=181)	Empirical permuted p-value
All	Number of CNVs	41	41	-
	Average number of CNVs	0.24	0.23	0.62
	Average kb covered by CNVs	50.45	46.4	0.65
	Avg. number of chr21 genes intersected by CNVs	0.20	0.19	0.62
Common (≥0.01)	Number of CNVs	31	30	-
	Average number of CNVs	0.18	0.17	0.67
	Average kb covered by CNVs	40.14	38.88	058
	Avg. number of chr21 genes intersected by CNVs	0.14	0.14	0.62
Rare (<0.01)	Number of CNVs	10	11	-
	Average number of CNVs	0.06	0.06	0.53
	Average kb covered by CNVs	77.31	90.28	0.38
	Avg. number of chr21 genes intersected by CNVs	0.05	0.05	0.61
Registered in DGV	Number of CNVs	37	35	-
	Average number of CNVs	0.21	0.19	0.71
	Average kb covered by CNVs	42.25	39.73	0.52
	Avg. number of chr21 genes intersected by CNVs	0.17	0.16	0.67
Not registered in DGV	Number of CNVs	4	6	-
	Average number of CNVs	0.02	0.03	0.41
	Average kb covered by CNVs	122.3	111.6	0.58

	Avg. number of chr21 genes intersected by CNVs	0.02	0.03	0.52

Table 2.4: Results of CNVs disrupting exons burden tests. No association was seen with CNVs disrupting genes and returning them to disomy.

	Number of Cilia Genes Intersected		
	Cases DS+AVSD	Controls DS+NH	p-value
Deletions	0	2	0.1
Duplications	2	0	0.2
Exon-disrupting CNV	1	2	0.25

Table 2.5: Counts of cilia genes on chromosome 21 intersected by CNVs. Gene Set Enrichment Analysis (GSEA) tests for enrichment of CNVs in 19 cilia genes compared to all other genic CNVs. Two cilia genes are intersected by deletions in controls, which suggests an enrichment from a GSEA permuted p-value of 0.1. We find the opposite effect in duplications where 2 genes are intersected only in controls, which while being in the hypothesized direction, is not statistically significant by GSEA. Considering that duplications with breakpoints within an exon likely have the same effect as a deletion and combining these duplications with exon-interrupting deletions, we no longer find a significant difference of cilia genes being affected by CNVs between cases and controls.

African Americans				Caucasians			
Deletions		Duplications		Deletions		Duplications	
Case Only	Control Only	Case Only	Control Only	Case Only	Control Only	Case Only	Control Only
LOC339622	C21orf56	ATP5J	SIK1	AGPAT3	CHODL	ADAMTS5	CBR3
	COL6A1	GABPA		MIR802	DYRK1A	C21orf49	CBR3-AS1
	COL6A2	JAM2		PDE9A	LINC00308	C21orf62	DOPEY2
	FTCD	LINC00158		SAMSN1	LINC00478	CRYZL1	LOC100133286
	HLCS	LINC00515		TRPM2	LINC00515	DONSON	SCAF4
	LSS	MIR155			PCP4	DSCAM	
	PCBP3	MIR155HG			PDXK	DSCAM-AS1	
		MRPL39			PKNOX1	GCFC1	
					TCP10L	GCFC1-AS1	
						ITSN1	
						KRTAP21-1	
						KRTAP21-2	
						KRTAP21-3	
						KRTAP7-1	
						KRTAP8-1	
						NCAM2	
						PRMT2	
						S100B	
						SON	
						SYNJ1	
						USP25	
						ZNF295	

Table 2.6: Genes provided to ClueGO for GO term pathway analysis. In each list, genes intersected by the respective CNV type only in DS+AVSD cases or only in DS+NH controls were submitted for pathway analysis. Respective cnv type + case/control status gene lists were combined between the two populations.

African Americans				Caucasians			
Deletions		Duplications		Deletions		Duplications	
Case Only	Control Only	Case Only	Control Only	Case Only	Control Only	Case Only	Control Only
	LOC101928796		SIK1	SAMSN1	LINC00308	USP25	SCAF4
	COL6A1		LOC102724426	MIR802	LINC00515	NCAM2	
	COL6A2			TRPM2	TCP10L	ADAMTS5	
	FTCD				DYRK1A	ZNF295	
	SPATC1L				PKNOX1		
	LSS				PDXK		

Table 2.7: Genes with exons intersected by deletions or duplications that do not envelope an entire gene. Genes across CNV type and across ethnicity were merged separately for cases and controls and provided to ClueGo for GO term pathway analysis.

African Americans		Caucasians	
Duplications		Duplications	
Case Only	Control Only	Case Only	Control Only
ATP5J		DSCAM-AS1	CBR3
GABPA		PAXBP1	CBR3-AS1
JAM2		PAXBP1-AS1	
LINC00158		PRMT2	
LINC00515		CRYZL1	
MIR155		DONSON	
MIR155HG		KRTAP21-1	
MRPL39		KRTAP21-2	
		KRTAP21-3	
		KRTAP7-1	
		KRTAP8-1	

Table 2.8: Genes completely enveloped by duplications. Genes across ethnicity were merged separately for cases and controls and provided to ClueGo for GO term pathway analysis.

		Sailani et al. (2013) CGH Results				Our CGH Results			
		Cases (n=53) DS+AVSD		Controls (n=55) DS+NH		Cases (n=198) DS+AVSD		Controls (n=222) DS+NH	
	Coordinates (hg19)	Deletion frequency	Duplication frequency	Deletion frequency	Duplication frequency	Deletion frequency	Duplication frequency	Deletion frequency	Duplication frequency
CNV 1	chr21:43,193,374 -43,198,244	0.18	0.07	0	0	0	0	0	0
CNV 2	chr21:43,411,411 -43,413,231	0.24	0.14	0	0.11	na	na	na	na

Table 2.9: Comparison of DS+AVSD significantly associated CNVs from Sailani et al. (2013) to our current study by aCGH. We did not replicate the previously reported significant association of common deletions and duplications at CNV1. Our CGH array did not have at least 6 probes inside CNV2 and thus was undetectable by our methodology.

	Coordinates (hg19)	Sailani p-value	Our Cohort p-value
CNV1			
Probe 2	chr21:43195101-43195176	0.001	0.812
Probe 3	chr21:43195664-43195743	0.017	0.007
Probe 6	chr21:43198103-43198173	0.036	0.08
CNV2			
Probe 1	chr21:43411026-43411115	0.014	0.366
Probe 2	chr21:43411401-43411473	0.018	0.577
Probe 4	chr21:43412130-43412219	0.001	0.054
Probe 5	chr21:43412564-43412653	0.004	0.387
Probe 6	chr21:43412999-43413088	0.016	0.056
Probe 7	chr21:43413251-43413340	0.008	0.783

Table 2.10: Comparison of DS+AVSD significantly associated CNVs from Sailani et al. (2013) to our current study using Nanostring technology. We performed Nanostring nCounter assays on 46 DS+AVSD cases and 45 DS+NH controls using the same probes used by Sailani et al. (2013) in their CNV replication experiment that included 49 cases and 45 controls. To maintain congruency, we applied their assessment strategy to test for mean differences in normalized count (CN) ratios using a one-sided Mann-Whitney U-test. In CNV1, we detected a significant difference in CN ratios for probe 3 (using Sailani et al. nomenclature), but did not find this relationship for the other two probes previously found significant by Sailani et al. (2013). In CNV2, two of the six previously significant probes were marginally significant in our experiment.

Probe Coordinates	Deletion frequency		Duplication frequency	
	Cases	Controls	Cases	Controls
CNV1				
P1.1: chr21:43195435-43195436	0.00	0.00	0.00	0.00
P1.2: chr21:43198105-43198106	0.00	0.00	0.00	0.02
CNV2				
P2.1: chr21:43411535-43411536	0.00	0.00	0.00	0.02
P2.2: chr21:43413074-43413075	0.00	0.00	0.00	0.02

Table 2.11: TaqMan copy number results of 46 DS+AVSD cases and 46 DS+NH controls. TaqMan copy number assays were performed targeting CNV1 and CNV2, each with two probe-sets. Copy number was estimated using CopyCaller® software. A single control sample had a duplication call at CNV2 by both probes and at CNV1 by one of the two probes. No other CNVs were called by TaqMan.

Inclusion Criteria	Cases	Controls
Arrays passed Agilent scanner QC of Derivative of Log2 Ratio < 0.3	236	283
Arrays with probe log2 variance < 1 SD over mean	205	239
Arrays with probe log2 mean inside 2 SD	198	229
Arrays with variant counts within 5 SD of mean	198	222
Array sample had Affymetrix genotyping data available	196	220
Array not population outlier based on Principle Component Analysis	196	211
FINAL | Caucasian	174	181
FINAL | African American	24	30

Table 2.12: Arrays were required to meet stringent quality control criteria to be eligible for downstream analyses. Analyzed arrays met 1. Agilent's recommended Derivative of Log_2 Ratio > 0.3; 2. Intra-array normalized log_2 ratio variance < one standard deviation above the mean; 3. Intra-array normalized log_2 ratio mean inside two standard deviations; 4. Contain fewer deletions or duplications than five standard deviations above the mean; 5. Have Affymetrix genotyping data available for Principle Component analysis (PCA) and 6. Not be population outliers identified by PCA.

Criteria	Deletions	Duplications
ADM2 called	486	479
GADA called	2,750	2,328
50% Reciprocal Overlap	323	428
Variant < 1 mb	309	421
Variant not inside positive control reference deletion	309	67
Variant on q-arm	307	66
Variant not in outlier sample	226	60
Variant not in sample without Affymetrix genotyping data	224	59
Variant not in population outler identified by Principle Component analysis	215	59
FINAL | Caucasian	137	54
FINAL | African American	78	5

Table 2.13: We followed a conservative CNV filtering paradigm to minimize false positives. Analyzed deletions and duplications were required to be called by both the ADM2 and GADA algorithms. CNVs over 1 mb were removed after visualizing log_2 plots and determining they were likely spurious calls. CNVs within our reference deletion were removed as we did not distinguish absolute copy number of gains or losses and thus could not interpret data over this known deletion in the reference. CNVs on the p-arm or pericentromeric (chr21:0-15,400,000) were removed as these regions are poorly mapped in the reference genome and are gene sparse. CNVs in clear outlier samples with more deletions or duplications than five standard deviations over the mean were dropped along with those samples. CNVs in samples without Affymetrix genotyping data were dropped as well as those CNVs residing in population outliers identified by PCA.

CNV coordinates (hg19)	Platform	Probe Coordinates
chr21:43,193,374-43,198,244	Nanostring	chr21:43195101-43195176
		chr21: 43195664-43195743
		chr21:43198103-43198173
	TaqMan®	chr21:43195435-43195436
		chr21:43198105-43198106
chr21:43,411,411-43,413,231	Nanostring	chr21:43411026-43411115
		chr21:43411401-43411473
		chr21:43412130-43412219
		chr21:43412564-43412653
		chr21:43412999-43413088
		chr21:43413251-43413340
	TaqMan®	chr21:43411535-43411536
		chr21:43413074-43413075

Table 2.14: Nanostring and TaqMan target probe coordinates for validation of Sailani associated CNVs.

Gene	Coordinates (hg19)
GART	chr21:34876237-34915198
MCM3AP	chr21:47655047-47705236
RSPH1	chr21:43892596-43916401
ABCG1	chr21:43619798-43717354
DYRK1A	chr21:38739858-38887679
PWP2	chr21:45527207-45551063
PCNT	chr21:47744035-47865682
SOD1	chr21:33031934-33041243
HSPA13	chr21:15743436-15755509
USP25	chr21:17102495-17252377
U2AF1	chr21:44513065-44527688
WDR4	chr21:44263203-44299678
CCT8	chr21:30428647-30446010
CBS	chr21:44473300-44496040
PDXK	chr21:45138977-45182188
COL18A1	chr21:46825096-46933634
ITSN1	chr21:35014783-35261609

Table 2.15: Cilia gene list used for Gene Set Enrichment Analysis. Assembled from Ripoll et al., 2012.

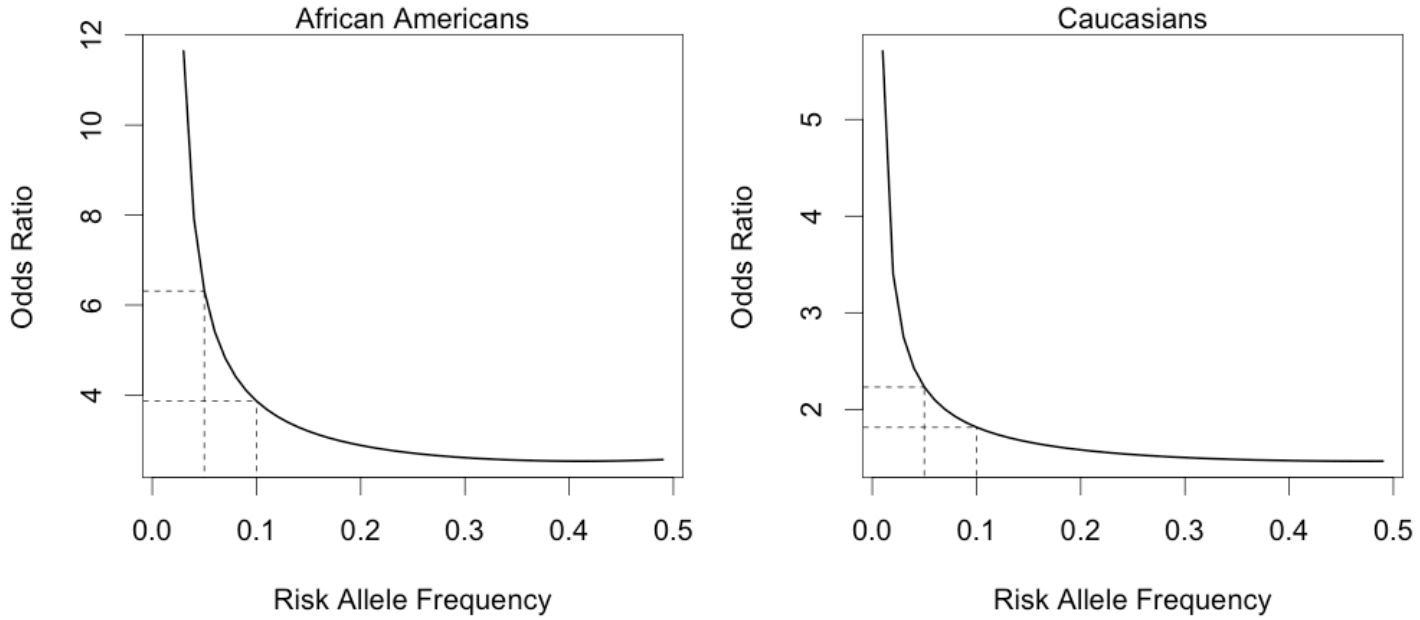


Figure 2.1: Power calculations assuming a nominal p-value = 0.05, additive model, and disease frequency of 0.18. We had 80% power to detect a risk variant with allele frequency = 0.05 with odds ratio (OR) of 6.3 and with an OR of 3.9 at allele frequency = 0.1. In Caucasian, we had 80% power to detect ORs of 2.2 and 1.8 for risk allele frequencies of 0.05 and 0.1 respectively.

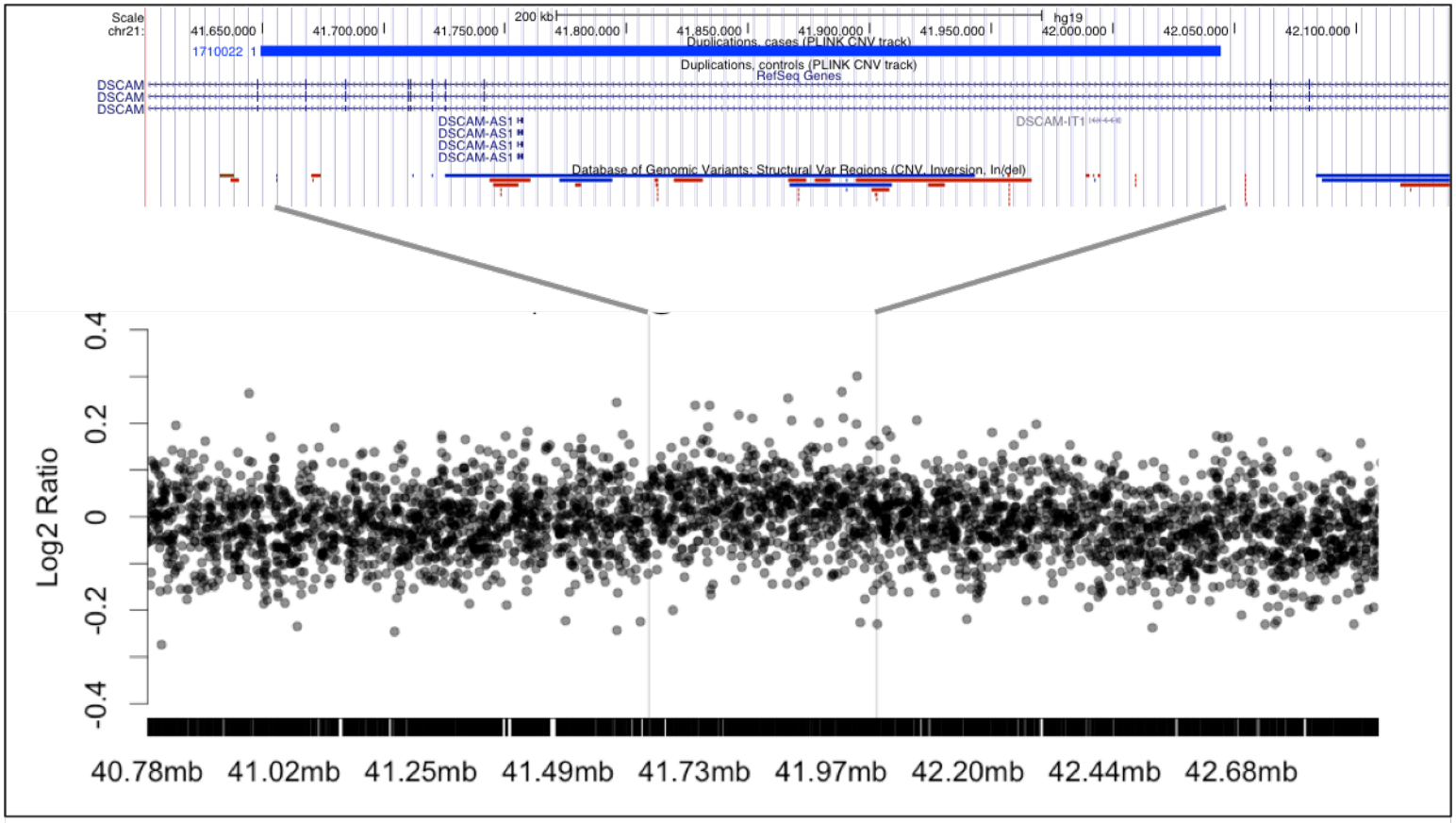


Figure 2.2.a: Duplication in a case at chr21:41649282-42044388 in a DS+AVSD case enveloping *DSCAM-AS1*

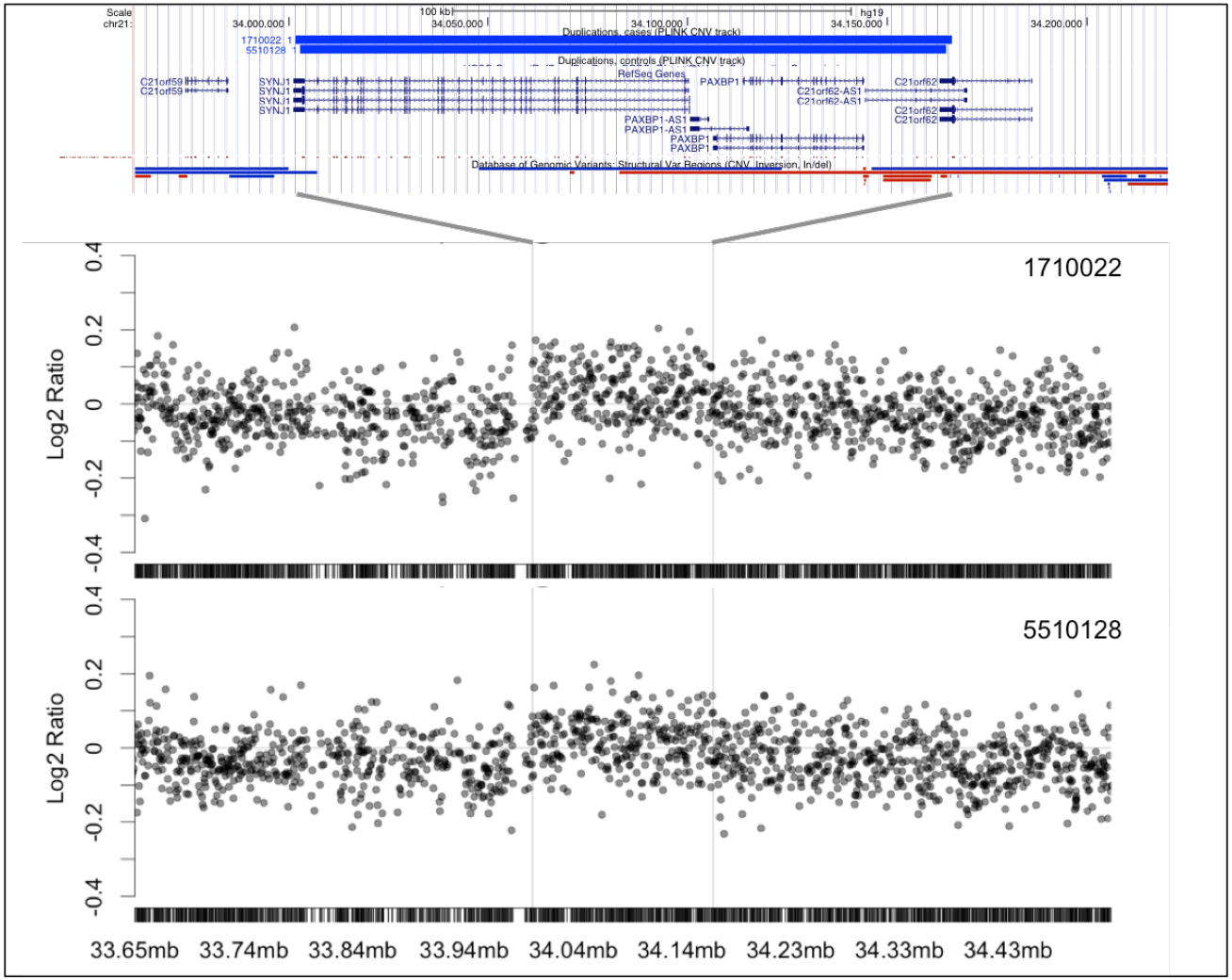


Figure 2.2.b: Duplications in two cases at chr21:34001528-3416643 and 34002843-34164679 enveloping *PAXBP-1*

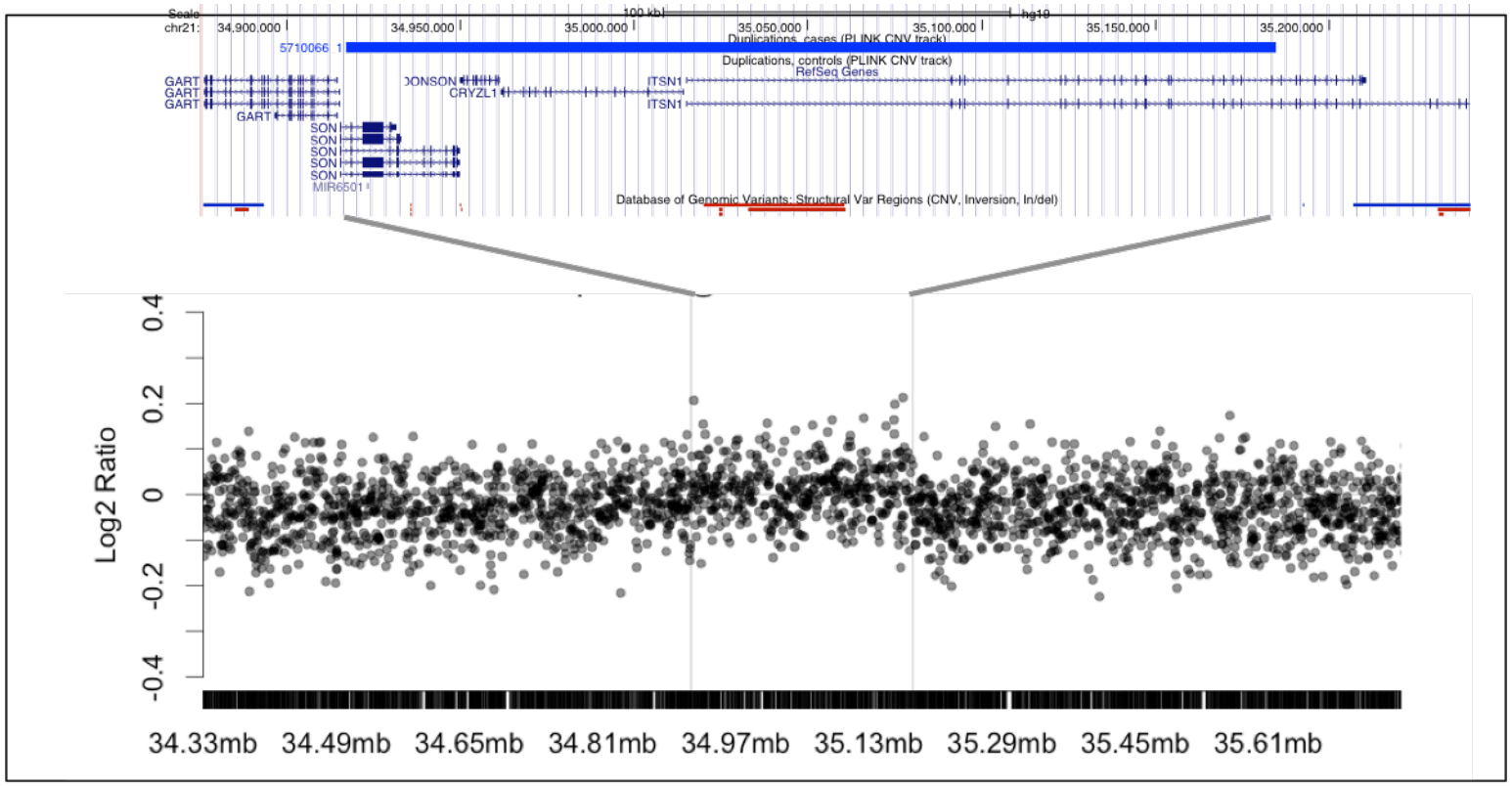


Figure 2.2.c: Duplication in a case at chr21:34916999-35184696 enveloping *DONSON* and *CRYZL1*.

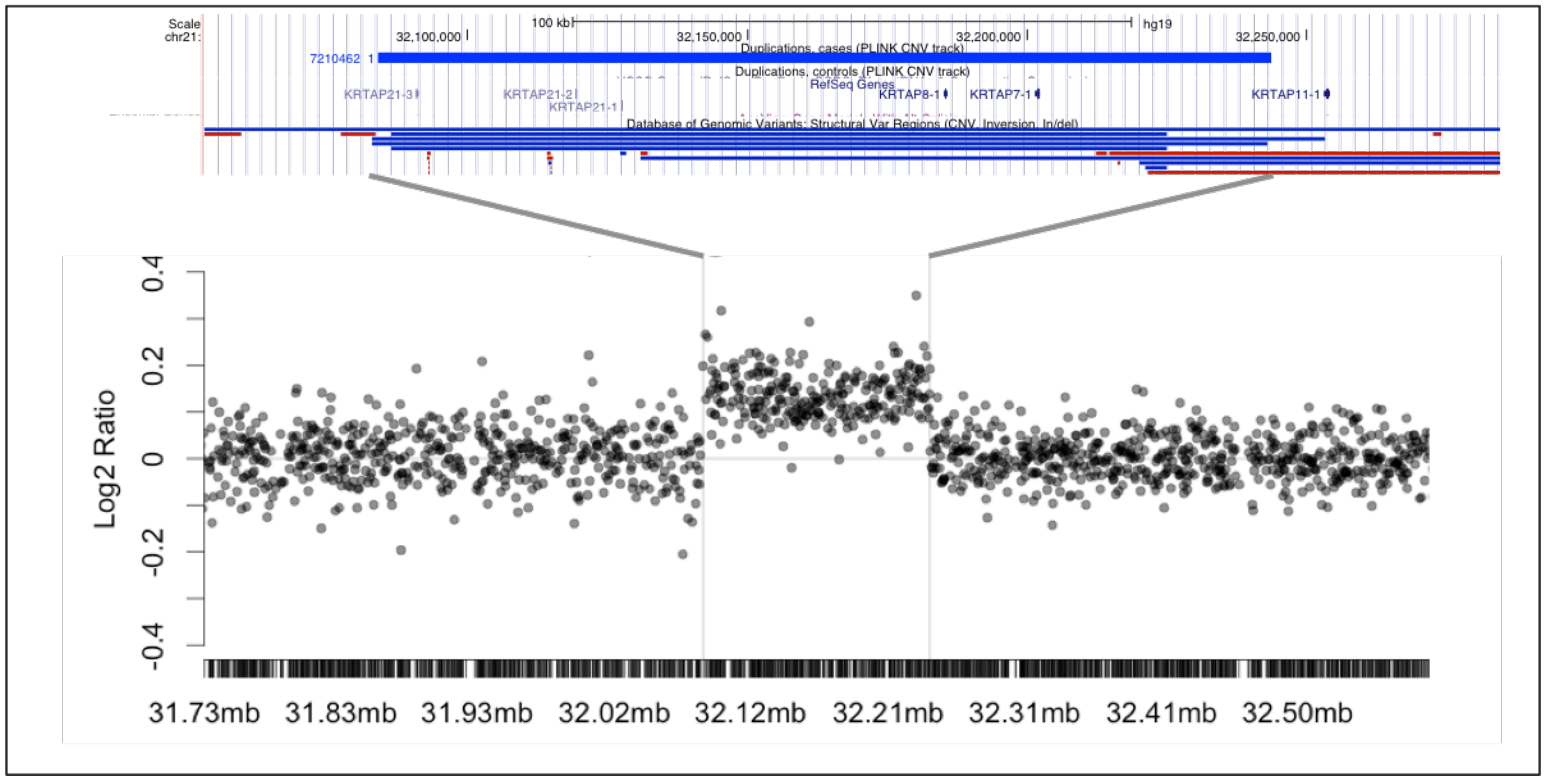


Figure 2.2.d: Duplication in a case at chr21:32083998-32243681 enveloping five *KRTAP* genes.

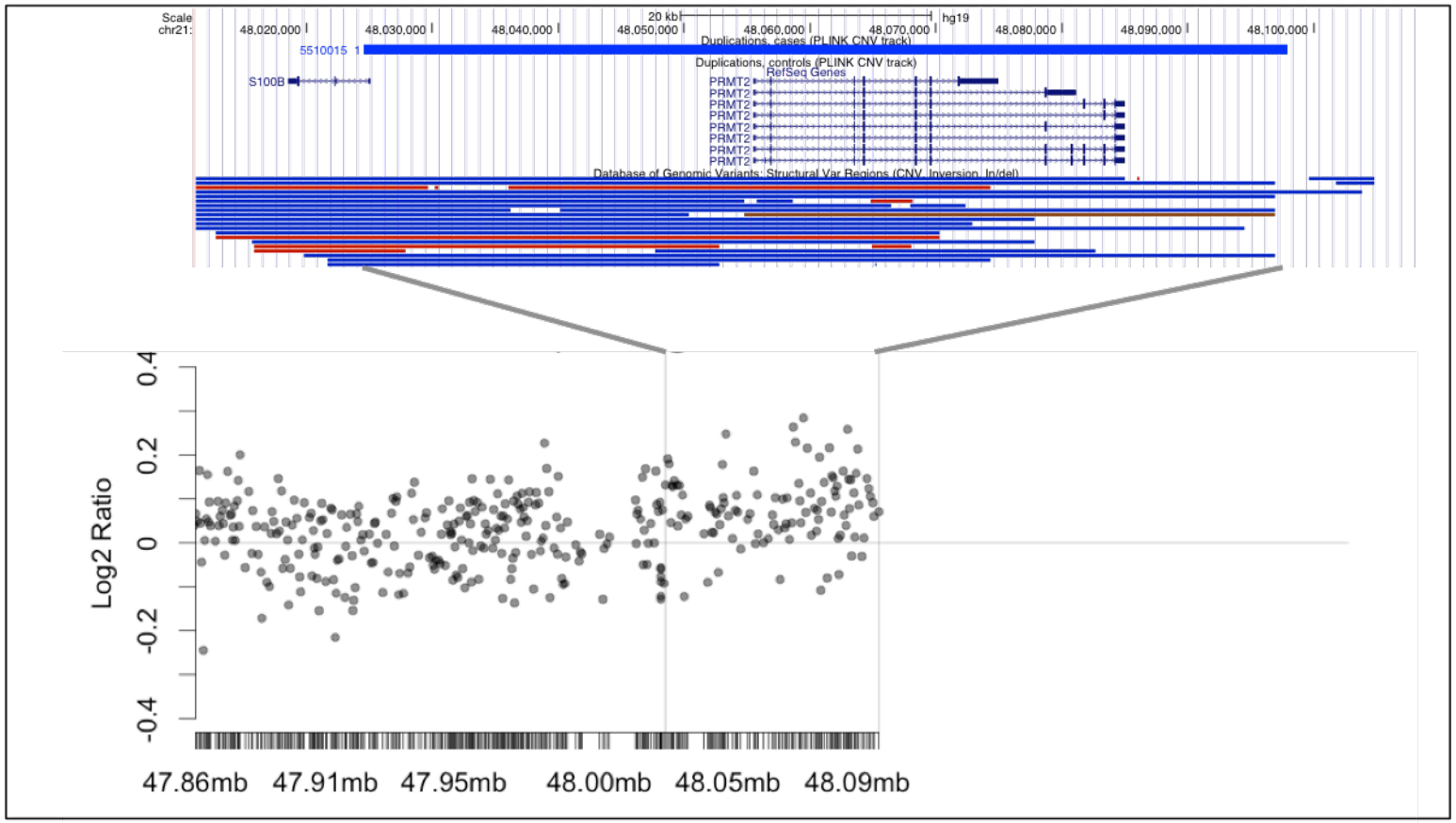


Figure 2.2.e: Duplication in a case at chr21:48024509-48097907 enveloping *PRMT2*.

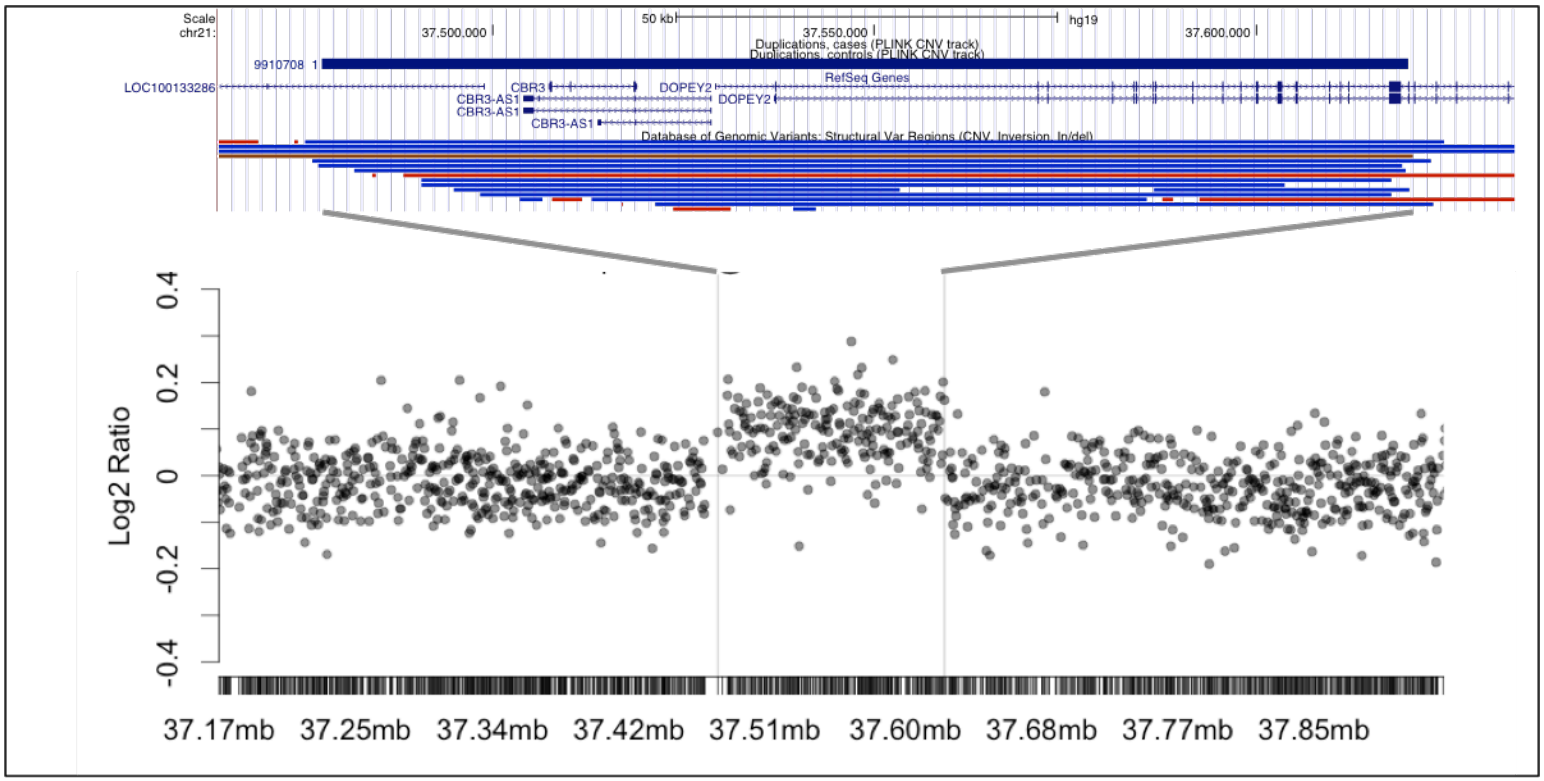


Figure 2.2.f: Duplication in a control at chr21:37477698-37619810 enveloping *CBR3*.

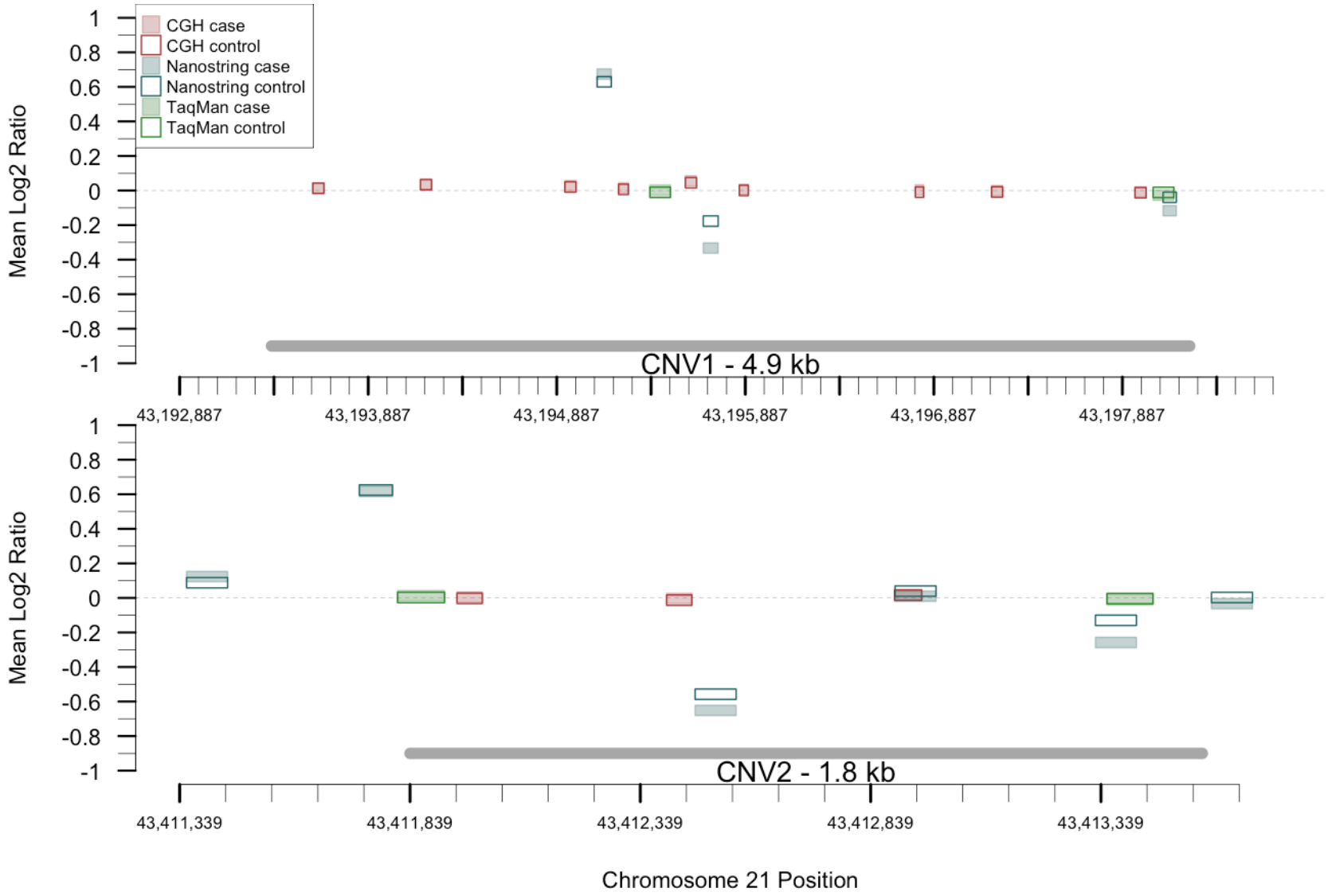

CGH case
CGH control
Nanostring case
Nanostring control
TaqMan case
TaqMan control
Mean Log2 Ratio
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
CNV1 - 4.9 kb
43,192,887
43,193,887
43,194,887
43,195,887
43,196,887
43,197,887
CNV2 - 1.8 kb
43,411,339
43,411,839
43,412,339
43,412,839
43,413,339
Chromosome 21 Position

Figure 2.3: Replication of previously reported DS+AVSD associated common CNVs. Mean Log2 ratios of cases (solid rectangles) and controls (outlined rectangles) are shown covering regions targeted for interrogation according to probe sequences (CGH and Nanostring) or expected amplicon (TaqMan). CGH probes (red rectangles) or TaqMan Copy Number assays (green rectangles) did not detect aberrant copy numbers or differences between cases and controls. Varying results were found across these loci with Nanostring probes (blue rectangles) with some probes showing differences in Log2 means between cases and controls for some probes and not others. Within these same proposed small CNV loci, Nanostring probes called all possible combinations of copy gain, loss and no change within the same small cohort. When compared to adjacent CGH and TaqMan probes, it is clear the Nanostring probes are not reliable predictors of copy number state at this locus.

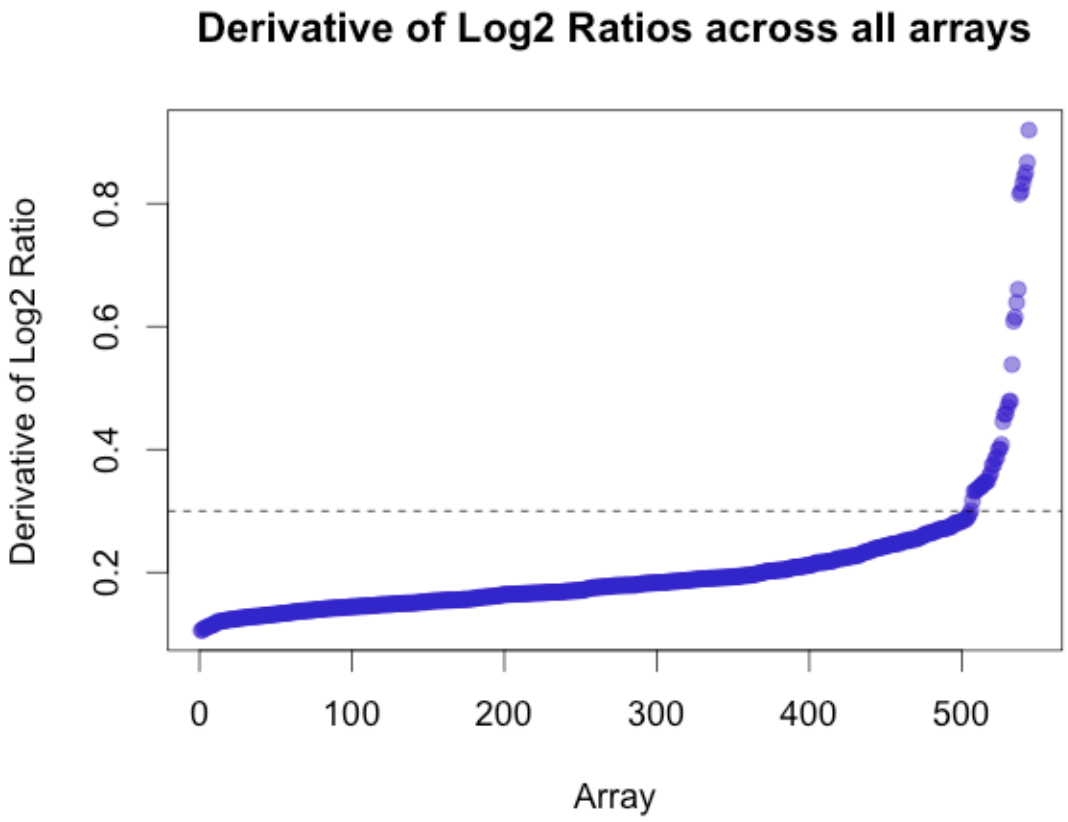


Figure 2.4: Derivative of Log2 Ratios (DLR) were calculated within Agilent's Feature Extraction software and arrays with DLR >= 0.3 were removed from further analyses.

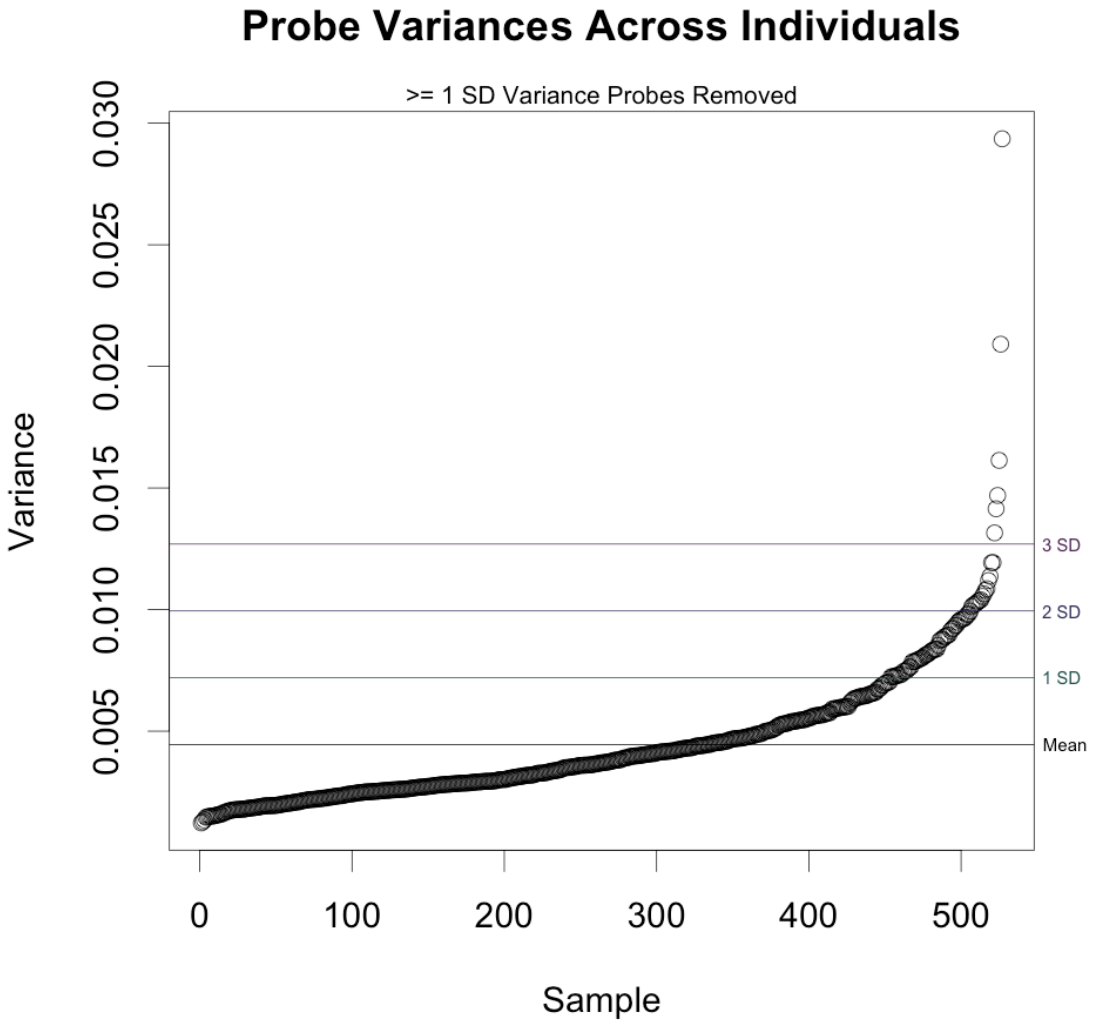


Figure 2.5: 74 arrays with intra-array probe log2 variance >= one standard deviation over the mean were removed from further analyses.

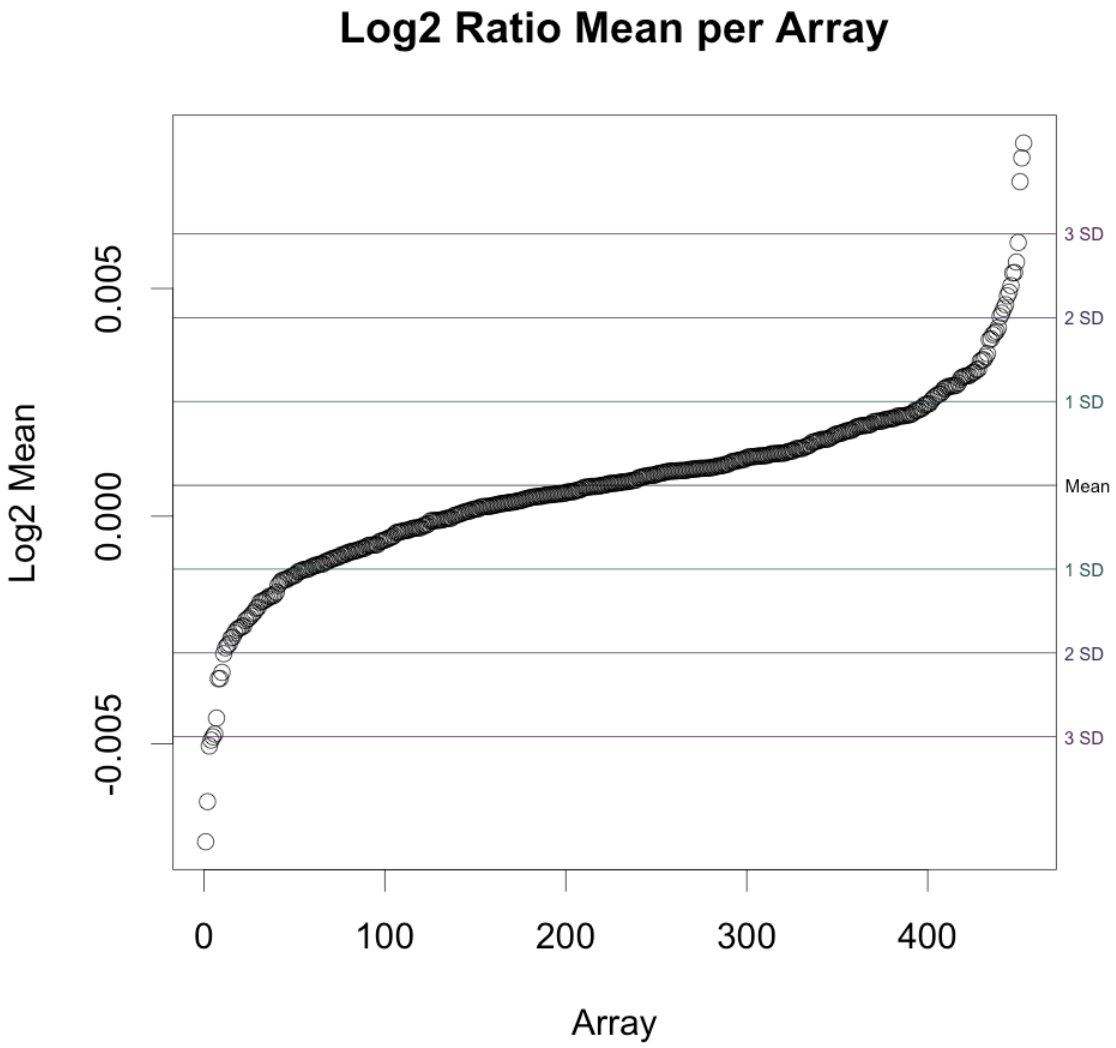


Figure 2.6: 25 arrays with intra-array log2 probe means outside of two standard deviations from the grand mean were removed from further analyses.

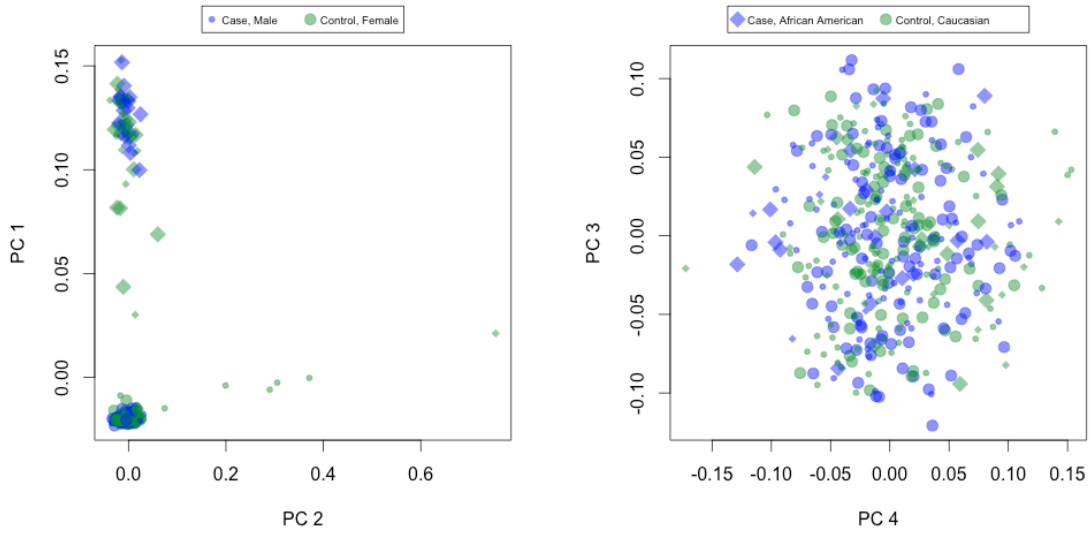


Figure 2.7.a Round 1 of Principle component analysis properly separates self-identifying African Americans from Caucasians.

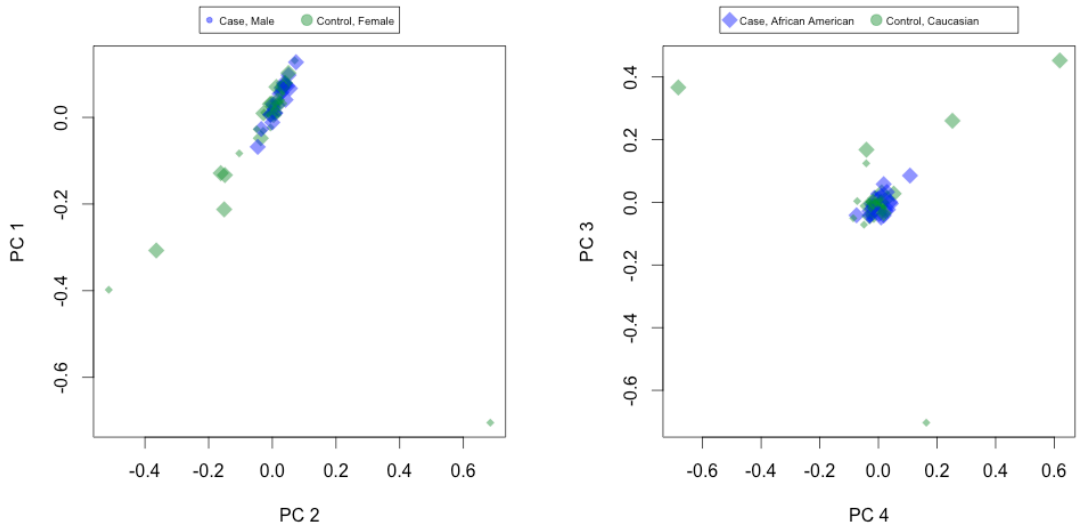


Figure 2.7.b: Round 2 of PCA in African Americans shows population outliers. African American individuals with PC1 eigenvalues < -0.127 were removed from further analyses.

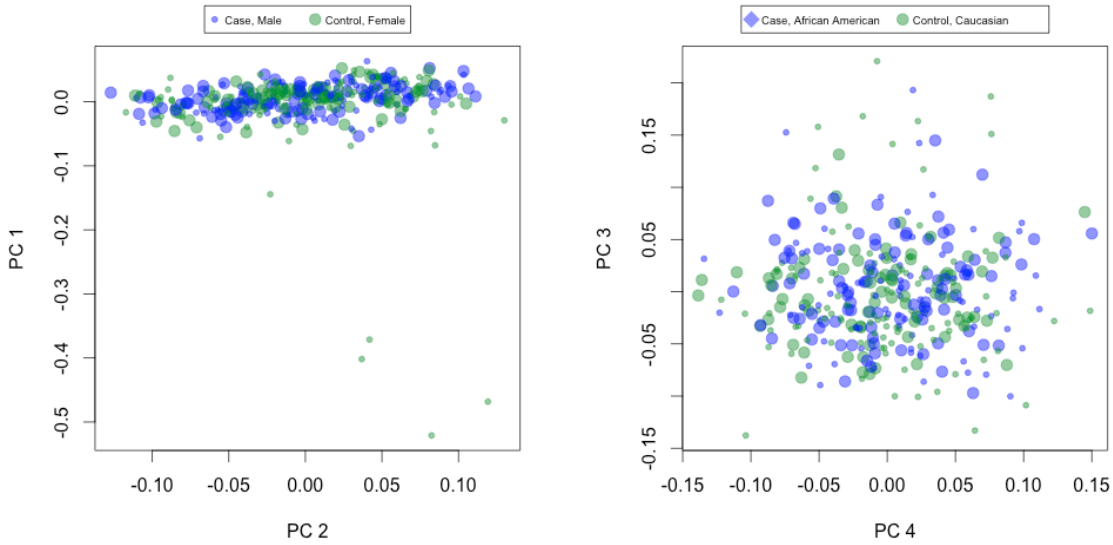


Figure 2.7.c: Round 2 of PCA in Caucasians shows population outliers. Caucasian individuals with PC1 eigenvalues < -0.1 were removed from further analyses.

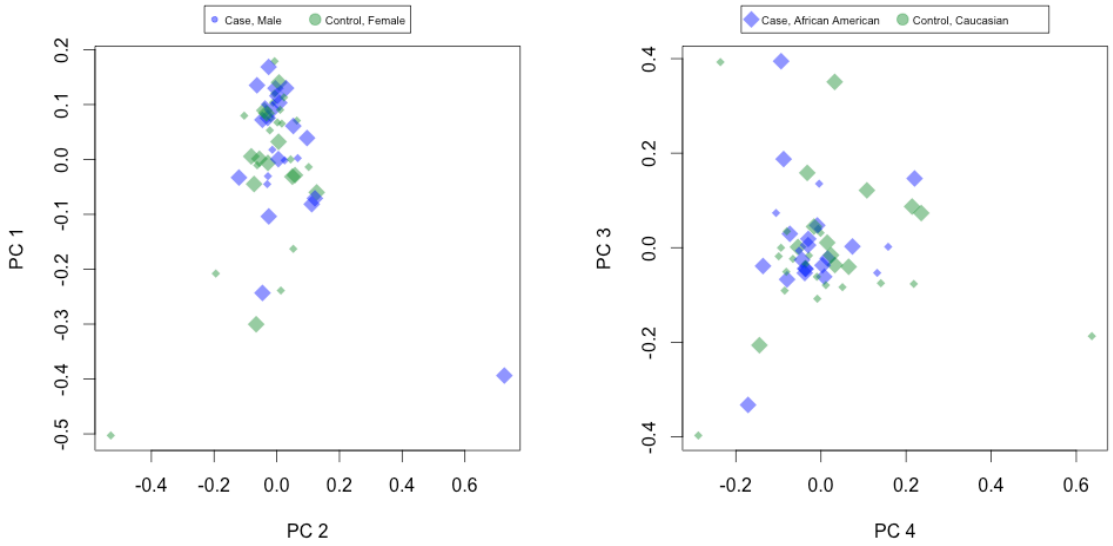


Figure 2.7.d: Final PCA plots of African Americans show random scattering of individuals without clear case/control biases.

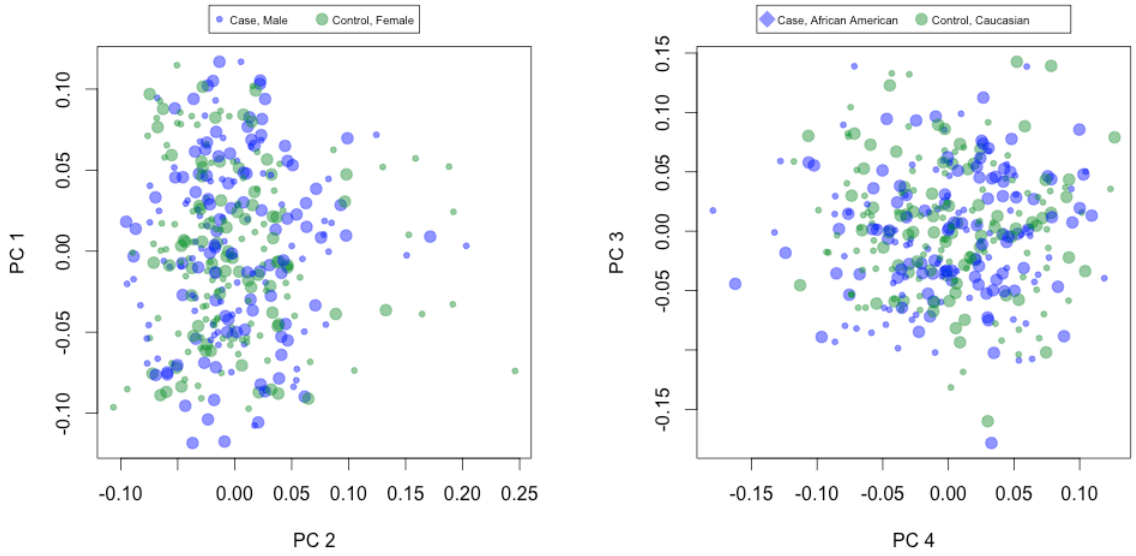


Figure 2.7.e: Final PCA plots of Caucasians show random scattering of individuals without clear case/control biases.

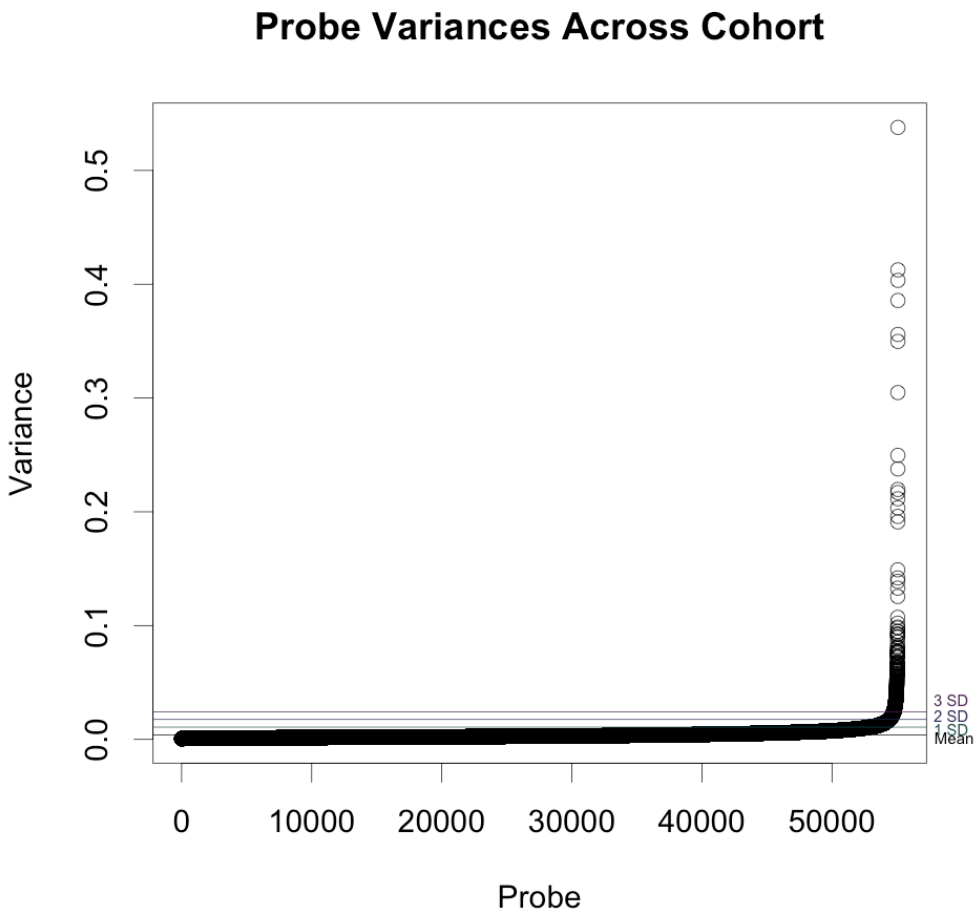


Figure 2.8: Data from 2,193 probes with an interarray normalized fluorescence log2 ratio variance >= 1 standard deviation over the mean were removed.

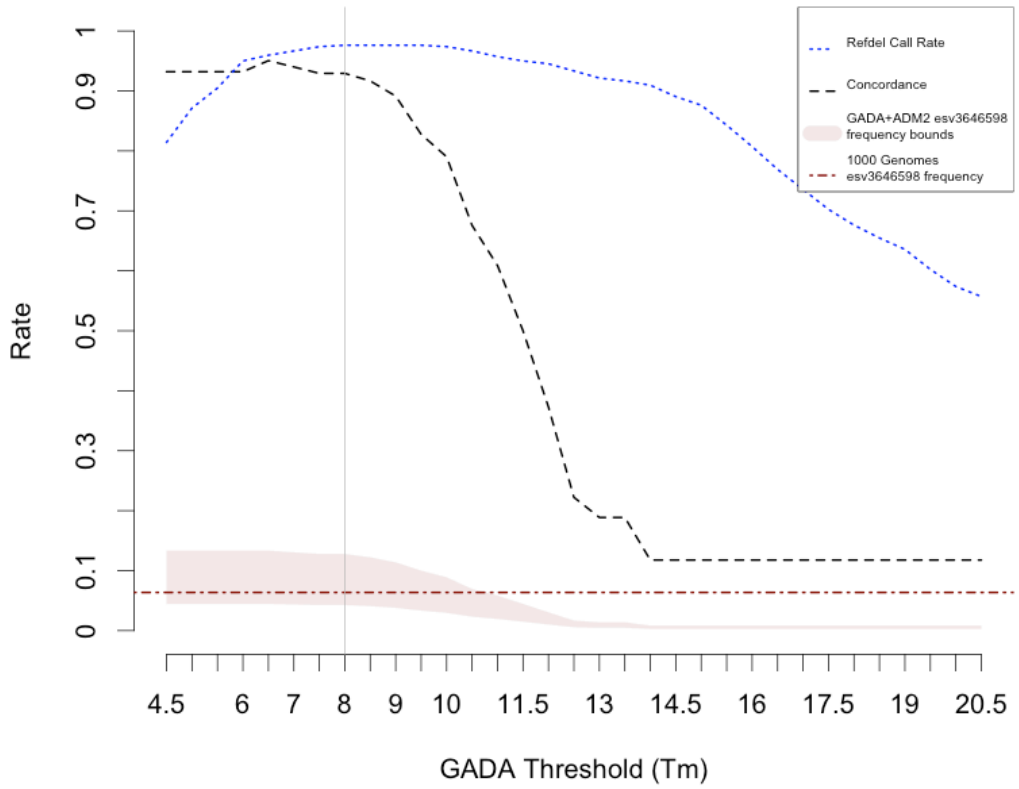


Figure 2.9.a: 1000 Genomes deletion esv3646598 detection across GADA Thresholds. To optimize the GADA Threshold parameter (T_m) for calling CNVs we aimed to optimize the detection of a known deletion (chr21:45,555,257-45,615,042) in our reference sample (called as a duplication in test samples) and to call the common deletion esv3646598 that was reported in 1000 Genomes Phase 3 Structural Variant release at a frequency of 0.0636 (red dash-dot line) in individuals with European Ancestry. The dotted blue line represents the consensus call rate of our reference deletion by ADM2 + GADA. The dashed black line is the concordance rate of ADM2 + GADA calling esv3646598. The red highlighted area displays the estimated frequency bounds of the esv3646598 deletion as detected by the two algorithms across increasing GADA T_m. We choose a GADA T_m of 8 that maximized our reference deletion call rate and called esv3646598 within the expected range.

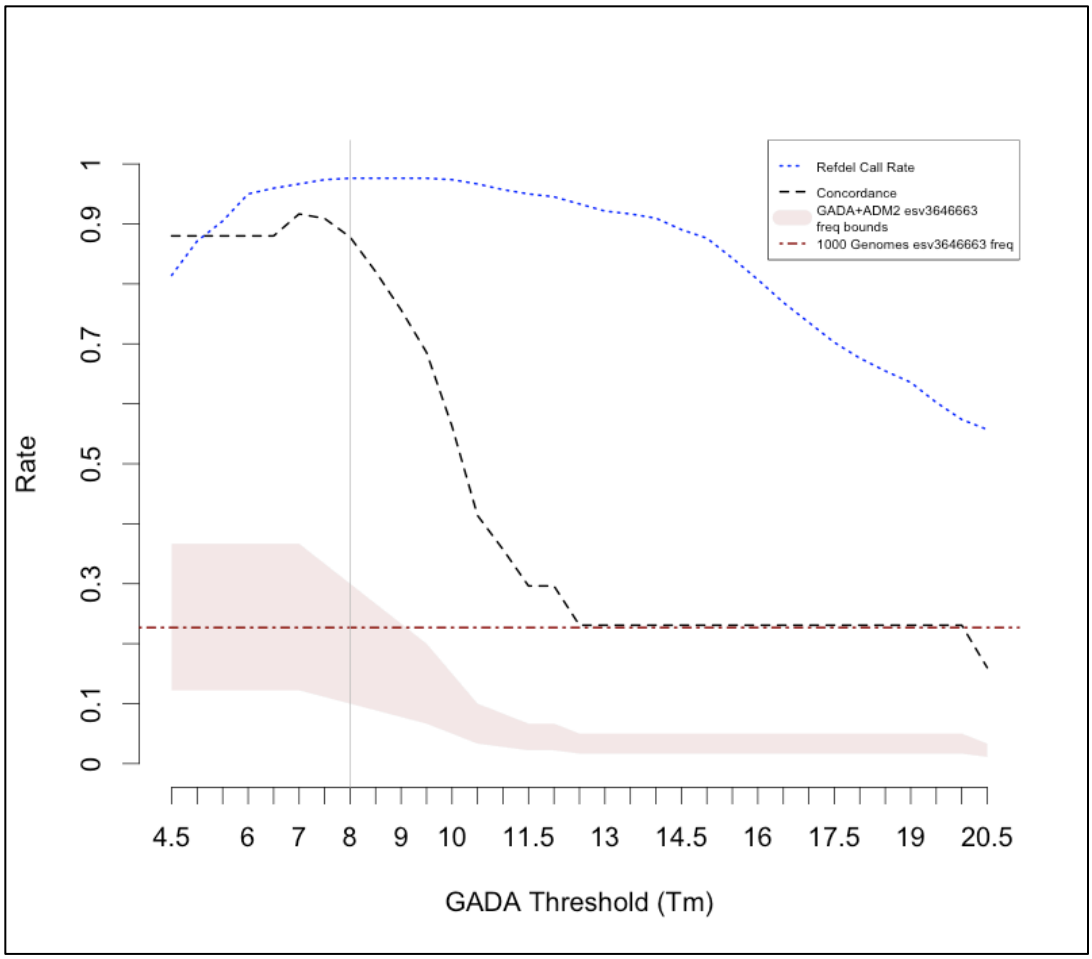


Figure 2.9.b: 1000 Genomes deletion esv3646663 detection across GADA Thresholds. To optimize the GADA Threshold parameter (T_m) for calling CNVs we aimed to optimize the detection of a known deletion (chr21:45,555,257-45,615,042) in our reference sample (called as a duplication in test samples) and to call the common deletion esv3646663 that was reported in 1000 Genomes Phase 3 Structural Variant release at a frequency of 0.2269 (red dash-dot line) in individuals with European Ancestry. The dotted blue line represents the consensus call rate of our reference deletion by ADM2 + GADA. The dashed black line is the concordance rate of ADM2 + GADA calling esv3646663. The red highlighted area displays the estimated frequency bounds of the esv3646663 deletion as detected by the two algorithms across increasing GADA T_m. We choose a GADA T_m of 8 that maximized our reference deletion call rate and called esv3646663 within the expected range.

INTRODUCTION

Our previous interrogations of single nucleotide polymorphisms (SNPs) and copy number variants (CNVs) in Down syndrome (DS)-associated atrioventricular septal defects (AVSD) has shown that no single common variant accounts for these individuals' high rate of heart defects (Ramachandran et al., 2014; 2015; Chapter 2). Sailani et al. (2013) is the singular report of a common CNV on chromosome 21 that associated with AVSD in individuals with DS. Our work detailed in Chapter 2 demonstrated how their finding appears to be a false positive. In the general euploid population, genome-wide association studies of thousands of individuals with and without congenital heart defects (CHD) has also proven that CHDs are not caused by single common variants in the general population either (Cordell et al., 2013; Hu et al., 2013).

This information gives us confidence that DS-associated AVSD is a complex disease involving numerous variants with variable effect sizes that act epistatically to disrupt heart development. Similar findings in other common complex diseases, combined with the dramatic reduction in the cost of human whole-exome and whole-genome sequencing, has led human genetics towards discovery of rare variants with larger effect sizes through direct sequencing. These variants are examined for their effects both singularly, in the case of rare variant with large effects, and in combination of sets of epistatic SNPs.

Exome sequencing of parent-offspring trios of CHD-affected euploid probands revealed a significant increase in protein-altering *de novo* variants in human-mouse orthologous genes highly expressed in developing mouse hearts (Zaidi et al., 2013). Of all genes harboring protein-altering *de novo* variants, they also found that these genes

were enriched in processes related to histone-3 lysine-4 methylation. Nonsyndromic AVSD has also been studied with whole exome sequencing. Al Turki et al. (2014) sequenced the exomes of 13 parent-offspring trios of probands with AVSD and 112 unrelated individuals with AVSD and compared their variation to that in 5,194 unrelated controls. They performed burden analyses by Fisher's exact test for rare alleles (minor allele frequency < 1%) in all genes between cases and controls and found *NR2F2* to harbor significantly more missense variants in cases than controls. Ackerman et al. (2012) took a targeted approach, sequencing 26 AVSD candidate genes in 141 case individuals with DS and AVSD (DS+AVSD cases) and 141 control individuals with DS and no heart defect (DS+NH controls). They filtered variants predicted to be the most deleterious and found a significant increase in these rare variants in genes enriched in the vascular endothelial growth factor pathway.

As whole exome and whole genome sequencing produces hundreds of thousands to millions of single nucleotide variants and insertions and deletions per individual, intelligent variant filtering is paramount to gain biological insight into the phenotype being studied. The three principal methods to help cull through the identified variants include filtering by minor allele frequency, by evolutionary conservation and by predicted protein disruption. The latter two have been joined into a single score called the Combined Annotation-Dependent Depletion (CADD) score, which have been assigned to all 8.6 billion possible SNVs in hg19 (Kricher et al., 2014). Their method combined 63 different annotations including conservation metrics, regulatory motifs, transcript expression and architecture, and protein-disruption predictions. Those annotation data and a limited number of significant interactions terms were used to train a support vector

machine (SVM) with a linear kernel to predict the status of a variant site 1) being fixed in the human population since the inferred human-chimpanzee shared ancestor or 2) being an alternative, simulated *de novo* variant. The trained SVM was then applied to all possible SNVs in hg19 and the resulting C scores for each possible SNV were then ranked and scaled from 1 to 99. These C-scaled scores represent a log-increasing prediction of how deleterious a variant could be, i.e., $C \geq 10$ denotes variants within the highest 10% of raw C score, $C \geq 20$ are in the highest 1%, and so on. CADD scores have a higher area under the curve in receiver operator curves for predicting known pathogenic mutations than that for previous deleterious-prediction algorithms. CADD scores also cover all possible human variants mapped in hg19, whereas other algorithms are limited to prescribed regions, such as exons.

The majority of variation revealed in exome and genome sequencing is rare. The scarcity of these variants does not allow for conventional statistical tests of association, as power will be limited in most ascertainable cohort sizes. To overcome these statistical limitations, variants can be grouped into sets based on genomic regions, such as those within individual genes or those in a specific pathway. A burden test can then be applied on the sum effects of all variants in the set to identify that set's contribution to the phenotype of interest. A burden test assumes that minor allele effects will all be in the same direction within a set, either all protective, or all risk. Often, the study design and phenotype of interest do not warrant this model assumption, as the underlying biology is not known and the elucidation of which is the goal. To overcome this limitation of burden testing, Wu et al. (2010) devised a sequence kernel association test (SKAT) that allows for modeling of the joint effect of risk and protective alleles within a set via a logistic

kernel-machine-based test. SKAT also allows for inclusion of covariates, such as principal components. While burden tests are less powerful than SKAT when there is a combination of protective and risk alleles within a set, they are more powerful than SKAT if most minor allele effects in the set are in the same direction; this may be the case for rare variants in some genes. An optimal unified test (SKAT-O) of both burden tests and SKAT was developed to maximize the value in both types of combined variant testing (Lee et al., 2012). SKAT-O models both SKAT and the burden test for each defined variant set and finds the optimal linear combination of SKAT and burden tests. Thus, it optimizes power for all scenarios. In small sample sizes less than 1,000 individuals, SKAT-O also estimates the sample variance and kurtosis allowing for proper reference distribution, minimizing type I errors.

Finally, the vast amount of novel data derived from genome-scale sequencing projects provides an opportunity to discover novel biological underpinnings of a phenotype through non-hypothesis driven queries. Pathway analysis allows for unsupervised clustering into biologically related genes of a large number of genes identified harboring different types of genetic variation in a studied cohort. The foundation of pathway analysis is in the ontology terms defined for each gene. The Gene Ontology Consortium is a group of laboratories supported by a P41 grant from the National Human Genome Research Institute (grant 5U41HG002273-14) that combine bioinformatic methods with expert review to annotate genes into defined ontology terms (The Reference Genome Group of the Gene Ontology Consortium, 2009). The computational assignment relies on a hidden Markoff Model to group protein sequences by motif allowing for GO assignment of nearly all mammalian proteins (Thomas et al.,

2003). A caveat to pathway analysis is the intersection of large pathways and subsequent multiplication of terms for genes with broad function. To address this, expert curating by Ph.D. scientists have further reduced these grand GO lists into "slim" versions (Mi et al., 2005).

In this chapter, we present the results of applying whole genome and whole exome sequencing to elucidating the perturbed mechanisms that lead to atrioventricular septal defects in the sensitized Down syndrome population. We begin with an exome-wide association study to confirm previous results showing that common SNPs of large effect do not account for AVSD in DS. We then apply SKAT-O to CADD-predicted-deleterious variants in two different gene-sets of cilia genes. The first gene list is a very narrow list of cilia genes with a stringent burden-of-proof for their ciliary involvement, and thus a low rate of false positives at the allowance of a missing many cilia genes. The second list is a broader cilia gene list containing genes with minimal evidence for cilia involvement, and thus a low rate of false negatives with possible non-cilia genes. Finally, we ask what pathways are enriched in genes containing rare variants that nominally associate with AVSD or no CHD in DS by Pathway analysis.

METHODS

Subjects

Participant samples were an overlapping set of those described in Chapter 2. They were collected as described previously (Freeman et al., 1998; Freeman et al., 2008; Locke et al., 2010; Ramachandran et al., 2015; Ramachandran et al., 2015; Chapter 2). Individuals diagnosed with full or translocation trisomy 21, documented by karyotype, were recruited from centers across the United States. Institutional review boards at each enrolling institution approved protocols and informed consent was obtained from a custodial parent for each participant. A single cardiologist (K. Dooley) identified cases from medical records as individuals with a complete, balanced AVSD diagnosed by echocardiogram or surgical reports (DS+AVSD). Controls were classified as individuals with a structurally normal heart, patent foramen ovale, or patent ductus arteriosus (DS+NH).

Genomic DNA was extracted from lymphoblastoid cell lines with the Puregene DNA purification kit by manufacturer's protocol (Qiagen, Valencia, CA). DNA quantity and quality were checked on a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, DE) and assessed for integrity on 0.8% agarose gels stained with ethidium bromide.

Data cleaning

Paired-ended whole genome sequencing (WGS) was performed by Hudson Alpha (Huntsville, AL) on 164 DS+AVSD cases and 38 DS+NH controls to a target depth of 30x. Raw fastq data were mapped and variants were called using Emory's PEmapper. Actual mean coverage depth ± standard deviation (sd) was 30.2 ± 4.1 and two cases and four controls were dropped for having mean coverage < 2 sd below the mean. Mean

transition/transversion ratio ± sd was 2.05 ± 0.07. Exomes were captured on 136 DS+AVSD cases and 138 controls with Nimblegen V2 exome target and single-end sequenced on an Illumina MiSeq to 63x coverage. Fastq files from WES were also mapped and had variants called with Emory's PEmapper. Mean average depth ± sd of exome sequencing was 61.9 ± 12.6 and mean transition/transversion ratio ± sd was 2.9 ± 0.06. No sample had average coverage < 2 sd below the mean and, thus, none were dropped. Variants were annotated using SeqAnt (http://seqant.genetics.emory.edu).

Data from WES and WGS were merged requiring that each base be covered in WES and WGS resulting in 375,221 single nucleotide variants (SNV) across 298 cases and 172 controls. To account for differences in platform efficiency and case/control ratios between WES and WGS, we removed 4,718 variants that common (MAF > 15%) in one platform but rare (MAF < 1%) in the other. First, gross SNV and sample failures were addressed by removing 23,611 SNVs with missingness > 20% and then six cases and one control with > 20% missing genotypes. Next we increased our stringency and removed 25,833 SNVs with missingness > 1% and two cases with > 1% missing genotypes. Finally, we removed 190 SNVs that failed the exact test for Hardy-Weinberg equilibrium at a p-value < 10^{-6}. Sex was then checked by computing F statistics for X chromosome heterozygosity, which were also used to impute sex on individuals from Coriell missing sex data (Weir and Cockerman, 1984; Chang et al., 2015). Two individuals (one case and one control) thought to be male were flagged as female and removed from further analyses (Figure 3.1).

Data were prepared for Principal Component Analysis (PCA) by taking common SNPs (MAF > 0.05) and pruning out SNPs in linkage disequilibrium with an $r^2 > 0.2$,

stepping along five SNPs at a time within 50kb windows. PCA was then performed using the SNPRelate package in R (Zheng et al., 2012). Round one correctly segregates self-identified African Americans, who were removed from further analyses (Figure 3.2). PCA was repeated a second round and nine case outlier samples were detected at PC1 < -0.1 and PC2 > 0.1 and removed (Figure 3.3). A third and final round of PCA was performed and indicated that no population outliers were present (Figure 3.4). The final dataset for genomic analyses contained 207,600 SNVs across 245 cases and 130 controls, providing a total genotyping rate of 0.9997.

Exome wide association analysis

35,071 common single nucleotide variants (MAF > 5%) were tested in PLINK v1.9 (Chang et al., 2015) for genotypic association with AVSD by logistic regression including the first five eigenvectors from the final round of PCA as covariates. None of the covariates were significant predictors of AVSD status (Figure 3.5); thus the association analysis was repeated without covariatesWith 35,071 SNPs across 245 DS+AVSD cases and 130 DS+NH controls of Caucasian ancestry, we had 80% power to detect a SNP with an additive effect, at 5% MAF with an odds ratio of ~17 or at 10% MAF with an OR of ~6 (Figure 3.6).

SKAT-O analysis

119,921 SNVs with CADD score ≥ 15 were assigned to two ciliome gene-set lists. The first, "van Dam Ciliome," includes 301 genes containing 3,289 SNVs. These genes were manually curated from the largest collection of ciliary datasets housed in Cildb (Arnaiz et al., 2009; 2014) and included in the list if they met the following criteria: 1. Data based on experimental methods; 2. One-to-one homolog of human gene; 3. Gene was shown to function in cilia in two different organisms; and 4. Gene was shown to function in cilia in at least two types of experiments (van Dam et al., 2013). A second cilia gene list, "Leroux Ciliome," includes 1,748 genes containing 21,426 SNVs. This gene list was assembled by Leroux et al. (2006) through an extensive literature review of both experimental and bioinformatically derived cilia gene candidates. McClintock et al. (2008) extended this list by bioinformatically identifying genes overrepresented in mouse tissues containing cell types enriched in cilia (olfactory epithelium, testes, vomeronasal organ, trachea, and lung). Using this list, Ripoll et al. (2013) found a statistically significant enrichment of genes differentially expressed between lymphoblast cell lines from individuals with Down syndrome and AVSD and individuals with DS and no congenital heart defect. We have also found a suggestion of both chromosome 21 CNVs (Chapter 2) and of genome-wide large deletions (Ramachandran et al., 2014) enriched in the Leroux Ciliome. 125 of the van Dam Ciliome genes are in the Leroux Ciliome list. Both gene-sets were then analyzed in R by SKAT-O (Lee et al., 2012).

Rare variant pathway analysis

172,532 rare SNVs (MAF < 5%) were tested in PLINK v1.9 for association with AVSD by four genetic models (genotypic (AA vs. Aa vs. aa), allelic (A vs. a), recessive (aa vs. Aa/AA) and dominant (AA/Aa vs. aa)), computing p-values by Fisher's exact test. 2,203 nominally significant SNVs with non-adjusted p-value ≤ 0.05 and CADD score ≥ 15 (annotated via SeqAnt) were assigned to cases or controls per the associative direction of the minor allele. For both case and control SNV lists, the gene nearest to each SNV, annotated by SeqAnt, was assembled into lists of AVSD gene candidates for cases and controls respectively. The lists contained 215 genes for cases and 890 for controls. Each gene list was than analyzed for enrichment in five pathways curated by PANTHER: PANTHER Pathways, PANTHER Proteins, PANTHER GO-slim Biological, PANTHER GO-slim Cellular and PANTHER GO-slim Molecular (Mi et al., 2005). The enrichment analysis was performed using Panther's overrepresentation test v.10.0 (release 20160321; http://pantherdb.org/tools/compareToRefList.jsp).

RESULTS

Exome wide association analysis

35,071 single nucleotide polymorphisms with minor allele frequency > 0.05 were tested for additive association to atrioventricular septal defects (AVSD) in individuals with Down syndrome (DS) by logistic regression. No SNP reached genome-wide or near

genome-wide significance (Figure 3.7). The quantile-quantile plot indicates there is no population stratification within the data (Figure 3.8).

SKAT analysis

van Dam Ciliome

3,289 SNVs grouped into 301 sets of genes in the van Dam Ciliome gene list (van Dam et al., 2013) were analyzed with SKAT-O. No gene-set reached Bonferroni-corrected significance, though 24 genes passed nominal significance levels (Figure 3.9, Table 3.1). The quantile-quantile plot indicates that a large proportion of van Dam Ciliome genes may be truly associated with AVSD in DS, as there is a consistent departure above the expected p-values (Figure 3.10). This departure is not due to population stratification as shown in the GWAS QQ plot (Figure 3.8) and the principal component analysis (Figure 3.4).

Leroux Ciliome

21,426 SNVs grouped in 1,748 sets of genes in the Leroux Ciliome gene list (Inglis et al., 2006; McClintock et al., 2008) were analyzed with SKAT-O. No gene-set reached Bonferroni-corrected significance, though a single gene, *GMPPA,* had a marginally significant p-value = 0.0002; 104 other genes passed nominal significance levels (Figure

3.11, Table 2.2). The QQ plot indicates modest departure of many Leroux Ciliome genes above the expected p-values (Figure 3.11).

Pathway analysis

Case pathways

215 genes harboring rare variants predicted to be deleterious by CADD scores with minor alleles associating with DS+AVSD cases were analyzed by the PANTHER overrepresentation test v.10.0 against five curated Panther lists (Panther Protein, Panther Pathways, GO-slim Biological, GO-slim Cellular and GO-slim Molecular). Most significant results are presented in Table 3.3.

The most significantly enriched protein set belongs to the actin family cytoskeletal proteins (PC00041), which had four times the number of genes than expected by chance (Bonferroni-corrected p-value = 0.003). The enriched levels of deleterious alleles associating with cases in cytoskeletal genes was also seen in the broader Panther Protein category of cytoskeletal (PC00085, Bon. p-value = 0.054), in the GO-slim Molecular sets of actin-binding (GO:0003779, Bon. p-value = 0.017) and cytoskeletal protein binding (GO:0008092, Bon. p-value = 0.051), and the GO-slim Cellular set of actin cytoskeleton (GO:0015629, Bon. p-value = 0.045).

Control pathways

The DS+NH control gene set was much larger than the case gene set, including 890 genes. Pathways enriched in these genes cover a wide swath of basic cell functioning (Table 3.4). Within the GO-slim Molecular set, control genes were found to be enriched in transporter activity (GO:0005215, Bonferonni p-value = 0.004) and hydrolase activity (GO:0016787, Bon. p-value = 0.023). Within the GO-slim Biological sets, genes containing deleterious alleles in DS+NH that are involved in cell movement were significantly enriched (GO:0006928, Bon. p-value = 0.0004, GO:0051179, Bon. p-value = 0.011 and GO:0006810, Bon. p-value = 0.02).

DISCUSSION

We have previously shown that common variation does not cause AVSD in the sensitized DS population (Ramachandran et al., 2014; 2015, Chapter 2). To investigate the contribution of rare single nucleotide polymorphisms, we performed and analyzed the data from whole genome and whole exome sequencing on 245 DS+AVSD cases and 130 DS+NH controls.

We tested for the contribution of genes functioning in cilia based on two different set of cilia genes, which have a minimal overlap (Figure 3.13). We did not find any gene within either set to be multiple-test correction-significant. *GMPPA* was the most associated gene within the Leroux Ciliome gene sets (Non-corrected p-value = 0.0002). Four out of the five SNVs within *GMPPA* are rare and have the minor allele associating

with DS+NH controls (Table 3.5). All changes are nonsynonymous except rs34873891, which is a serine to proline mutation in a single control individual. All variants do have high CADD scores and, as they are silent variants, likely reflect a high level of evolutionary conservation. Though no individual cilia gene stood out as having a large effect on the AVSD phenotype in DS, the general inflation of p-values over expectation might indicate the significance of these genes being part of the highly heterogeneous fabric of AVSD.

After performing pathway analysis, we found that rare variants in DS+AVSD cases, predicted to be deleterious, are enriched in genes involved with actin binding and remodeling. Actin is an essential protein found in nearly all eukaryotic cells and it is the major subunit in microfilaments. Microfilaments are critical cellular scaffolds for organelles and cell movement. Actin also plays a necessary role in muscle contraction, including cardiac muscle.

An analysis of transcripts up-regulated in mouse heart development showed an enrichment in actin remodeling proteins (Gan et al., 2015). Dominant mutations in the actin subunit *ACTC1* have been shown to cause atrial septal defect in humans (Matsson et al., 2007; Augiere et al 2015). Numerous actin remodeling proteins such as *zyxin* (Mori et al., 2009) and *β2-spectrin* (Lim et al., 2013) are necessary for proper heart development, including for the critical epithelial to mesenchymal transition that gives rise to the population of cells that migrate to form the endocardial cushion. Furthermore, actin is intimately involved in cilia. Nodal cilia are required during early gastrulation to establish a left-right axis (Koefoed et al., 2016) at least partially via the gradient expression of the actin-binding protein *Ablim1* (Stevens et al., 2010). Ciliogenesis itself relies on actin-

remodeling. Knockdown of the actin regulatory kinases *LIMK2* and *TESK1* in human epithelial cells significantly increased the number of ciliated cells (Kim et al., 2015). Mice harboring mutations in *Pk1* fail to properly orient actin filaments in the septal mesenchyme resulting in cardiac outflow defects as well as fewer and stunted cilia in embryonic fibroblasts (Gibbs et al., 2016).

The multitude of signaling pathways mediated by, and required for, proper cilia functioning provides a large target for mutations to accumulate across the population. While it is clear that no variants of large effect cause AVSD in the DS population, this group remains sensitized to AVSD by the presence of trisomy 21. It appears that deleterious variation in actin remodelers that can be tolerated in a euploid state, may lead to AVSD in DS.

References

Ackerman, C., Locke, A. E., Feingold, E., Reshey, B., Espana, K., Thusberg, J. et al. (2012). An Excess of Deleterious Variants in VEGF-A Pathway Genes in Down-Syndrome-Associated Atrioventricular Septal Defects. *American journal of human genetics*, *91*(4), 646-659.

Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M.-P., Lindsay, S. et al. (2014). Rare Variants in NR2F2 Cause Congenital Heart Defects in Humans. *American journal of human genetics*, *94*(4), 574-585.

Arnaiz, O., Malinowska, A., Klotz, C., Sperling, L., Dadlez, M., Koll, F. et al. (2009). Cildb: a knowledgebase for centrosomes and cilia. *Database*, *2009*.

Arnaiz, O., Cohen, J., Tassin, A.-M., & Koll, F. (2014). Remodeling Cildb, a popular database for cilia and links for ciliopathies. *Cilia*, *3*(1), 1-10.

Augiere, C., Megy, S., El Malti, R., Boland, A., El Zein, L., Verrier, B. et al. (2015). A Novel Alpha Cardiac Actin (<italic>ACTC1</italic>) Mutation Mapping to a Domain in Close Contact with Myosin Heavy Chain Leads to a Variety of Congenital Heart Defects, Arrhythmia and Possibly Midline Defects. *PLoS ONE*, *10*(6), e0127903.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 1-16.

Cordell, H. J., Bentham, J., Topf, A., Zelenika, D., Heath, S., Mamasoula, C. et al. (2013). Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nature genetics*, *45*(7), 822-824.

Gan, J., Sonntag, H.-J., Tang, M. K., Cai, D., & Lee, K. K. H. (2015). Integrative Analysis of the Developing Postnatal Mouse Heart Transcriptome. *PLoS ONE*, *10*(7), e0133288.

Gibbs, B. C., Damerla, R. R., Vladar, E. K., Chatterjee, B., Wan, Y., Liu, X. et al. (2016). Prickle1 mutation causes planar cell polarity and directional cell migration defects associated with cardiac outflow tract anomalies and other structural birth defects. *Biology Open*, *5*(3), 323-335.

Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., Lin, Y. et al. (2013). A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nature genetics*, *45*(7), 818-821.

Inglis, P. N., Boroevich, K. A., & Leroux, M. R. (2006). Piecing together a ciliome. *Trends in Genetics*, *22*(9), 491-500.

Kim, J., Jo, H., Hong, H., Kim, M. H., Kim, J. M., Lee, J.-K. et al. (2015). Actin remodelling factors control ciliogenesis by regulating YAP/TAZ activity and vesicle trafficking. *Nat Commun*, *6*.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, *46*(3), 310-315.

Koefoed, K., Veland, I. R., Pedersen, L. B., Larsen, L. A., & Christensen, S. T. (2014). Cilia and coordination of signaling networks during heart development. *Organogenesis*, *10*(1), 108-125.

Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A. et al. (2012). Optimal Unified Approach for Rare-Variant Association Testing with Application to Small-Sample Case-Control Whole-Exome Sequencing Studies. *The American Journal of Human Genetics*, *91*(2), 224-237.

Li, D., Sinha, T., Ajima, R., Seo, H.-S., Yamaguchi, T. P., & Wang, J. (2016). Spatial regulation of cell cohesion by Wnt5a during second heart field progenitor deployment. *Developmental Biology*, *412*(1), 18-31.

Lim, J. A., Baek, H. J., Jang, M. S., Choi, E. K., Lee, Y. M., Lee, S. J. et al. (2013). Loss of Œ≤2-spectrin prevents cardiomyocyte differentiation and heart development. *Cardiovascular Research*, *101*(1), 39-47.

Matsson, H., Eason, J., Bookwalter, C. S., Klar, J., Gustavsson, P., Sunneg√•rdh, J. et al. (2008). Alpha-cardiac actin mutations produce atrial septal defects. *Human Molecular Genetics*, *17*(2), 256-265.

McClintock, T. S., Glasser, C. E., Bose, S. C., & Bergman, D. A. (2008). Tissue expression patterns identify mouse cilia genes. *Physiological Genomics*, *32*(2), 198-206.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S. et al. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Research*, *33*(suppl 1), D284-D288.

Mori, M., Nakagami, H., Koibuchi, N., Miura, K., Takami, Y., Koriyama, H. et al. (2009). Zyxin Mediates Actin Fiber Reorganization in Epithelial‚ÄìMesenchymal Transition and Contributes to Endocardial Morphogenesis. *Mol Biol Cell Molecular Biology of the Cell*, *20*(13), 3115-3124.

Ramachandran, D., Mulle, J. G., Locke, A. E., Bean, L. J. H., Rosser, T. C., Bose, P. et al. (2014). Contribution of copy-number variation to Down syndrome-associated atrioventricular septal defects. *Genetics in medicine : official journal of the American College of Medical Genetics*, October), 1-7.

Ramachandran, D., Zeng, Z., Locke, A. E., Mulle, J. G., Bean, L. J. H., Rosser, T. C. et al. (2015). Genome-Wide Association Study of Down Syndrome-Associated Atrioventricular Septal Defects. *G3: Genes$_{Genomes}$Genetics*, *5*(10), 1961-1971.

Ripoll, C., Rivals, I., Ait Yahya-Graison, E., Dauphinot, L., Paly, E., Mircher, C. et al. (2012). Molecular signatures of cardiac defects in down syndrome lymphoblastoid cell lines suggest altered ciliome and hedgehog pathways. *PloS one*, *7*(8), e41616.

Sailani, M. R., Makrythanasis, P., Valsesia, A., Santoni, F. A., Deutsch, S., Popadin, K. et al. (2013). The complex SNP and CNV genetic architecture of the increased risk of congenital heart defects in Down syndrome. *Genome research*.

Stevens, J., Ermakov, A., Braganca, J., Hilton, H., Underhill, P., Bhattacharya, S. et al. (2010). Analysis of the asymmetrically expressed Ablim1 locus reveals existence of a lateral plate Nodal-independent left sided signal and an early, left-right independent role for nodal flow. *BMC Developmental Biology*, *10*(1), 1-12.

The, R. G. G. O. T. G. O. C. (2009). The Gene Ontology's Reference Genome Project: A Unified Framework for Functional Annotation across Species. *PLoS Comput Biol*, *5*(7), e1000431.

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R. et al. (2003). PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Research*, *13*(9), 2129-2141.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358-1370.

Wu, M. C., Kraft, P., Epstein, M. P., Taylor, D. M., Chanock, S. J., Hunter, D. J. et al. (2010). Powerful SNP-Set Analysis for Case-Control Genome-wide Association Studies. *The American Journal of Human Genetics*, *86*(6), 929-942.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D. et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, *498*(7453), 220-223.

Gene	SNVs in Gene	p-value
IFT46	6	0.003
DNAH1	75	0.003
TRAPPC9	21	0.007
CNGA4	9	0.007
GLI2	16	0.009
AHI1	17	0.010
KIF19	12	0.013
TOPORS	17	0.014
CBY1	2	0.019
SEPT2	6	0.023
NINL	11	0.023
PDE6D	1	0.024
MDM1	10	0.026
PACRG	2	0.027
RPGRIP1L	23	0.028
TTC8	8	0.030
CEP89	8	0.030
TPPP2	4	0.032
NPHP1	7	0.037
WDPCP	6	0.037
SMO	17	0.038
HSPB11	3	0.043
DYX1C1	8	0.045
PCM1	25	0.046

Table 3.2: Top SKAT-O associated van Dam cilia genes.

Gene	SNVs in Gene	p-value	Gene	SNVs in Gene	p-value
GMPPA	5	0.0002	ATP9B	16	0.0284
UBQLN1	5	0.0011	TTC8	8	0.0292
KDELC2	9	0.0020	GPD2	11	0.0293
DBR1	7	0.0021	MTMR3	15	0.0294
LONP2	5	0.0028	TSGA10	8	0.0298
CAMK2D	4	0.0034	MYEF2	7	0.0299
MAD2L1BP	6	0.0042	NXN	5	0.0300
BMPER	11	0.0049	APEH	7	0.0300
INPP4A	11	0.0050	METTL6	1	0.0301
ARMC3	8	0.0053	ADAT2	3	0.0302
HIST1H4J	2	0.0054	UXS1	3	0.0304

TMBIM1	7	0.0058	SAE1	5	0.0308
ARHGAP29	11	0.0063	NBN	10	0.0310
MCRS1	3	0.0071	UBXN11	9	0.0312
PIGN	13	0.0072	TRAF5	6	0.0314
GLI2	16	0.0078	ACOT12	10	0.0344
KDELC1	3	0.0081	CH25H	6	0.0360
SRI	4	0.0083	SYDE1	3	0.0365
TUBA1B	4	0.0086	DUSP12	4	0.0366
LBR	9	0.0089	PIPOX	3	0.0366
COG4	11	0.0090	LATS2	6	0.0370
MRPL49	3	0.0096	RYK	5	0.0372
GPSM1	9	0.0106	NPHP1	7	0.0373
WNT16	9	0.0110	THOC1	4	0.0379
GPATCH1	13	0.0111	TFDP1	4	0.0384
XPO1	6	0.0117	SRRT	11	0.0393
PDCD6	1	0.0121	CDT1	9	0.0399
KCNJ1	5	0.0123	PROM1	9	0.0402
NPR2	9	0.0124	SUGT1	5	0.0405
ANAPC1	14	0.0124	DCUN1D3	2	0.0409
RCAN3	2	0.0128	ATP5C1	2	0.0415
GK5	2	0.0134	WDR48	9	0.0415
REXO4	2	0.0149	SMO	17	0.0420
CUL1	5	0.0150	LZIC	1	0.0422
GNAS	7	0.0163	POLK	8	0.0424
SPATA17	7	0.0178	SDR9C7	5	0.0427
KIF15	15	0.0190	CDC5L	6	0.0428
RNF146	3	0.0195	RPS3	6	0.0435
INPP5F	2	0.0206	ANKS1A	16	0.0435
FGFR1	10	0.0209	TXNDC9	3	0.0438
YIF1A	5	0.0209	DYSF	50	0.0439
RSAD2	3	0.0224	STYXL1	3	0.0442
TXLNA	2	0.0230	PSAP	6	0.0447
HDHD2	2	0.0242	CSTF3	4	0.0451
MAPK14	3	0.0254	RPL7	3	0.0452
ELOVL6	2	0.0260	SH3YL1	2	0.0453
KCNA6	2	0.0260	ATG10	6	0.0453
NCKAP1	9	0.0269	DYX1C1	8	0.0454
RPS15	3	0.0270	PLD3	10	0.0456
PACRG	2	0.0274	PCM1	25	0.0465
MRPS10	2	0.0274	GAS2L1	1	0.0480
ZDHHC20	6	0.0277	SOD1	1	0.0484
			RAB3GAP2	13	0.0485

Table 3.3: Top associated Leroux Ciliome genes by SKAT-O.

Annotation Set	Term	Total Genes in Pathway	Case Genes in List	Case Genes Expected	Over/Under enrichment	Fold Enrichment	Bonferroni p-value
Panther Protein	actin family cytoskeletal protein (PC00041)	395	15	4.04	+	3.71	0.003
	non-motor actin binding protein (PC00165)	190	9	1.94	+	4.63	0.035
	hydrolase (PC00121)	1511	31	15.46	+	2	0.036
GO-slim Molecular	lipid transporter activity (GO:0005319)	112	8	1.15	+	6.98	0.004
	actin binding (GO:0003779)	176	9	1.8	+	5	0.017
	cytoskeletal protein binding (GO:0008092)	251	10	2.57	+	3.89	0.051
GO-slim Cellular	actin cytoskeleton (GO:0015629)	345	11	3.53	+	3.12	0.045
GO-slim Biological	developmental process (GO:0032502)	2456	45	25.13	+	1.79	0.017

	anatomical structure morphogenesis (GO:0009653)	596	17	6.1	+	2.79	0.035

Table 4.3: Pathway analysis results of 215 genes nominally associated with DS+AVSD case individuals from PANTHER overrepresentation testing. Listed are all ontology terms with Bonferroni-corrected p-values ≤0.1.

Annotation Set	Term	Total Genes in Pathway	Case Genes in List	Case Genes Expected	Over/Under enrichment	Fold Enrichment	Bonferroni p-value
Panther Protein	hydrolase (PC00121)	1511	97	63.59	+	1.53	0.006
	transporter (PC00227)	1020	70	42.93	+	1.63	0.012
GO-slim Molecular	transporter activity (GO:0005215)	1102	76	46.38	+	1.64	0.004
	transmembrane transporter activity (GO:0022857)	1010	70	42.51	+	1.65	0.007
	hydrolase activity (GO:0016787)	2205	128	92.8	+	1.38	0.023
	catalytic activity (GO:0003824)	5209	265	219.23	+	1.21	0.045
GO-slim Cellular	cell part (GO:0044464)	3090	173	130.05	+	1.33	0.002
GO-slim Biological	cellular component movement (GO:0006928)	476	44	20.03	+	2.2	4.00E-04

	cellular process (GO:0009987)	6708	338	282.32	+	1.2	0.009
	localization (GO:0051179)	2607	150	109.72	+	1.37	0.011
	transport (GO:0006810)	2473	142	104.08	+	1.36	0.020

Table 3.5: Pathway analysis results of 890 genes nominally associated with DS+NH control individuals from PANTHER overrepresentation testing. Listed are all ontology terms with Bonferroni-corrected p-values ≤ 0.1.

Position	chr2:219501527	chr2:219501550	chr2:219502393	chr2:219505470	chr2:219506313
SNP ID	rs34873891	rs41272703	rs150386940	rs146215853	rs1046474
Minor Allele	T	T	G	T	G
Major Allele	C	C	A	C	C
AA change	Ser -> Pro	Nonsyn.	Nonsyn.	Nonsyn.	Nonsyn.
cadd	30	24	20	26.5	29
phastCons	0.823	0.984	0.988	0.961	0.996
phyloP	0.709	0.945	0.945	0.945	0.945
MAF Cases	0.000	0.010	0.006	0.002	0.184
MAF Controls	0.004	0.046	0.019	0.008	0.135

Table 3.5: Variant statistics for the most associated SKAT-O gene, *GMPPA,* from the Leroux Ciliome list.

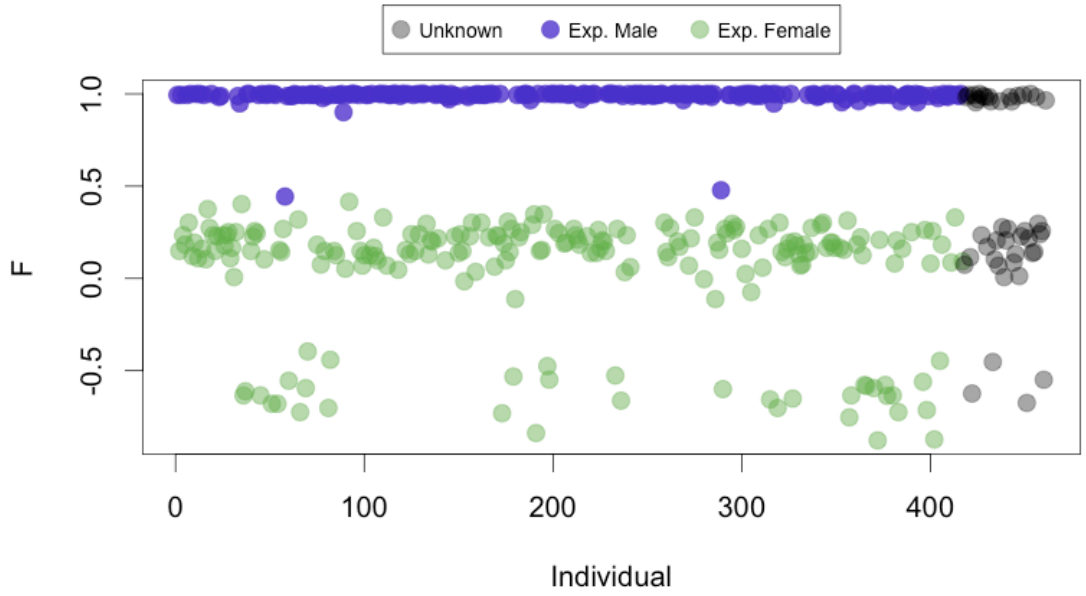


Figure 3.1: F statistics for heterozygosity estimates were computed for the X chromosome. Males are expected to have F = 1 indicating complete homozygosity. Colors indicated reported sex (blue = male, green = female, black = unknown). Two individuals thought to be male had F values predicting they were females. As this indicates either sample contamination or sample mix-up, these two individuals were removed. Females with F ≃ -0.5 indicates higher levels of heterozygosity, which is expected as these individuals are African Americans.

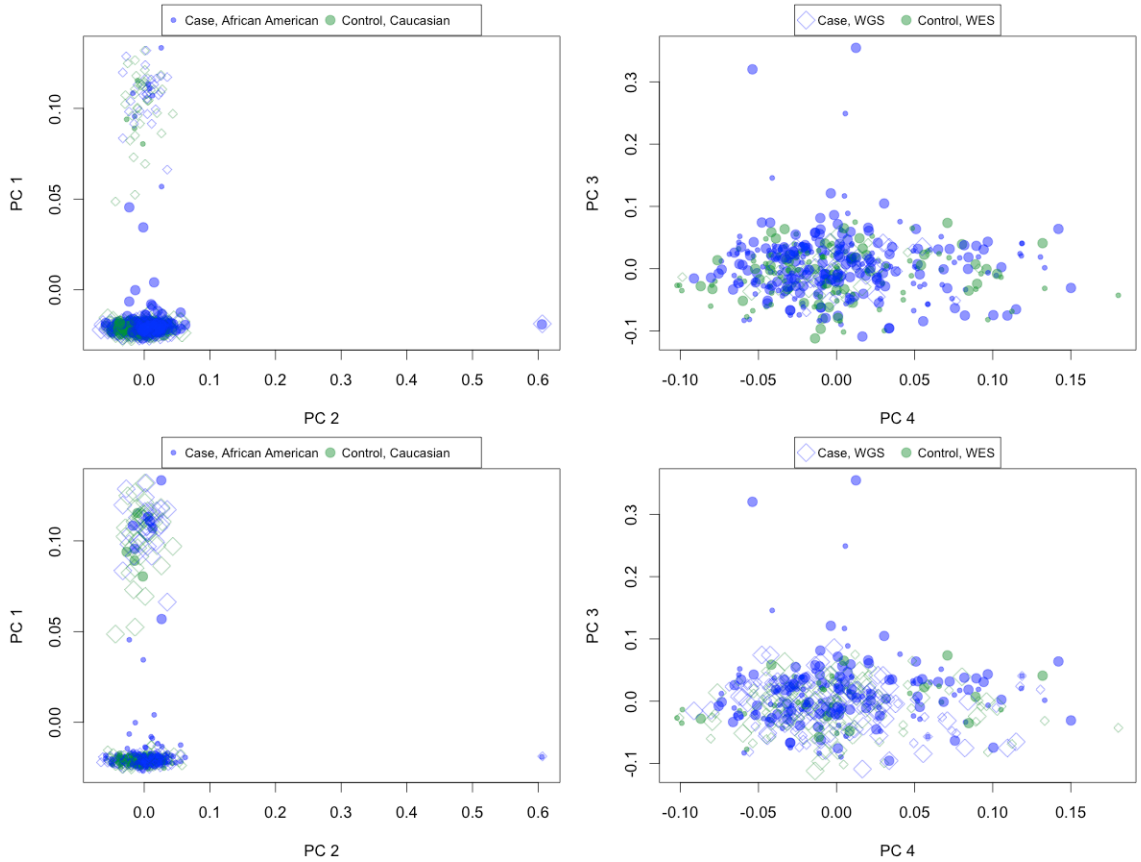


Figure 3.2: Principal components analysis round one properly segregates self-reported African Americans from Caucasians. DS+AVSD cases are in purple and DS+NH controls are in green. African Americans are small points and Caucasians are large points. Unfilled diamonds are samples from whole genome sequencing and filled circles are samples from whole exome sequencing.

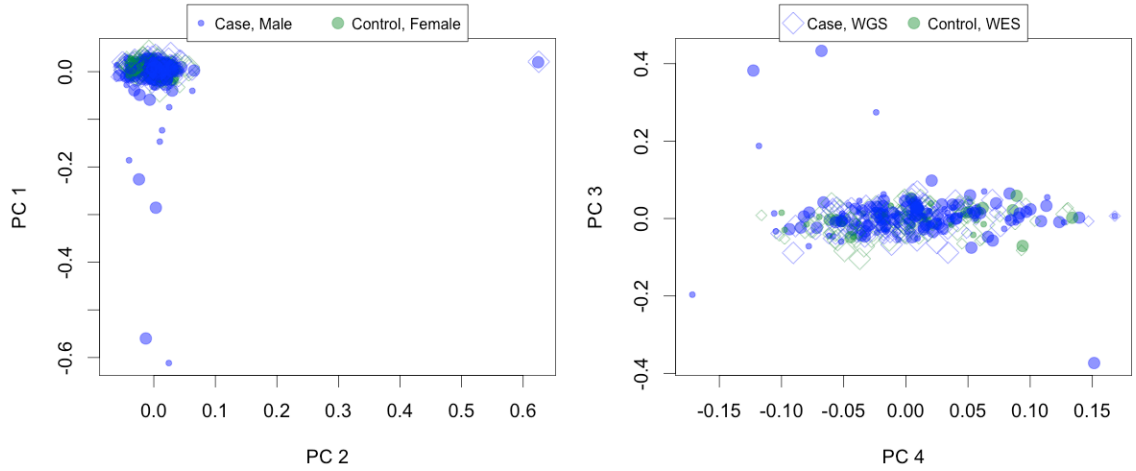


Figure 3.3: Round two of principal component analysis indicates nine DS+AVSD case outlier samples at PC1 < -0.1 and PC2 > 0.1 who were removed from further analyses. DS+AVSD cases are in purple and DS+NH controls are in green. Males are small points and females are large points. Unfilled diamonds are samples from whole genome sequencing and filled circles are samples from whole exome sequencing.

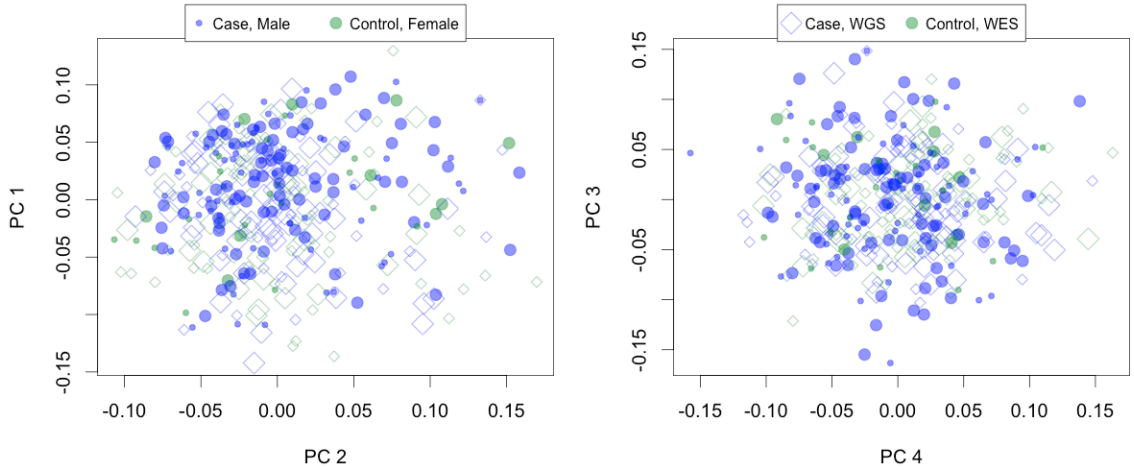


Figure 3.4: Round three of principal component analysis indicates no major population outliers. All of these individuals were retained for genomic analyses. DS+AVSD cases are in purple and DS+NH controls are in green. Males are small points and females are large points. Unfilled diamonds are samples from whole genome sequencing and filled circles are samples from whole exome sequencing.

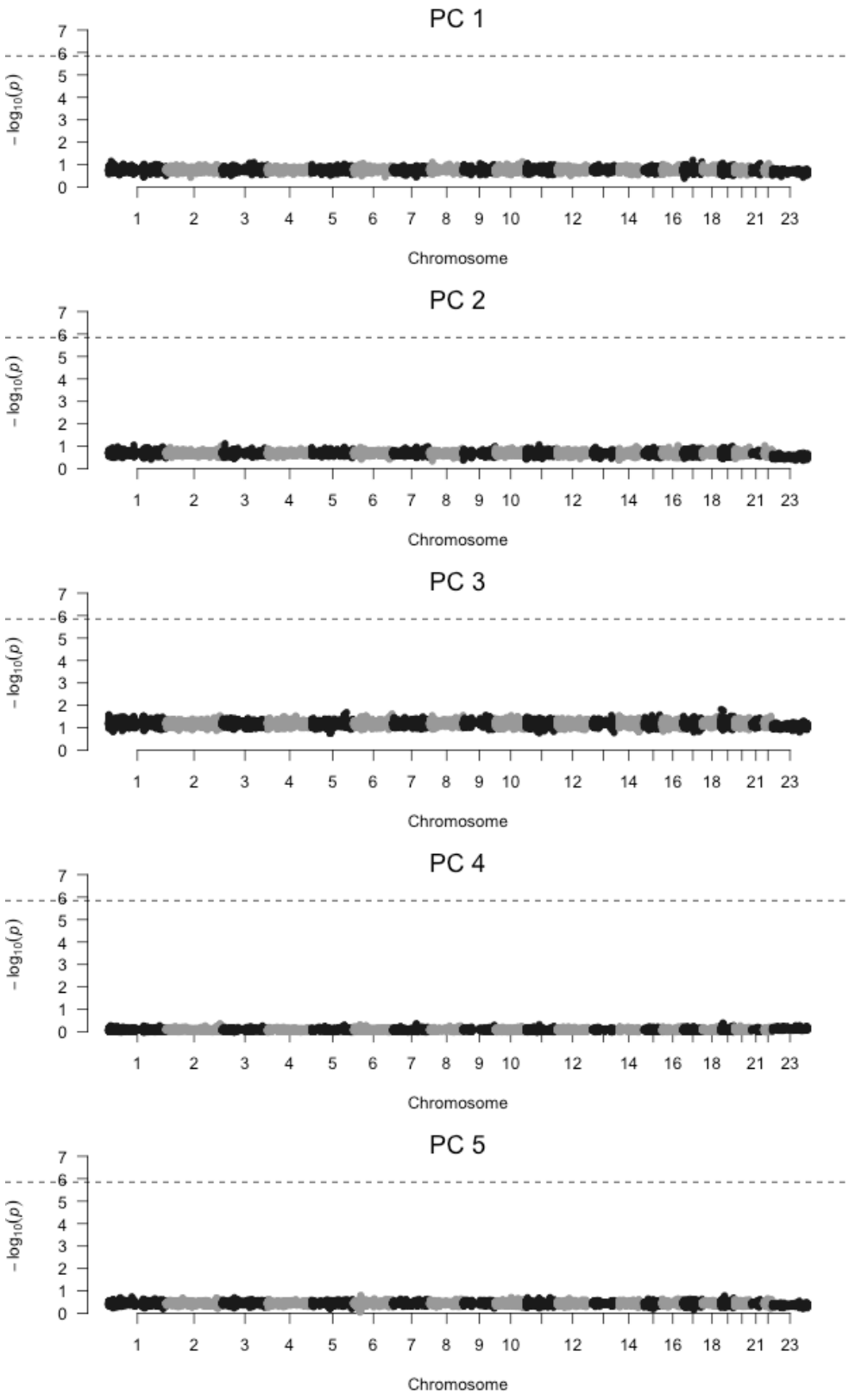

PC 1
PC 2
PC 3
PC 4
PC 5
$-\log_{10}(p)$
0 1 2 3 4 5 6 7
1 2 3 4 5 6 7 8 9 10 12 14 16 18 21 23
Chromosome

Figure 3.5: Manhattan plots of first five principal components as covariates to individuals SNP modeling by logistic regression. Dashed line is genome-wide significance level. No PC covariate is significant for any SNP.

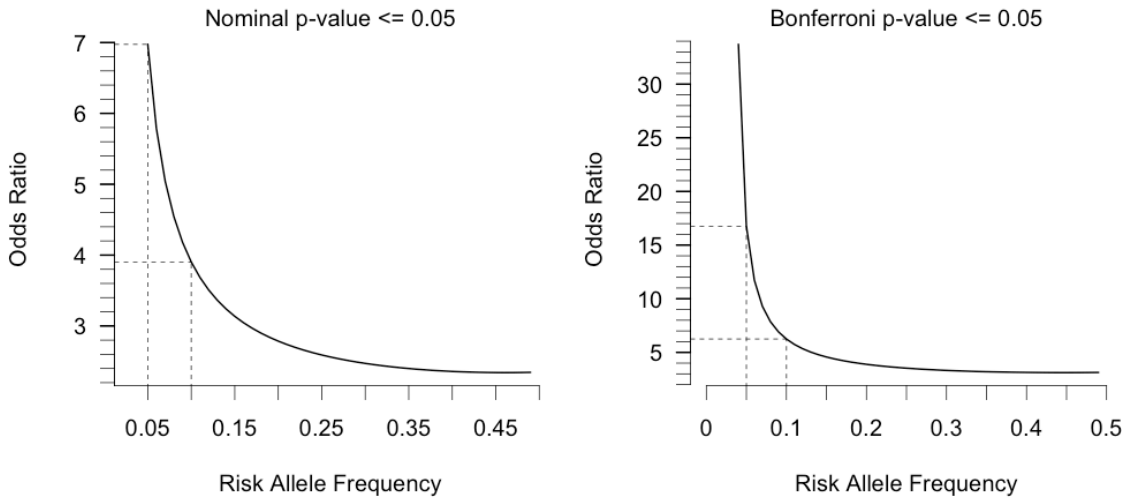


Figure 3.6: 80% power analysis for an additive model shows our ability to detect AVSD-associated variants with minor allele frequency of 0.05 with an odds ratio of ~17 and variants with MAF = 0.1 and OR of ~6 with a type 1 error of 0.05, after Bonferroni correction.

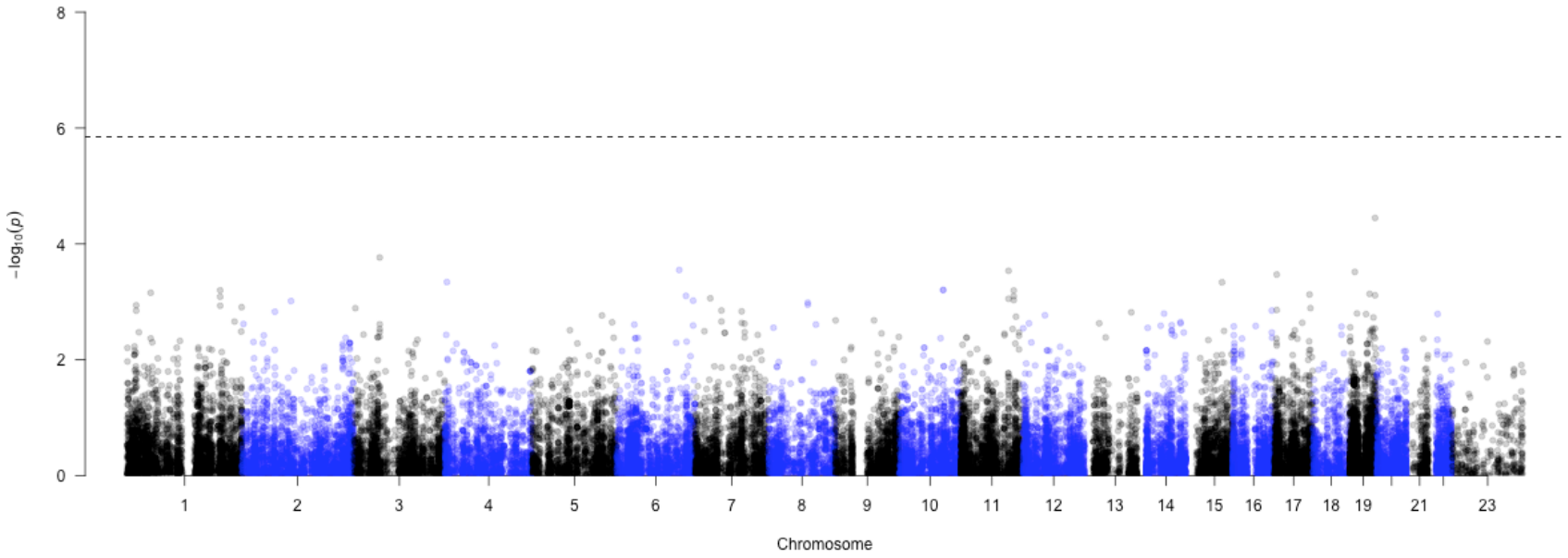


Figure 3.7: Manhattan plot of GWAS results from 35,071 common SNPs. No SNP reached genome-wide (dashed line) or near genome-wide statistical significance.

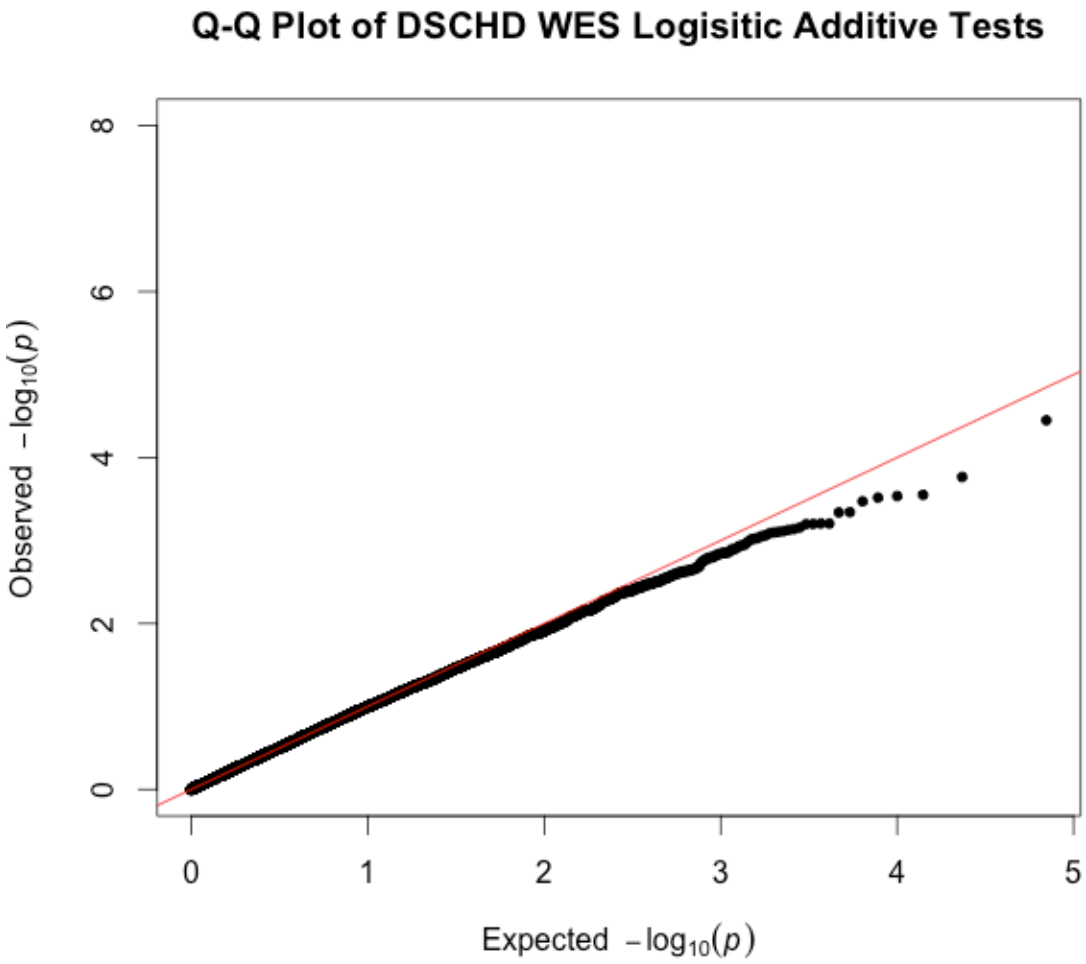


Figure 3.8: Quantile-quantile plot demonstrates that the associated SNP data follow the expectation, indicating no population stratification.

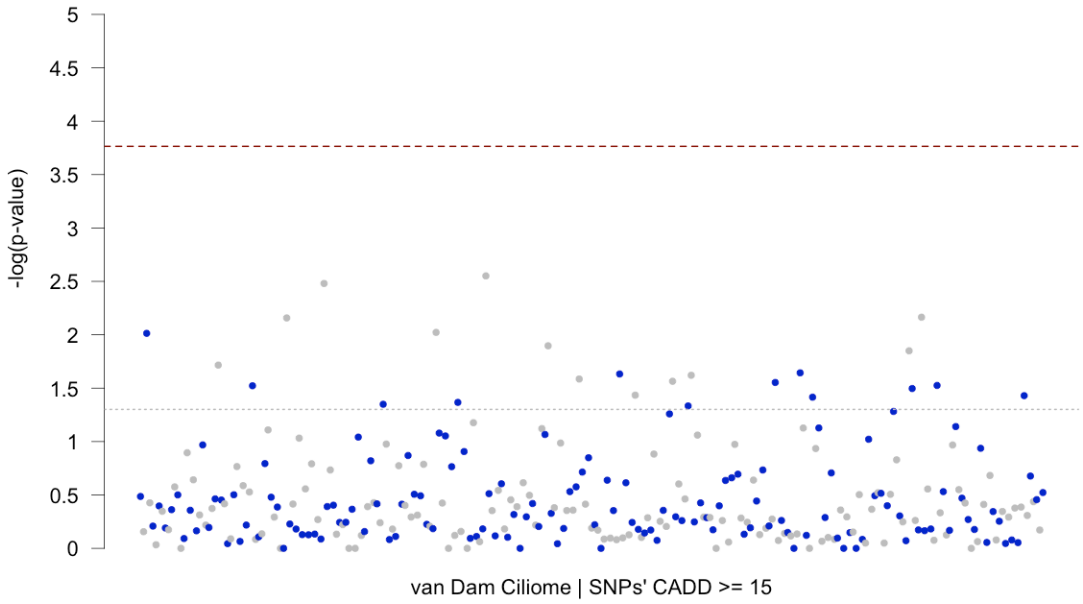


Figure 4.9: Manhattan plot of SKAT-O results of van Dam Ciliome Genes. No gene reached Bonferroni-corrected significant association levels (red dashed line) though many passed nominal significance thresholds (gray dotted line).

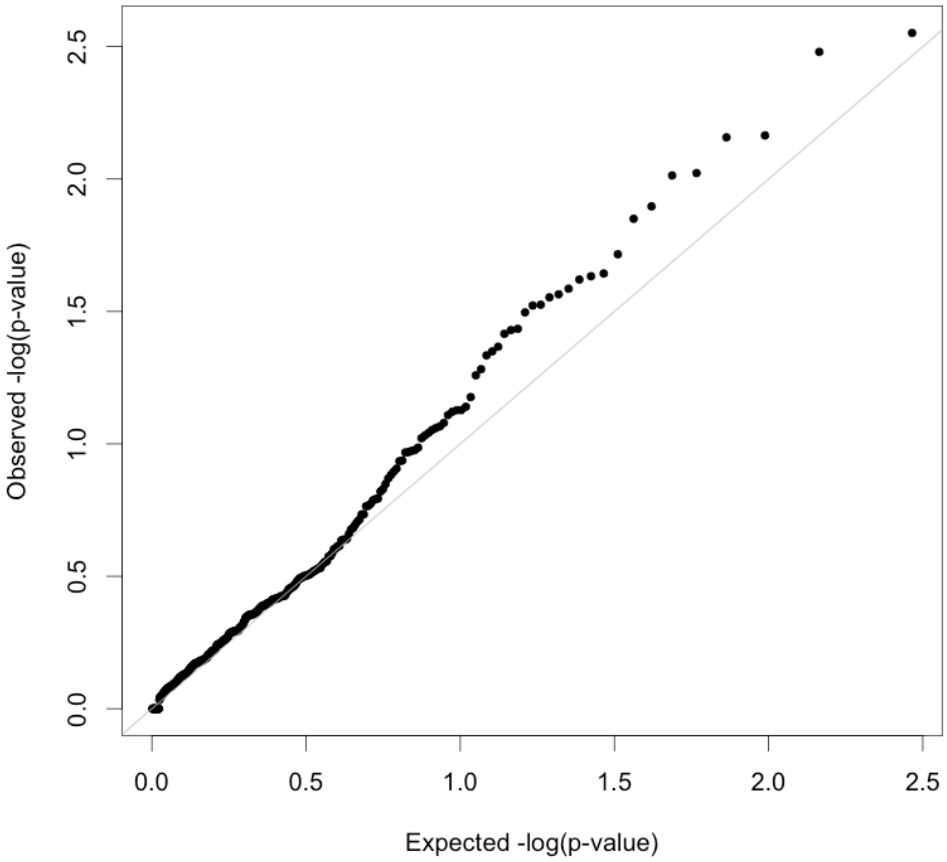


Figure 3.10: Quantile-quantile plot of van Dam Ciliome genes analyzed by SKAT-O. Many of these genes depart above the expected p-values indicating possible true association with AVSD in DS.

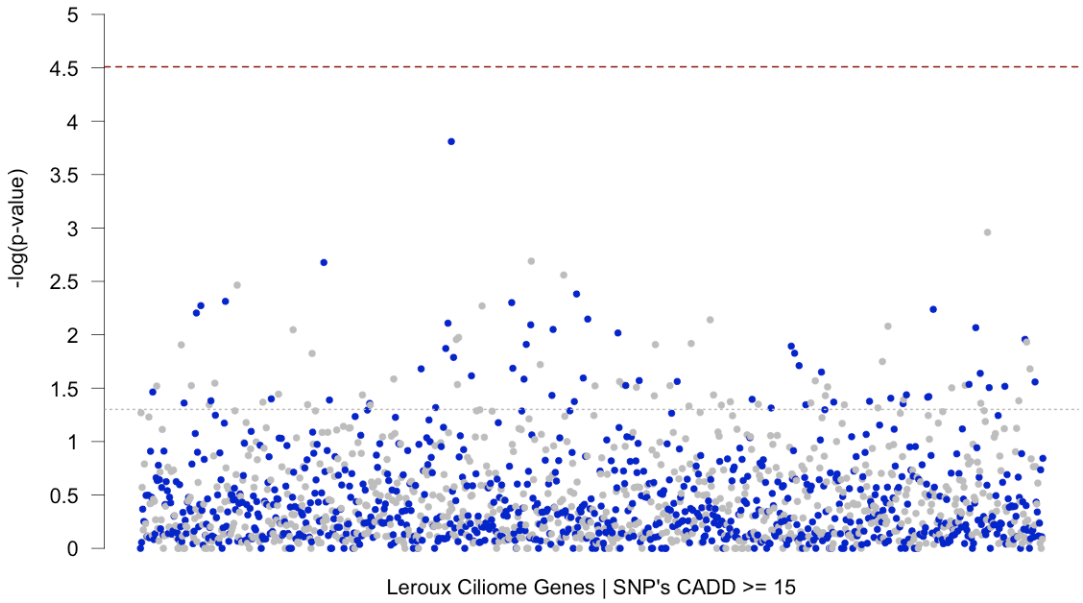


Figure 3.11: Manhattan plot of SKAT-O results of Leroux Ciliome Genes. No gene reached Bonferroni-corrected significant association levels (red dashed line) though many passed nominal significance thresholds (gray dotted line).

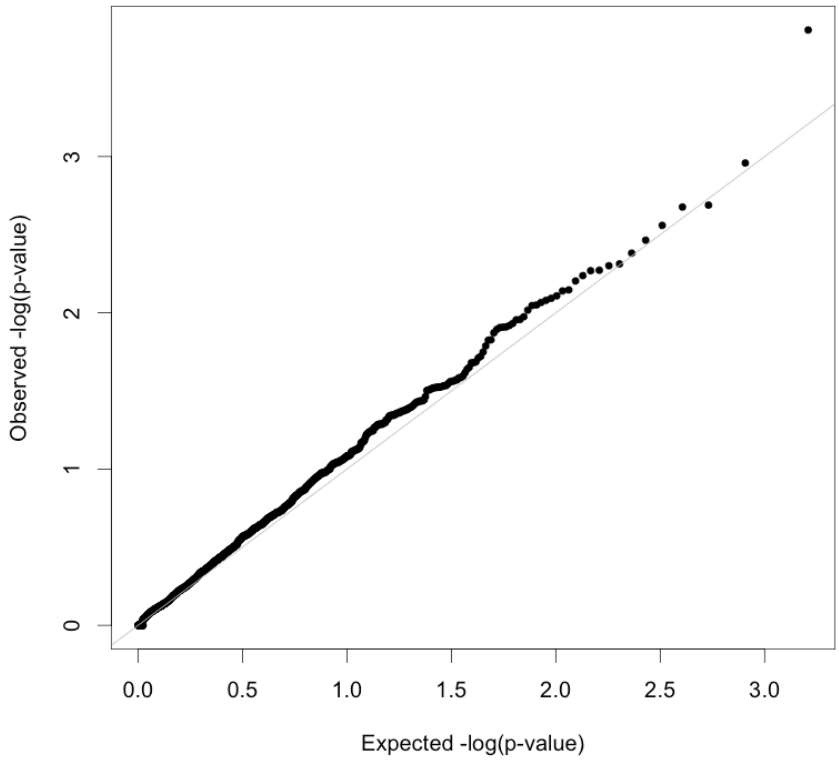
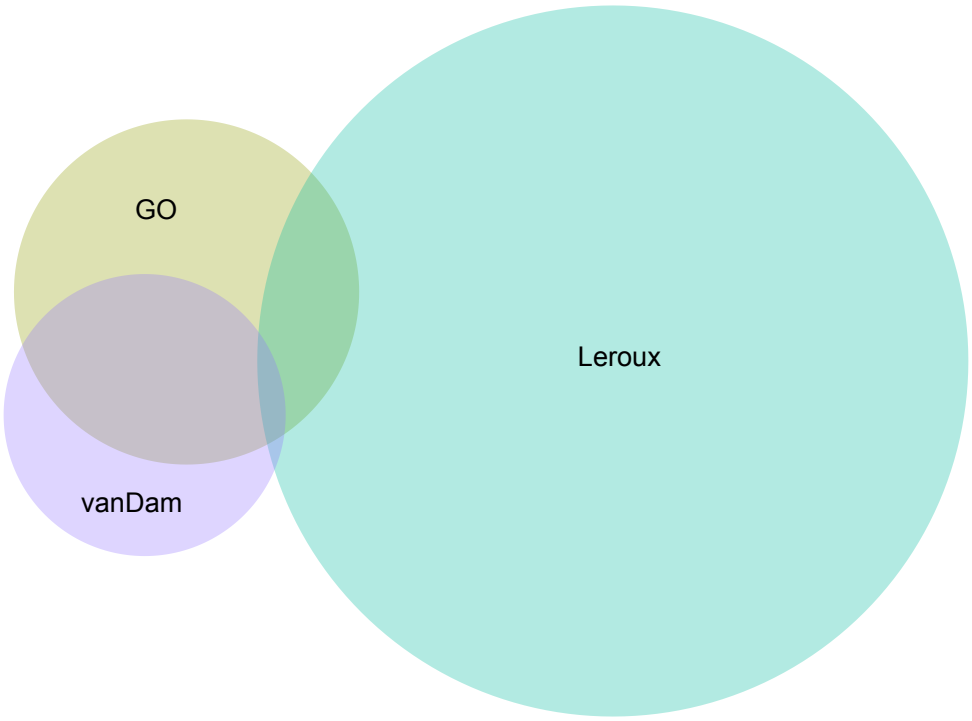


Figure 3.12: Quantile-quantile plot of Leroux Ciliome genes analyzed by SKAT-O. Many of these genes minimally depart above the expected p-values indicating possible true association with AVSD in DS.



	GO	Leroux	vanDam
GO	149	-	-
Leroux	116	1587	-
vanDam	165	125	91

Figure 3.13: Overlap between Leroux and vanDam cilia gene lists used for SKAT-O analysis. Also shown is the overlap between these gene lists and the Gene Ontology cilia list GO:0005929.

DISCUSSION

Conclusion and Future Directions

Millions of people are born with congenital heart defects and incur substantial costs, not only monetarily, but also in time spent seeking care, risk of sequelae and in the intangible psychological burden of having a chronic disease. Understanding the underpinning genetic contribution to CHDs began simultaneously by mapping breakpoints of syndromes caused by chromosomal aberrations like Holt-Oram syndrome as well as family-based linkage mapping of highly penetrant, rare mutations (Li et al., 1997; Scheffield et al., 1997; Green et al., 2000; Rupp et al., 2002; Robinson et al., 2003). These strategies were fruitful and brought the field a collection of genes required for heart development.

As technology progressed, new models could be tested. The common-disease common-variant hypothesis reasoned that common diseases not segregating in an obvious familial manor could be due to variants of relatively common frequency of moderate effect—penetrant in most, but not all. These effects could be found by performing genome-wide association studies using microarrays that could detect common variants, and thus resolve genotypes linked to disease. This hypothesis turned out to be largely false, including for CHDs, as was predicted by parsimonious classic population genetics theory that any moderately deleterious allele could never rise to a frequency common enough to be detected by obtainable cohort sizes (Pritchard, 2001; Hu et al., 2012; Cordell et al., 2012). Moving past the common-

disease common-variant hypothesis, we now expect that most common diseases without known genetic cause are heterogeneous—caused by many different rare variants of large effect and common variants of very small effect acting epistatically. This model can be tested by uncovering the entire gamut of variation via whole genome sequencing analysis. The field is currently finding success here in discovering a plethora of genes involved in CHDs and illuminating the genetic networks required for a functioning heart (Zaidi et al., 2013; Al Turki et al., 2014; Priest et al., 2016).

In Chapter 2, I have tested the hypothesis that common copy number variants (CNVs) on chromosome 21 cause Down syndrome (DS)-associated atrioventricular septal defects (AVSD). I measured small copy number variants on chromosome 21 in a large cohort of individuals with DS with and without AVSD. We reasoned that the 2,000-fold increased risk for AVSD that is incurred by DS individuals could be a result of either: 1. common deletions reducing AVSD-loci back to a disomic state and protect these people with DS from AVSD, or 2. common duplications could increase AVSD-loci over a tolerable threshold of three copies and cause AVSD. This is not that case at least for CNVs larger than 1.7 kb that were detectable by my methods. As others have seen, I reported a suggestive effect of cilia genes playing a role in AVSD. Pathway analysis of genes intersected by CNVs indicated the involvement of vesicle trafficking defects.

To test for the combined effects of rare and common variants, in Chapter 3, I analyzed merged single nucleotide variant (SNV) data from whole exome and whole sequencing data. I tested for the association of epistatic effects within cilia genes by

SKAT and again found a suggestion that aberrant cilia genes are involved in DS-associated AVSD. Pathway analysis of the sequencing data revealed a significant enrichment of rare variants in DS+AVSD cases in genes involved with actin remodeling, thus providing a novel candidate pathway for AVSD.

As an explanation for the incomplete penetrance of genes associated with CHDs, recent evidence from mouse models with partial synteny to Hsa21 suggests unknown environmental variance. Lana-Elola et al. (2016) used chromosome engineering to build seven different mouse models trisomic for sections of Mmu16, each with increasing resolution of Hsa21 syntenic regions from 23 Mb to 1.5 Mb. Mice bred with the largest trisomic region successfully modeled similar rates of penetrance of CHDs as in humans with DS at 61.5%, with a similar distribution of varied CHD types. In these mouse models, genetic modifiers cannot easily explain this incomplete penetrance as all mice were inbred and isogenic. Their mice with smaller Hsa21 syntenic trisomies within the largest 23 Mb region failed to produce similar rates of CHD, though most did show significantly increased rates. That two broods of mice, each with different, non-overlapping Hsa21 syntenic trisomic regions, can both increase risk of CHDs highlights the heterogeneity in proper heart development.

Other lines of evidence suggest that environmental factors contribute to increased risk of CHD. Maternal smoking, particularly during the first trimester, contributes to elevated risk of right ventricular outflow tract obstructions and septal defects including atrial, ventricular, and atrioventricular. This risk varies with increasing use of cigarettes (Malik et al., 2008). In mice haploinsufficient for the

CHD-associated transcription factor *NKX2-5*, the risk of ventricular septal defects increases with maternal age and can be attenuated with exercise provided by *ad libitum* exercise on a running wheel during three months before conception (Schulkey et al., 2015). Maternal folic acid supplementation is associated with DS-associated AVSD where individuals with DS+AVSD are 1.7 times more likely to be born to mothers without folic acid supplementation (Bean et al., 2011). However, we should recognize that the heterogeneity found in the genetic component of CHD is likely to be present in the environmental component. As potential environmental factors are infinitely vaster than genetic contributions, we can expect that genomics will drive the resolution of heart development's underlying biology.

Moving forward, genetic studies of CHD in the DS population, as well as nonsyndromic CHDs, should be designed knowing that there is considerable genetic heterogeneity. It is clear that, while DS alone increases the risk for AVSD 2,000 fold, it alone is not sufficient and no single common variant, SNV or CNV, contributes a large effect. Thus we will need larger cohorts of individuals with DS with and without CHDs to find susceptibility loci of measurable effects. As these cohorts continue to grow, efforts should focus on exome and whole genome sequencing approaches that identify rare variants, whose effects can be tested for burdening candidate genetic pathways of cardiogenesis. Finally, environmental factors need greater consideration and resources should be prioritized to gather broad epidemiological data and linking it to genomic resources.

REFERENCES

Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M.-P., Lindsay, S. et al.

(2014). Rare Variants in NR2F2 Cause Congenital Heart Defects in Humans. *American journal of human genetics*, *94*(4), 574-585.

Bean, L. J. H., Allen, E. G., Tinker, S. W., Hollis, N. D., Locke, A. E., Druschel, C. et al. (2011). Lack of maternal folic acid supplementation is associated with heart defects in Down syndrome: a report from the National Down Syndrome Project. *Birth defects research. Part A, Clinical and molecular teratology*, *91*(10), 885-893.

Cordell, H. J., Bentham, J., Topf, A., Zelenika, D., Heath, S., Mamasoula, C. et al. (2013). Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nature genetics*, *45*(7), 822-824.

Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., Lin, Y. et al. (2013). A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nature genetics*, *45*(7), 818-821.

Lana-elola, E., Watson-scales, S., Slender, A., Gibbins, D., Martineau, A., Douglas, C. et al. (2016). Genetic dissection of Down syndrome- associated congenital heart defects using a new mouse mapping panel. *1880*, 1-20.

Malik, S., Cleves, M. A., Honein, M. A., Romitti, P. A., Botto, L. D., Yang, S. et al. (2008). Maternal smoking and congenital heart defects. *Pediatrics*, *121*(4), e810-6.

Priest, J. R., Osoegawa, K., Mohammed, N., Nanda, V., Kundu, R., Schultz, K. et al. (2016). De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. *PLoS Genet*, *12*(4), e1005963.

Pritchard, J. K. (2001). Are Rare Variants Responsible for Susceptibility to Complex Diseases. *The American Journal of Human Genetics*, *69*(1), 124-137.

Schulkey, C. E., Regmi, S. D., Magnan, R. A., Danzo, M. T., Luther, H., Hutchinson, A. K. et al. (2015). The maternal-age-associated risk of congenital heart disease is modifiable. *Nature*, *520*(7546), 230-233.

Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D. et al. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, *498*(7453), 220-223.