The Phenotyping and Genetic Basis of Cardiovascular Defects in Syndromic and Non-syndromic Individuals

By

Holly Corbitt

Maslen Lab

A DISSERTATION

Presented to
The Department of Molecular and Medical Genetics and the
Oregon Health & Science University
School of Medicine
In partial fulfillment, of
the requirements for the degree of

Doctor of Philosophy

December 2018

TABLE OF CONTENTS

Acknowledgements i

List of Abbreviations ii

Abstract 1-2

Introduction 2-12

References 12-15

Chapter I: Phenotyping Turner syndrome aortic disease: **16-34**
Allometric considerations when assessing aortic aneurysms in Turner syndrome: Implications for activity recommendations and medical decision-making

- Abstract 17
- Introduction 17-19
- Methods 19-21
- Results 21-24
- Discussion 24-27
- Supplemental 28-32
- References 33-34

Chapter II: Turner syndromic aortic disease sequencing study: **35-65**
TIMP3 and *TIMP1* are risk genes for bicuspid aortic valve and aortopathy in Turner syndrome

- Abstract 36
- Introduction 36-39
- Results 40-47
- Discussion 47-49
- Materials & Methods 49-57
- Supplemental 58-63
- References 64-65

Chapter IIb: In silico and functional studies of TIMPs and MMPs **66-86**

Chapter IIc: *CHADL* is associated with TAA **87-94**

Chapter III: Down syndrome AVSD sequencing study: **95-111**
Rare deleterious variants in a primary cilium component, *CEP290*, is associated with atrioventricular septal defects in Down syndrome

Abstract 96
Introduction 96-98
Methods 98-100
Results 100-102
Discussion 102-106
Supplemental 107-108
References 109-111

Chapter IV: Non-syndromic AVSD sequencing study: **112-148**
AVSD has heterogeneous genetic contributions that include genes important in cilia function and cardiovascular system development

Abstract 113
Introduction 113-115
Methods 116-121
Results 121-130
Discussion 130-134
Supplemental 135-145
References 146-148

Conclusions **149-150**

Acknowledgements

I would like to thank Dr. Cheryl Maslen for the continuous support during my time at OHSU. She afforded me the freedom and guidance to chase my interest, develop projects, and become an independent scientist. To my committee members who gave me their time, wisdom, and foresight on countless projects.

The funding I received was invaluable to my research and ability to train without hindrance: NIH Ruth L Kirschstein T32 training grant, The American Heart Association Predoctoral Fellowship, Friends of Doernbecher grant, The Roche/ARCS Foundation Scholar Award, and the Nicholas L. Tartar Research Fellowship.

Thanks to my best friend and co-worker Rebecca for always being patient, thoughtful, and up for a brainstorming session. My fellow classmates and chosen family who were always there to listen or to distract from the everyday grind.

In particular, I want to thank Dr. Elizabeth Cohen from Mt. Hood Community College for being the first person to give me a research opportunity and really catapult my interest in science.

Last but not least, I would like to thank my family: my parents and my brother for the unconditional love and support through this journey.

Abbreviations

AV	Atrioventricular
AVSD	Atrioventricular septal defect
BAV	Bicuspid aortic valve
BSA	Body surface area
CHD	Congenital heart defects
DS	Down syndrome
ECM	Extracellular matrix
MMP	Matrix metalloproteinase
PCA	Principal component analysis
SHF	Second heart field
SKAT	Sequence kernel association test
SHH	Sonic hedgehog
TS	Turner syndrome
TAA	Thoracic aortic aneurysm
TAD	Thoracic aortic dilation
TIMP	Tissue inhibitors of metalloproteinase
WES	Whole exome sequencing

Abstract

Infants with Turner syndrome (TS) and Down syndrome (DS) account for the vast majority of patients born with congenital heart defects (CHD), where those with TS have a 1 in 3 chance of having a bicuspid aortic valve (BAV) and aortopathy and those with DS have a 1 in 5 chance of having an atrioventricular septal defect (AVSD). These chromosomal aneuploidy's associations with CHDs and observations of Mendelian inheritance patterns of CHDs in families provides strong evidence of a genetic component. Although the chromosomal aneuploidies that give rise to these syndromic forms of CHDs are well established, the additional genetic factors that determine which subset will develop a CHD is not understood. Additionally, CHDs can also occur sporadically, with no family history, and without any other defect as non-syndromic CHDs. Therefore, we also studied AVSD in the euploid population utilizing a case-control gene-burden analysis and a rare variant analysis. The information presented here is a collection of whole exome sequencing studies pertaining to the investigation of the genetic basis of congenital heart defects and diseases in both syndromic (TS and DS) and non-syndromic populations.

For TS in particular, the diagnosis, understanding, and medical decision making related to aortopathy is still inconsistent within the cardiology field. Therefore, before studying the genetic underpinnings of this phenotype, it was important to first investigate which metrics should be used to determine the degree of aortopathy, specifically thoracic aortic dilation. We provide evidence to clinicians and the research community that TS specific z-scores should be used in assessing aortopathy in those with TS. These z-scores were then used in our studies to elucidate risk genes (*TIMP3*, *TIMP1*, and *CHADL*) and key pathways (regulation of the extracellular matrix and TGFß1 signaling) in TS related aortopathy.

For both syndromic and non-syndromic forms of AVSD, the second heart field and Sonic hedgehog (Shh) signaling through the primary cilia is the primary pathophysiological mechanism leading to the

defect. Thus, we took a hypothesis-driven approach to identify ciliome genes associated with AVSD in both the DS (syndromic) and the euploid (non-syndromic) genomic context. The DS AVSD case-control study revealed *CEP290*, which is a critical component of the primary cilia, as a significant candidate gene contributing to syndromic-AVSD. In addition, the non-syndromic studies identified several risk genes in the ciliome and Shh signaling pathways including: *FGF10*, *GLI3*, *GLI1*, *SMO*, and *PTCH1*.

Introduction

Overview

Chromosomal aneuploidies, or abnormal chromosomal number, accounts for a significant proportion of CHDs and cardiovascular diseases. Based on the observation that these syndromes alone are not sufficient to cause the CHDs and aortic diseases, we hypothesize a simple threshold model. In this threshold model, we propose that a chromosome gain or loss confers a significant increase risk of their respective CHDs, but additional genetic hit(s) are necessary to reach the disease threshold, Figure 1.

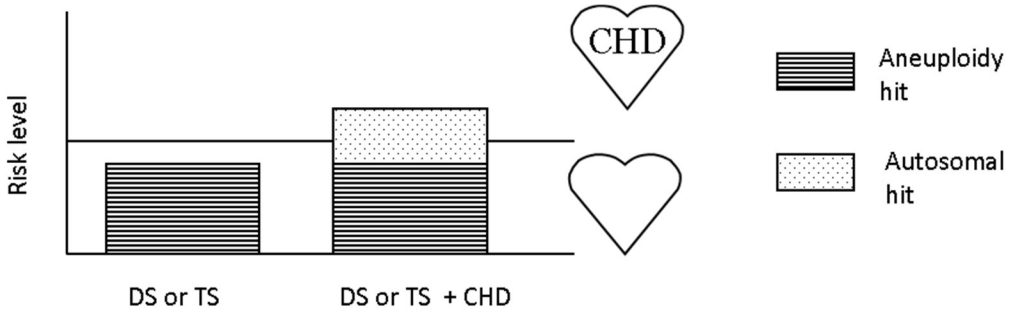


Figure 1: Theoretical threshold model of aneuploidy risk level and additional hits to reach CHD.

Although these defects occur in chromosomal aneuploidies at high rate, they also occur in the general euploid or non-syndromic population. Whether in association with a syndrome or alone, these defects and diseases are often fatal, require surgery, and have life-long impacts[3, 4]. Our goal is to not only understand the secondary hit(s) in the syndromic populations, but to also relate our genetic findings of AVSD, BAV, and TAD to the general population. We hypothesize that studying

syndromic and non-syndromic cardiovascular defects and diseases together may help us more fully understand the molecular pathogenesis which will lead to drug targets and improved treatments.

Down syndrome

DS is the most common chromosomal aneuploidy in humans with an incidence of 1 in every 700 live births, with over 400,000 individuals living with DS in the US[5]. It has long been known that DS is the result of trisomy 21 caused by in error in cell division called nondisjunction. This nondisjunction event primarily occurs in the oocyte in which 70% originate in meiosis I[6]. Because of the long tenure of oocyte development, the incidence of a nondisjunction event greatly increases with increased maternal age. Clinically, a maternal age of over 35 is considered high risk for having a child with DS[7]. Individuals with DS are at increased risk for a variety of phenotypes and disorders during their lives, including dysmorphic facial features, hearing loss, CHDs, intellectual disability, and leukemia to name a few[8]. The genetic causes of the defects associated with DS are poorly understood. Of interest to our study, the risk of an AVSD is 2000-fold higher in DS compared to the general non-syndromic population[1]. The partial prevalence of an AVSD, 1 in 5 infants with DS, supports that trisomy 21 acts as a significant risk factor, but other factors must also be contributing to disease onset, which we hypothesize are genetic factors.

Turner syndrome

TS is the result of a complete or partial loss of the second sex chromosome and notably is the only viable monosomy in humans[9]. While the frequency of TS is about 1 in every 2,500 live female births, it has been predicted that only 1% of fetuses with TS (45,X karyotype) survive to birth. This observation has led to the belief that the 1% of 45,X individuals that survive to term has some low level of mosaicism, although hard to detect through conventional methods[10]. TS is not related to increased maternal or paternal age and interestingly 80% of the single X chromosomes in a 45,X karyotype are of maternal origin, which argues against a maternal meiotic error[11]. Besides the 45,X

karyotype, there are numerous other karyotypes that result in TS, all of which can also have some level of mosaicism. A general breakdown of observed karyotypes is ~50% have monosomy X, ~48% have either a partial loss, rearrangement, or mosaicism of a second X chromosome, and ~ 2% have some Y chromosome material[10]. It is hypothesized that these variations in the status of the second sex chromosome can lead to varying phenotypes and severity of TS. These include short stature, webbed neck, premature ovarian failure, lymphedema, CHDs and cardiovascular disease, etc. Of these, the CHDs and related cardiovascular disease affecting the left side of the heart (including BAV and TAD) are recognized as the primary decrease of life expectancy in TS[12].

The genetics of congenital heart defects - non-syndromic and syndromic

CHDs are the main occurrence of birth defects in the world and account for most deaths in the first year of life, excluding infections, and are a significant component of pediatric cardiology[13]. Therefore, it is critical that we understand the origin and etiology of CHDs. Several studies looking at mouse knockouts and mutant lines with cardiac developments defects have identified signaling molecules, transcriptional regulators, and components of the extracellular matrix (ECM) that have proven to be critical for cardiac development. Additionally, several linkage studies of non-syndromic familial forms of CHDs have yielded genes associated with human CHDs, including *TBX5*, *NOTCH1*, *GATA4, GATA5, GATA6*, *NKX2-5*, *NR2F2*, and *TBX20*[14-18]. Because of these studies and other landmark epidemiological studies (Baltimore-Washington Infant Study of 1980 and a population based study in Denmark) it is widely accepted that CHDs are multifactorial and extremely heterogeneous, due to genetic susceptibility and inheritance[19-20]. Another key piece of evidence that CHDs have a strong genetic basis is that 25-40% of CHDs occur in association with genetic syndromes (macro or micro aneuploidies or single gene defects)[21]. While these syndromes have a high rate of CHD occurrence, they are complex in presentation and are not 100% penetrant nor do they affect 100% of the individuals. Of interest to the studies presented below are two of the most common aneuploidies in humans, Down syndrome (DS) and Turner syndrome (TS).

The genetics of AVSD in the context of DS is poorly understood. Our lab and others operate under the threshold hypothesis in which the dosage effect of trisomy 21 acts a large risk factor, but that additional risk factors must also be contributing. Our lab has supporting evidence for this threshold model using the Ts65Dn trisomy 21 mouse model. *CRELD1* was one of the first genes identified as a risk factor for AVSD[22]. Subsequent mouse studies showed that having a null allele of *CRELD1* was benign on a euploid background, but had a high incidence rate of AVSD on the Ts65Dn background[23]. Investigating how genes on chromosome 21, in a trisomy context, contribute to AVSD is an ongoing field of study. This is further complicated by the imperfect Ts65Dn mouse model which only has two-thirds of genes orthologous to the human chromosome 21[24]. Despite this, a gene on the Ts65Dn chromosome, *JAM2*, has been shown to interact with *CRELD1* and lead to an increase in septal defects, highlighting the efficacy of using this model[23]. To expand on this, our approach uses a cases-control whole exome sequencing (WES) analysis of DS patients with and without an AVSD. We hypothesize that we will identify genetic modifiers, within the ciliome, outside of chromosome 21 that could act as additional hits needed to reach the disease threshold. Additionally, we have performed an expansion study of AVSD in the euploid population, non-syndromic AVSD, which occurs at a rate of 1 in 10,000 live births[25]. We hypothesize that we will identify deleterious variants in our non-syndromic study that highlight the molecular similarities in the pathogenesis of AVSD between DS and non-syndromic populations.

Comparable to DS, the genetics of aortopathy in TS are also poorly understood. Of specific interest to the studies presented hereafter are TS associated BAV and TAD. While BAV can occur as an isolated defect, in those with TS, BAV most often occurs in combination with TAD and is increased 50-100 fold compared to the general population[2]. Most of what is currently known about the epidemiology of BAV and TADs is from the non-syndromic population in which there is a male dominated prevalence (BAV of 3:1 [26] and TAD leading to aneurysms of 2:1[27]). This sex bias serves as additional evidence for the influence of the sex chromosome genetics on the presence and

progression of aortic disease and suggests that the lack of a second X chromosome predisposes both males and Turner syndrome females to have BAV and TAA (Figure 2).

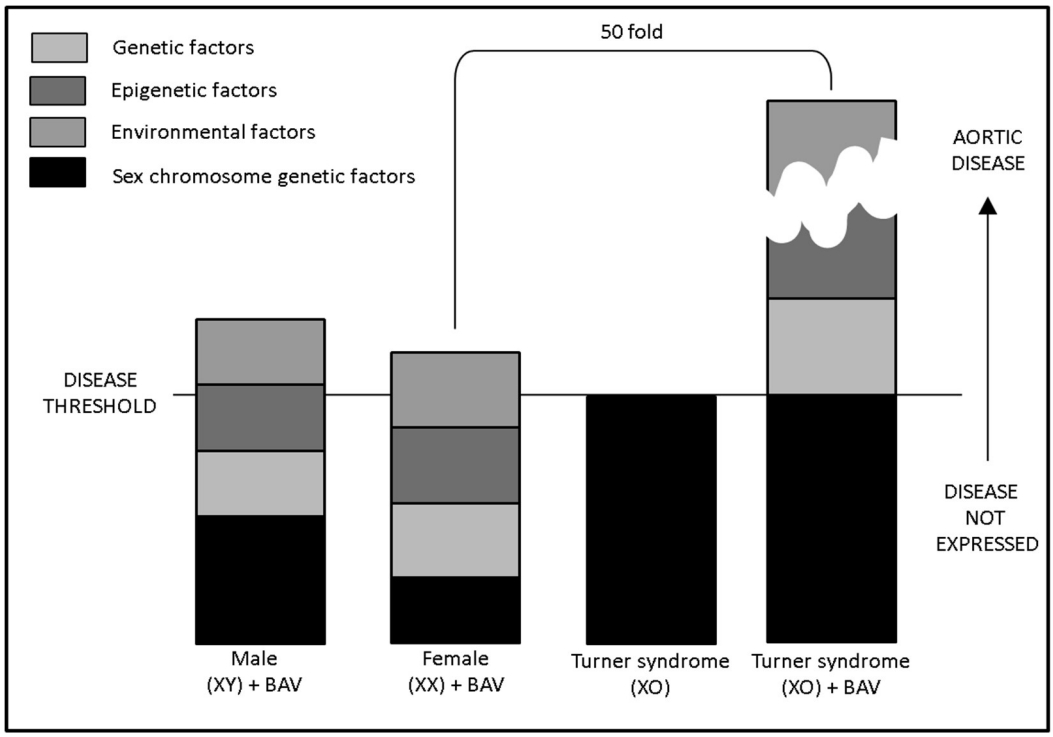


Figure 2: Threshold model to show the theoretical level of risk conferred by sex chromosome status and the additional risk factors that contribute to aortic disease onset. The relative risk for having a BAV is shown by comparing euploid males (46, XY), euploid females (46, XX), and Turner syndrome (45, X0). The solid horizontal line depicts the theoretical threshold for disease, with the black bars representing sex chromosome associated risk and the color bars representing other additional risk factors (genetic, epigenetic, and/or environmental). There is a ~50 fold increased risk for BAV in Turner syndrome compared to the risk of BAV in euploid females.

Family pedigree studies have established that BAV has a high heritability rate of 89%, indicating that the disease has a major genetic component[29-31]. Studies of familial BAV pedigrees have found genetic associations with *TGFßR1*, *TGFBß2*, *NOTCH1*, *NKX2.5*, *GATA5, ACTA2*, and *KCNJ2*[32]. The genetic heritability of TAD is less known, but recently mutations in *NOTCH1* and *TGFß2R* have been reported[33,34]. Because of the common neural crest lineage and shared developmental pathways of the aortic root and ascending aorta, it is hypothesize that genes involved in BAV are also implicated in BAV associated TAD. Supporting this, mutations in *NOTCH1* and *ACTA2* have also been described in BAV associated TAD[35,36]. Other syndromes associated with TADs leading to aneurysms, such as vascular Ehlers-Danlos syndrome (*COL3A1)*, Marfan syndrome (*FBN1)*, and

Loeys-Dietz syndrome (*TGFßR1* and *TGFßR2*), are autosomal dominant single gene disorders[37,38]. This has led the field to hypothesize that there is substantial polygenic and genetic heterogeneity leading to aortopathy. Because the lack of a second X chromosome acts as a highly sensitizing factor in TS and sex bias towards euploid males, we hypothesize that a region or gene will be identified on the X chromosome leading to a dosage effect and therefore allow autosomal variants with smaller effect sizes in the non-syndromic population to be highly penetrant in the TS population.

The molecular etiology of atrioventricular septal defects

An AVSD occurs when there is a failure of fusion between the mesenchymal cap of the primary atrial septum, the dorsal mesenchymal protrusion, and the atrioventricular (AV) endocardial cushions. The formation of the AV septum is one of the key steps in the organogenesis of the heart. A recent paradigm shift in the origin of the cells that form the AV septum was discovered in Dr. Moskowitz's lab. They discovered that progenitor cells (termed the second heart field or SHF) receive Sonic hedgehog (Shh) signals that trigger their migration and differentiation to populate the primary atrial septum and dorsal mesenchymal protrusion (Figure 3)[39]. The coordinated development of these two regions gives rise to the atrial and AV septal structures[40]. A conditional knockout of Shh signaling resulted in all of the pups having an AVSD, while their wild type litter mates had normal hearts. This exhibited that Shh signaling is both necessary and sufficient for the SHF development into the AV septa[39]. Thus, Shh signaling is a critical molecular event of AV septation, the defect in AVSD.

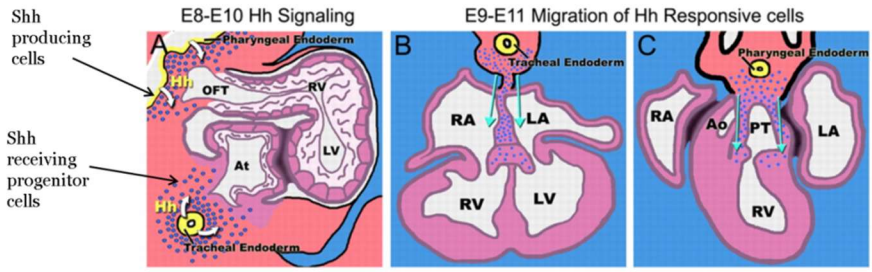


Figure 3: Shh signaling controls migration and differentiation, from Hoffmann et al., 2009.

Shh signaling is transduced exclusively through the primary cilium. The primary cilium is a microtubule based protrusion that plays multiple roles in chemical and mechanical sensing of the environment. As figure 4 depicts, the Shh signaling pathway is in its off state when no Shh ligand is present. Upon Shh paracrine signaling, Shh binds to Ptch1 and releases its inhibition on Smo. As Ptch1 exits the cilium, Smo enters and activates the Gli transcription factors into their active form and translocate to the nucleus to transcribe its downstream targets[41].

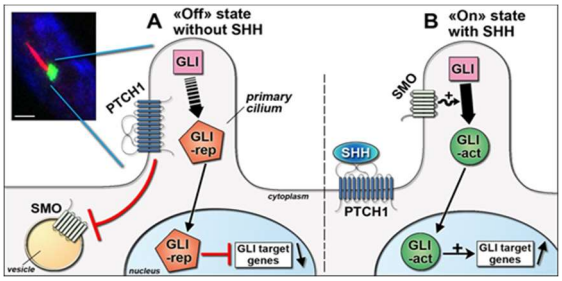


Figure 4: Shh signaling pathway through the primary cilium, from Mailleux et al., 2013.

The expansive networks of proteins that make up the primary cilium are known as the ciliome. Since Shh signaling is dependent on the integrity of the ciliome, it is rational to assume that proteins within the ciliome could also contribute to CHDs. In fact, the results of a large ENU recessive forward genetic screen established the ciliome's importance in the pathogenesis of CHDs. In this screen, 91 recessive mutations in 61 genes lead to a CHD. Of the 61 genes almost all of them played some role in the function of the ciliome[42]. Therefore, we hypothesized that we would identify variants in genes that were within the ciliome to be associated with AVSD in our WES case-control studies of both syndromic and non-syndromic cohorts.

The molecular etiology of bicuspid aortic valve and thoracic aortic dilation

The aorta is the main artery that transports blood out of the heart. The aortic valve guards the entrance to the aorta and regulates the directional pumping of blood out of the heart to the rest of the body. The BAV defect is when the aortic valve has two leaflets instead of the usual three leaflets and is more common than all other congenital heart defects combined[43]. Over time, a BAV commonly

causes obstructions to the blood flow out of the heart and can become leaky causing regurgitation of blood back into the heart. BAV often occurs in combination with an enlarged aorta, or TAD, which is the main clinical cause of concern because it can result in catastrophic aneurysm, aortic dissection, rupture, and death[43]. It has been observed that patients with BAV develop TAD leading to an aneurysm at least 10-15 years earlier than patients with a tricuspid valve[44].

While the genetics of BAV and associated aortopathy are largely unknown, animal models and molecular approaches have implicated a large number of genes, including those involved in transcription, components of the ECM, and components of proliferation and apoptosis[38,45-49]. These molecular and cellular events which result in BAV and associated TAD are multifactorial. However, ECM degradation (specifically, increased fragmentation of elastin and collagen) in the aortic media layer appears to be a consistent histopathological and biochemical feature of TAD[48,50-52]. Along with the ECM, the aortic media layer is primarily made up by vascular smooth muscle cells (VSMCs) and is the main structural layer in the aortic wall. In aneurysmal associated Vascular Ehlers-Danlos syndrome, over 70 different mutations in a collagen gene *COL3A1* have been identified, leading to decreased strength of the media layer. Also a part of the ECM is a microfibrillar network, mostly composed of fibrillin-1. Fibrillin-1 interacts with elastin, collagen, and the VSMCs to provide structural support and regulates growth factors and signaling molecules by sequestering them, such as TGFß1. In aneurysms associated with Marfan syndrome, the mutations in *FBN1* cause disorder to the fibrillin-1 network and the inability to effectively sequester TFGß1 leading to over activation, which propagates more ECM degradation[38,53-55]. Further supporting the molecular source of the ECM degradation in the pathogenesis of TADs, an increased expression of matrix metalloproteinases (MMPs, particularly MMP-2 and MMP-9) and an imbalance between MMP and their inhibitors (tissue inhibitors of metalloproteinases, TIMPs) has been demonstrated in the dilated aortic wall[49,56,57]. Another hallmark of histological analysis of thoracic aneurysms is the presence of inflammatory cells in the adventitia layer infiltrating the media layer. This recruitment and stimulation of T-lymphocytes

and macrophages stimulates the release of more MMPs and VSMC apoptosis[58-60]. A model describing these pathogenic events is presented below (Figure 5).

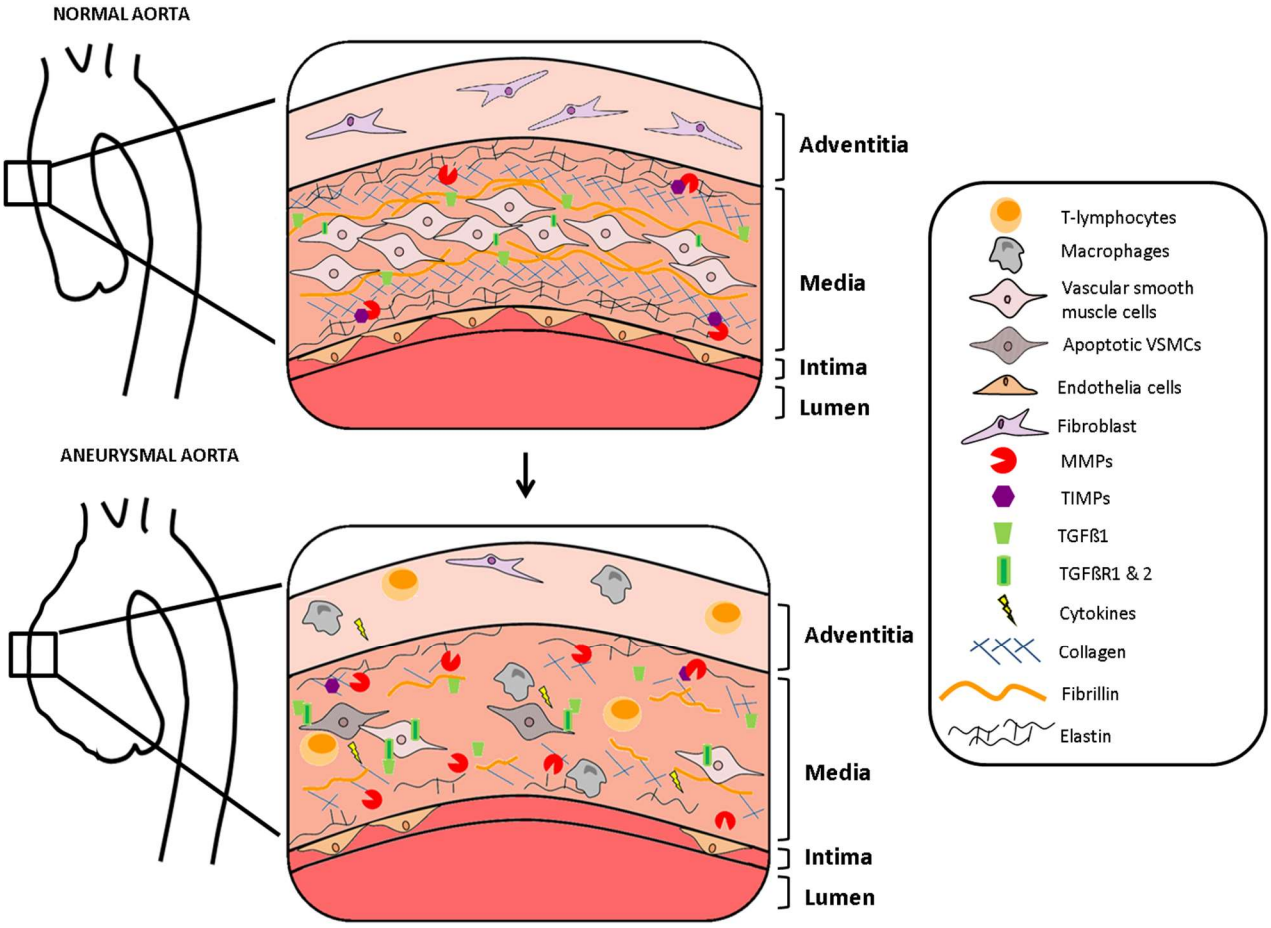


Figure 5: Model of thoracic aorta aneurysm (TAA) disease progression from the normal state. **Top)** In the normal state of the ascending aorta, the ECM in media layer is highly structured and densely surrounding VSMCs. There is little MMP activity because of their low expression and their inhibition by TIMPs. TFGß1 is present, but is sequestered by fibrillin-1. **Bottom)** In the aneurysmal state, the ECM is highly fragmented and sparse, leading to the release of TGFß1 and cytokines and to the apoptosis of VSMCs. An immune response is elicited to the damaged tissue by T-lymphocytes and macrophages which further degrade ECM components by expressing MMPs and further apoptosis. TAA is a progressive disease in which the initial pathogenic event is determined by the gene or process impacted and then feeds into a feedback cycle which propagates the disease further.

MMP's natural inhibitors, TIMPs, are well recognized for their role in ECM remodeling by controlling the activity of their respective MMPs by binding irreversibly to inhibit their proteolytic activity. The human genome has four genes encoding TIMPs (TIMP1, TIMP2, TIMP3, and TIMP4). This is of particular importance to the study of aortic disease in TS because *TIMP1* is located on the X

chromosome and polymorphically escapes from X-inactivation[61,62]. A complex feedback loop between the degradation of the ECM, TFGß signaling, MMP/TIMP imbalance, VSMC apoptosis, and inflammation play a role in the overall pathogenic events of aortopathy. In TS associated aortopathy, we hypothesize that it enters the disease cycle through the MMP/TIMP imbalance (Figure 6).

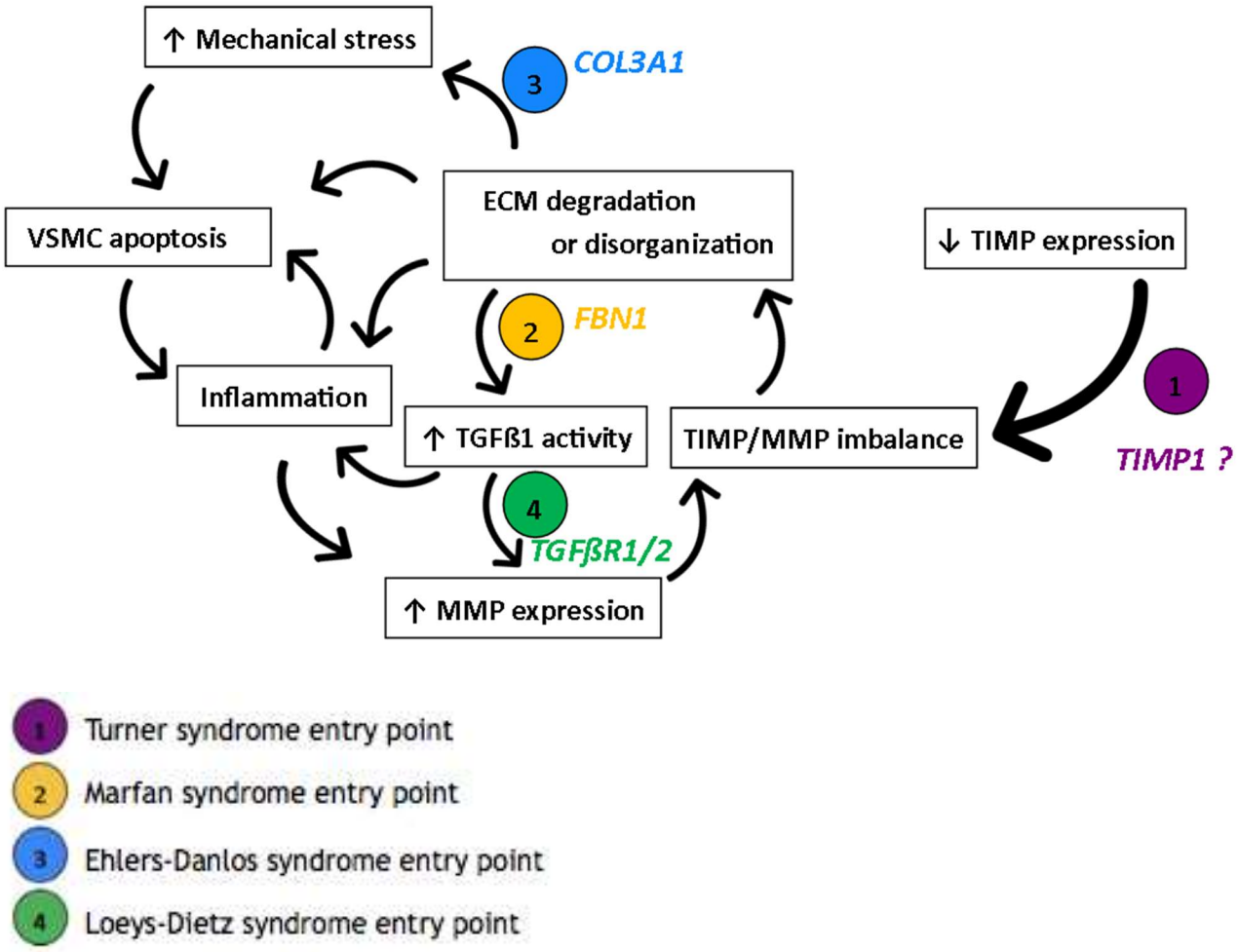


Figure 6: Feedback cycle of the pathogenic events in TAA. We propose that in Turner syndrome (TS) the disease cycle is entered through decreased TIMP expression leading to an imbalance in the TIMP/MMP ratio.

Phenotyping TAD in Turner syndrome

TAD is a progressive disease in which the aorta becomes more dilated over time. The growth progression is not uniform over time and the point of enlargement that results in an aneurysm, dissection, or rupture is not consistent between individuals. Therefore, clinicians depend on routine

monitoring and Z-score calculations, the number of standard deviations their aortic measurement is from the mean of a reference population, to predict who is at risk. These Z-scores are used to restrict activity or to opt for preemptive surgery to repair their aortas. These Z-score equations were created using regression models on based on aortic dimensions and body surface area (BSA) where a Z-score greater than 2 is considered at risk[63]. Historically, the Z-score calculations based off of the general, non-syndromic, population have also been used to predict dissection and rupture risk for the TS population. It was recognized, that due to the non-linear growth and characteristic short stature of those with TS, a new Z-score equation should be calculated based on those with TS as the reference population[64]. The analysis presented in this thesis, further explores the hypothesis that the new TS specific Z-score equations are more accurate and vastly different from the conclusions and recommendations clinicians would have reach from the general population Z-score calculator. Additionally, we thought it was critical to first understand how to phenotype TAD in the context of TS before using it as a variable in the case-control WES study.

1. Freeman, S.B. *et al.* Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. *Genet Med* 10, 173-80 (2008).
2. Weinsaft, J.W. *et al.* Aortic Dissection in Patients With Genetically Mediated Aneurysms: Incidence and Predictors in the GenTAC Registry. *J Am Coll Cardiol* 67, 2744-54 (2016).
3. Pradat, P., Francannet, C., Harris, J.A. & Robert, E. The epidemiology of cardiovascular defects, part I: a study based on data from three large registries of congenital malformations. Pediatric Cardiology 24, 195-221 (2003).
4. Harris, J.A., Francannet, C., Pradat, P. & Robert, E. The epidemiology of cardiovascular defects, part 2: a study based on data from three large registries of congenital malformations. Pediatric Cardiology 24, 222-35 (2003).
5. Antonarakis 2017 Down syndrome and the complexity of genome dosage imbalance.
6. Ghosh 2015 Genetic Etiology of Chromosome 21 Nondisjunction and Down syndrome Birth: Aberrant Recombination and Beyond.
7. Nemer 2012 Genetic Causes of Syndromic and Non-Syndromic Congenital Heart Disease
8. Korenberg, J.R. *et al.* Down syndrome phenotypes: the consequences of chromosomal imbalance. *Proceedings of the National Academy of Sciences of the United States of America* 91, 4997-5001 (1994).
9. Nielsen J, Wohlert M 1991 Chromosome abnormalities found among 34,910 newborn children: results from a 13-year incidence study in Arhus, Denmark. Hum Genet 87:81–83
10. Hook EB, Warburton DHum Genet. 1983; 64(1):24-7.The distribution of chromosomal genotypes associated with Turner's syndrome: livebirth prevalence rates and evidence for diminished fetal mortality and severity in genotypes associated with structural X abnormalities or mosaicism.
11. Zhong, Q., & Layman, L. C. (2012). Genetic Considerations in the Patient with Turner Syndrome—45,X with or without Mosaicism. Fertility and Sterility, 98(4), 775–779. http://doi.org/10.1016/j.fertnstert.2012.08.021

12. Schoemaker MJ, Swerdlow AJ, Higgins CD, Wright AF, Jacobs PA 2008 Mortality in women with Turner syndrome in Great Britain: a national cohort study. J Clin Endocrinol Metab 93:4735– 4742
13. Richards AA, Garg V. Genetics of Congenital Heart Disease. Current Cardiology Reviews. 2010;6(2):91-97. doi:10.2174/157340310791162703.
14. Schott JJ, Benson DW, Basson CT, et al. Congenital heart disease caused by mutations in the transcription factor NKX2-5. Science 1998;281:108–11.]
15. Garg V, Kathiriya IS, Barnes R, et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 2003; 424:443–7.
16. Basson CT, Bachinsky DR, Lin RC, et al. Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome [published correction appears in Nat Genet 1997;15:411]. Nat Genet 1997;15:30–5.
17. Garg V, Muth AN, Ransom JF, et al. Mutations in NOTCH1 cause aortic valve disease. Nature 2005; 437:270–4.
18. Kirk EP, Sunde M, Costa MW, et al. Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. Am J Hum Genet 2007;81:280–91.
19. Ferencz C, Rubin JD, Loffredo CA, Magee CM. Perspectives in Pediatric Cardiology. Vol. 4. Mount Kisco NY: Futura Publishing Co.Inc; 1993. The epidemiology of congenital heart disease, The Baltimore-Washington Infant Study (1981-1989)
20. Oyen N, Poulsen G, Boyd HA, Wohlfahrt J, Jensen PK, Melbye M. 2009. Recurrence of congenital heart defects in families. Circulation 120: 295–301
21. Bernstein D. In: Evaluation of the cardiovascular system. Behrman RE, Kliegman RM, Jenson HB, editors. Nelson Textbook of Pediatrics: hiladelphia, Saunders; 2004. pp. 1481–8.
22. Molecular genetics of atrioventricular septal defects. CL Maslen - Current opinion in cardiology, 2004 - europepmc.org
23. Li, H. et al. Genetic modifiers predisposing to congenital heart disease in the sensitized Down syndrome population. Circ Cardiovasc Genet 5, 301-8 (2012).
24. Down syndrome mouse models Ts65Dn, Ts1Cje, and Ms1Cje/Ts65Dn exhibit variable severity of cerebellar phenotypes LE Olson, RJ Roper, LL Baxter… - Developmental …, 2004 - Wiley Online Library
25. Hartman, R.J. *et al.* Descriptive study of nonsyndromic atrioventricular septal defects in the National Birth Defects Prevention Study, 1997-2005. *Am J Med Genet A* 155A, 555-64 (2011).
26. Giusti B, Sticchi E, De Cario R, Magi A, Nistri S and Pepe G (2017). Genetic Bases of Bicuspid Aortic Valve: The Contribution of Traditional and High-Throughput Sequencing Approaches on Research and Diagnosis. Front. Physiol. 8:612. doi: 10.3389/fphys.2017.00612
27. McClure, R.S. *et al.* Epidemiology and management of thoracic aortic dissections and thoracic aortic aneurysms in Ontario, Canada: A population-based study. *Journal of Thoracic and Cardiovascular Surgery* 155, 2254-2264 e4 (2018).
28. Cripe, L., Andelfinger, G., Martin, L. J., Shooner, K., and Benson, D. W. (2004). Bicuspid aortic valve is heritable. J. Am. Coll. Cardiol. 44, 138–143. doi: 10.1016/j.jacc.2004.03.050
29. Lewin, M. B., McBride, K. L., Pignatelli, R., Fernbach, S., Combes, A., Menesses, A., et al. (2004). Echocardiographic evaluation of asymptomatic parental and sibling cardiovascular anomalies associated with congenital left ventricular outflow tract lesions. Pediatrics 114, 691–696. doi: 10.1542/peds.2003-0782-L
30. Clementi, M., Notari, L., Borghi, A. & Tenconi, R. Familial congenital bicuspid aortic valve: a disorder of uncertain inheritance. *American Journal of Medical Genetics* 62, 336-8 (1996).
31. Freeze, S. L., Landis, B. J., Ware, S. M., and Helm, B. M. (2016). Bicuspid aortic valve: a review with recommendations for genetic counseling. J. Genet. Couns. 25, 1171–1178. doi: 10.1007/s10897-016-0002-6
32. Martin, L.J. et al. Evidence in favor of linkage to human chromosomal regions 18q, 5q and 13q for bicuspid aortic valve and associated cardiovascular malformations. Hum Genet 121, 275-84 (2007).
33. Garg, V. *et al.* Mutations in NOTCH1 cause aortic valve disease. *Nature* 437, 270-4 (2005).
34. Hasham, S.N. *et al.* Mapping a locus for familial thoracic aortic aneurysms and dissections (TAAD2) to 3p24-25. *Circulation* 107, 3184-90 (2003).

35. McKellar, S.H. *et al.* Novel NOTCH1 mutations in patients with bicuspid aortic valve disease and thoracic aortic aneurysms. *Journal of Thoracic and Cardiovascular Surgery* 134, 290-6 (2007).
36. Guo, D.C. *et al.* Mutations in smooth muscle alpha-actin (ACTA2) lead to thoracic aortic aneurysms and dissections. *Nat Genet* 39, 1488-93 (2007).
37. Masuno, M. *et al.* Ehlers-Danlos syndrome, vascular type: a novel missense mutation in the COL3A1 gene. *Congenit Anom (Kyoto)* 52, 207-10 (2012).
38. Dietz, H.C. *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 352, 337-9 (1991).
39. Hoffmann, A.D., Peterson, M.A., Friedland-Little, J.M., Anderson, S.A. & Moskowitz, I.P. sonic hedgehog is required in pulmonary endoderm for atrial septation. Development 136, 1761-70 (2009).
40. Bertrand, N. et al. Hox genes define distinct progenitor sub-domains within the second heart field. Dev Biol 353, 266-74 (2011).
41. Mailleux, A.A., Moshai, E.F. & Crestani, B. Sonic Hedgehog signaling in pulmonary fibrosis: a spiky issue? Am J Physiol Lung Cell Mol Physiol 304, L391-3 (2013).
42. Li, Y. et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature 521, 520-4 (2015)
43. Wang, L., Ming Wang, L., Chen, W. & Chen, X. Bicuspid Aortic Valve: A Review of its Genetics and Clinical Significance. *J Heart Valve Dis* 25, 568-573 (2016).
44. Phillippi, J.A. *et al.* Mechanism of aortic medial matrix remodeling is distinct in patients with bicuspid aortic valve. *Journal of Thoracic and Cardiovascular Surgery* 147, 1056-64 (2014).
45. Laforest, B., and Nemer, M. (2012). Genetic insights into bicuspid aortic valve formation. Cardiol. Res. Pract. 2012:180297. doi: 10.1155/2012/180297
46. De Cario, R., Sticchi, E., Giusti, B., Abbate, R., Gensini, G. F., Nistri, S., et al. (2014). Bicuspid aortic valve syndrome and fibrillinopathies: potential impact on clinical approach. Int. Cardiovasc. Forum J. 4, 167–174. doi: 10.17987/icfj.v1i4.45]
47. J.R. Barbour, F.G. Spinale, J.S. Ikonomidis. Proteinase systems and thoracic aortic aneurysm progression. J Surg Res, 139 (2007), pp. 292–307]
48. E.M. Isselbacher. Thoracic and abdominal aortic aneurysms. Circulation, 111 (2005), pp. 816–82
49. J.S. Ikonomidis, J.A. Jones, J.R. Barbour, R.E. Stroud, L.L. Clark, B.S. Kaplan, et al. Expression of matrix metalloproteinases and endogenous inhibitors within ascending aortic aneurysms of patients with Marfan syndrome. Circulation, 114 (1 Suppl) (2006), pp. I365–I370
50. He, R. *et al.* Characterization of the inflammatory and apoptotic cells in the aortas of patients with ascending thoracic aortic aneurysms and dissections. *Journal of Thoracic and Cardiovascular Surgery* 131, 671-8 (2006).
51. Martufi, G. *et al.* Case Study: Intra-Patient Heterogeneity of Aneurysmal Tissue Properties. *Front Cardiovasc Med* 5, 82 (2018).
52. Didangelos, A. et al. Extracellular matrix composition and remodeling in human abdominal aortic aneurysms: a proteomics approach. Mol Cell Proteomics 10, M111 008128 (2011).
53. Sherratt, M.J. *et al.* Fibrillin microfibrils are stiff reinforcing fibres in compliant tissues. *Journal of Molecular Biology* 332, 183-93 (2003).
54. Chaudhry, S.S. et al. Fibrillin-1 regulates the bioavailability of TGFbeta1. Journal of Cell Biology 176, 355-67 (2007).
55. Lindsay, M.E. & Dietz, H.C. Lessons on the pathogenesis of aneurysm from heritable conditions. *Nature* 473, 308-16 (2011).
56. J.S. Ikonomidis, J.A. Jones, J.R. Barbour, R.E. Stroud, L.L. Clark, B.S. Kaplan, et al. Expression of matrix metalloproteinases and endogenous inhibitors within ascending aortic aneurysms of patients with bicuspid or tricuspid aortic valves. J Thorac Cardiovasc Surg, 133 (2007), pp. 1028–1036]
57. L. Verschuren, J.H. Lindeman, J.H. van Bockel, H. Abdul-Hussien, T. Kooistra, R. Kleemann. Up-regulation and co-expression of MIF and matrix metalloproteinases in human abdominal aortic aneurysms Antioxid Redox Signal, 7 (2005), pp. 1195–1202]
58. El-Hamamsy, I. & Yacoub, M.H. Cellular and molecular mechanisms of thoracic aortic aneurysms. *Nat Rev Cardiol* 6, 771-86 (2009).
59. Qin, C. *et al.* Dynamic monitoring of platelet activation and its role in post-dissection inflammation in a canine model of acute type A aortic dissection. *J Cardiothorac Surg* 11, 86 (2016).

60. He, R. *et al.* Characterization of the inflammatory and apoptotic cells in the aortas of patients with ascending thoracic aortic aneurysms and dissections. *Journal of Thoracic and Cardiovascular Surgery* 131, 671-8 (2006).
61. Anderson, C.L. & Brown, C.J. Polymorphic X-chromosome inactivation of the human TIMP1 gene. *Am J Hum Genet* 65, 699-708 (1999).
62. Tukiainen, T. *et al.* Landscape of X chromosome inactivation across human tissues. *Nature* 550, 244-248 (2017).
63. Pettersen, M. D., Du, W., Skeens, M. E., & Humes, R. A. (2008). Regression equations for calculation of z scores of cardiac structures in a large cohort of healthy infants, children, and adolescents: An echocardiographic study. Journal of the American Society of Echocardiography,21(8), 922–934.
64. Quezada, E., Lapidus, J., Shaughnessy, R., Chen, Z., & Silberbach, M. (2015). Aortic dimensions in Turner syndrome. American Journal of Medical Genetics Part A, 167A, 2527–2532.

Chapter I:

Allometric considerations when assessing aortic aneurysms in Turner syndrome: Implications for activity recommendations and medical decision-making

Holly Corbitt[1], Cheryl Maslen[1], Siddharth Prakash[2], Shaine A. Morris[3], Michael Silberbach[4]

1. Knight Cardiovascular Institute, Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon
2. Department of Internal Medicine, McGovern Medical School, University of Texas Health Science Center, Houston, Texas
3. Department of Pediatrics, Baylor College of Medicine, Houston, Texas
4. Department of Pediatrics, Doernbecher Children's Hospital, Oregon Health & Science University, Portland, Oregon

Correspondence:
Michael Silberbach, MD, Department of Pediatrics, Doernbecher Children's Hospital, Oregon Health & Science University, CDRC-P, 3181 SW Sam Jackson Park Road, Oregon Health & Science University, Portland, OR 97239
Email: silberbm@ohsu.edu

Funding information:
Turner Syndrome Foundation of the United States; Leaping Butterfly Ministries; Philips Diagnostic Systems

Published:
Am J Med Genet. 2018;176A:277–282
https://doi.org/10.1002/ajmg.a.38584

Abstract

In Turner syndrome, the potential to form thoracic aortic aneurysms requires routine patient monitoring. However, the short stature that typically occurs complicates the assessment of severity and risk because the relationship of body size to aortic dimensions is different in Turner syndrome compared to the general population. Three allometric formula have been proposed to adjust aortic dimensions, all employing body surface area: aortic size index, Turner syndrome-specific Z-scores, and Z-scores based on a general pediatric and young adult population. In order to understand the differences between these formula we evaluated the relationship between age and aortic size index and compared Turner syndrome-specific Z-scores and pediatric/young adult based Z-scores in a group of girls and women with Turner syndrome. Our results suggest that the aortic size index is highly age-dependent for those under 15 years; and that Turner-specific Z-scores are significantly lower than Z-scores referenced to the general population. Higher Z-scores derived from the general reference population could result in stigmatization, inappropriate restriction from sports, and increasing the risk of unneeded medical or operative treatments. We propose that when estimating aortic dissection risk clinicians use Turner syndrome-specific Z-score for those under fifteen years of age.

Introduction

Catastrophic aortic dissection and rupture has been described in Turner syndrome in girls as young as 4 years of age (Lippe & Kogut, 1972) and in women up to 64 years old (Carlson & Silberbach, 2007). Clinicians therefore must monitor the size of the aorta in order to make decisions about permitting sports participation, initiating medications, and determining the need for operations. Accordingly, aortic measurements throughout the life span have been recommended (Gravholt et al., 2017).

Clinical geneticists, pediatric caregivers, and development biologists continue to grapple with the complex relationship between overall body size and the variable growth of individual organs.

Different allometric approaches have been suggested to adjust aortic measurements for body size. For example, Davies et al. (2006) stratified risk in patients with aneurysmal disease using the aortic size index (ASI = aortic diameter (cm)/body surface area (m^2). In adults, the ASI is now routinely employed for medical and operative decision making. We and others have observed that in adult women with Turner syndrome an ascending ASI >2.5 cm/m^2 is associated with an increased risk of aortic dissection (Carlson, Airhart, Lopez, & Silberbach, 2012; Matura, Ho, Rosing, & Bondy, 2007). However, ASI has never been studied in children. Importantly, ASI calculations in children may be misleading because of the non-constant variance (heteroscedasticity) associated with rapid somatic growth.

The need for scaling based on body size has led to the practice in pediatric patients of transforming absolute measurements of the aorta into Z-scores employing body surface area (BSA) (Pettersen, Du, Skeens, & Humes, 2008; Sluysmans & Colan, 2005). The Z-score, also known as the standardized score, is the number of standard deviations that an individual's aortic diameter is above or below the mean value of a reference population. In pediatrics, the aortic Z-score is typically calculated by comparison to a healthy general reference population. However, the characteristic short stature, small BSA, and differences in somatic growth trajectories of those with Turner syndrome raise the concern that aortic measurements adjusted for norms based on the general population could lead to exaggerated estimates of aortic dimensions. Employing a non-specific reference population could potentially result in unnecessary restriction of physical activity and/or inappropriate treatment. Accordingly, Quezada et al. (2015) recently reported Turner syndrome-specific aortic Z-score formula by using healthy girls and women with Turner syndrome as the reference population. However, the relationship between Z-scores derived from children with Turner syndrome versus a general population (hereafter referred to as “pediatric/young adult based Z-scores”) has not been previously reported.

Therefore, the goals of the present study are: (i) to evaluate the relationship between age and ASI in Turner syndrome and (ii) to compare Turner syndrome-specific Z-scores and pediatric/young adult based Z-scores.

Methods

2.1 | Study population

2.1.1 | Patient population and echocardiograms

All data utilized for the current analysis has been reported previously (Quezada et al., 2015). Briefly, 458 subjects with Turner syndrome ages 2–65 had focused echocardiograms performed as part of the Turner syndrome Healthy Heart Project, and all were performed at the annual meetings of the Turner syndrome Society of the United States. The Oregon Health & Science Institutional Review Board approved the protocol and continues to oversee this ongoing longitudinal study. Written consent from adult subjects or legal guardians and assent from children was obtained in all cases. Subjects included in the study were healthy females with Turner syndrome. Bicuspid aortic valve without significant aortic stenosis were included (Doppler-determined aortic valve velocity less than 2 m/sec). Exclusion criteria included greater than trace aortic insufficiency, un-operated, or structural congenital heart disease (other than unobstructed bicuspid aortic valve), and those who had elective surgery because of a dilated aorta, or who had a history of aortic dissection. The focused echo protocol used in this study has been previously described and was performed in accordance with the standard guidelines (Lopez et al., 2010; Quezada et al., 2015).

Aortic size index of the ascending aorta was calculated for all subjects (cm/m^2) using the Haycock body surface area equation (Haycock, Schwartz, & Wisotsky, 1978). For patients <age 30 years, Turner syndrome-specific Z-scores and pediatric/young adult based Z-scores (Colan, 2009; Sluymans & Colan, 2009) were also calculated. The pediatric/young adult aortic Z-scores calculator can be accessed on the web (http://zscore.chboston.org/) as is the calculator for Turner syndrome-specific Z-scores (http://www.parameterz.com/refs/quezada-ajmg-2015). Analyses focused on

ascending but not root ASI because we found considerable overlap between healthy subjects and subjects with aortic dissection when aortic root ASI were compared (data not shown). The Turner syndrome-specific and pediatric/young adult based Z-scores for the ascending aorta for each subject are provided in the Supplementary Data.

2.2 | Statistical analysis

Plots demonstrating the relationship between ASI and subject age were subjected to a breakpoint analysis using the Segmented package in R. The breakpoint of 15 years of age was used for all downstream analyses (<15 years and ≥15 years) (Supplementary Figure S1). Mean ascending ASI values for both groups (<15 years and ≥15 years) were compared using a non-parametric Mann–Whitney–Wilcoxon test, where the significance level was <0.05. A chi-squared analysis, with Yate's correction, was performed to assess differences in proportions between < or ≥ 15 years and ASI values > or < 2.5 cm/m^2. Linear regression models were fit for both groups against the outcome variable ascending ASI, where the lines with 95%confidence intervals were plotted and the adjusted R^2 values and significance levels are reported. Plots for quality control of each model are in Supplementary Figures S2 and S3.

For each subject, data points for ascending aorta measurements (cm) versus BSA (m^2) were plotted and lines for Z-scores corresponding to 0 and ±2 standard deviations were overlaid on the plot, where blue solid lines represent Turner syndrome-specific equations and red dashed lines represent the pediatric/young adult population equations (Figure 2). For all subjects <30 years old: ascending aorta Z-scores using the pediatric/young adult population equations and the Turner syndrome-specific equations were compared by scatter plot and means were compared using the Wilcoxon signed rank test. The difference between (delta) ascending aorta Z-scores were calculated for each subject (pediatric/young adult population Z-scores minus Turner syndrome-specific Z-score), and plotted against BSA (m^2), age (years), AAO dimensions (cm), and BMI (kg/m^2). For all plots, a linear regression model or quadratic regression model was fit to the data to look for an

association between delta Z-scores as the outcome variable and either BSA, age, AAO dimension, or BMI as the predictor variable.

Results

3.1 | Aortic size index

The plot of ascending ASI versus age (Figure 1) shows there is a breakpoint at 15 years age where the relationship between ascending ASI and age changes, (Supplementary Figure S1). In children with Turner syndrome <15 years, age is strongly negatively correlated with ASI values (R^2 = 0.485, p < 0.0001). For those ≥15 years with Turner syndrome, the correlation of ASI and age is minimal and is in the opposite direction. ASI values in those <15 years is significantly higher than in those ≥15 years (2.0 ± 0.5 vs. 1.6 ± 0.3, p < 0.0001). An ascending ASI >2.5 occurred in 14.0% of those <15 years and 23.2% <10 years. Whereas, in those ≥15 years an ascending ASI >2.5 occurred in 1.4% of subjects (p < 0.0001 chi squared, compared to <15 years).

Figure 1

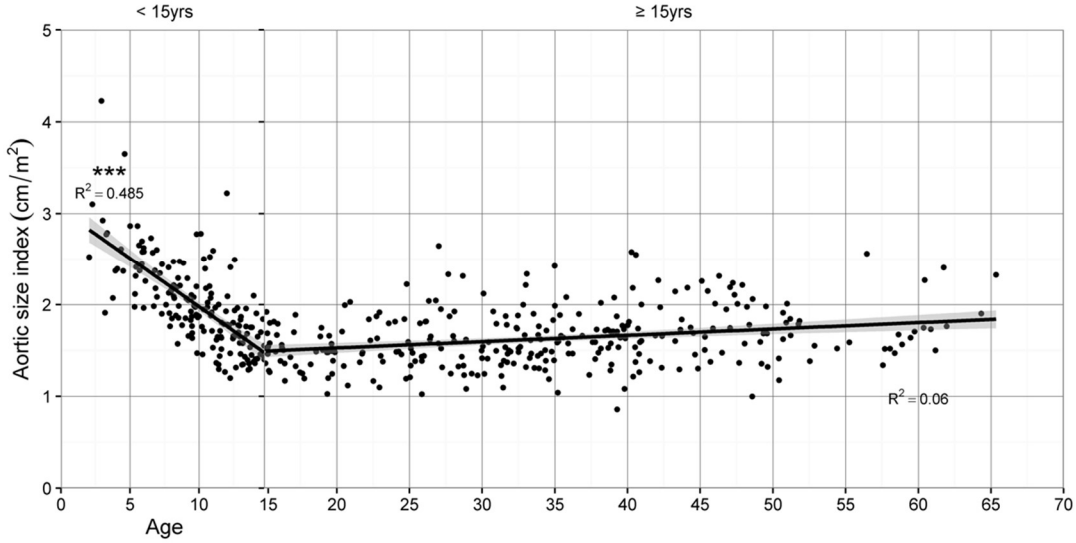


Figure 1: Aortic size index (cm/m^2) is independent of age only after 15 years. Regression lines are plotted with shading representing the 95%CI and with adjusted R^2 values of each model (***p < 0.0001)

3.2 | Ascending aorta Z-scores

Regression lines using both Turner syndrome-specific and pediatric/young adult based Z-scores corresponding to 0, +2, and −2 standard deviations are plotted for 246 measurements for 171 girls <15 years (Figure 2). Some subjects had multiple measurements performed over the years of the study.

Figure 2

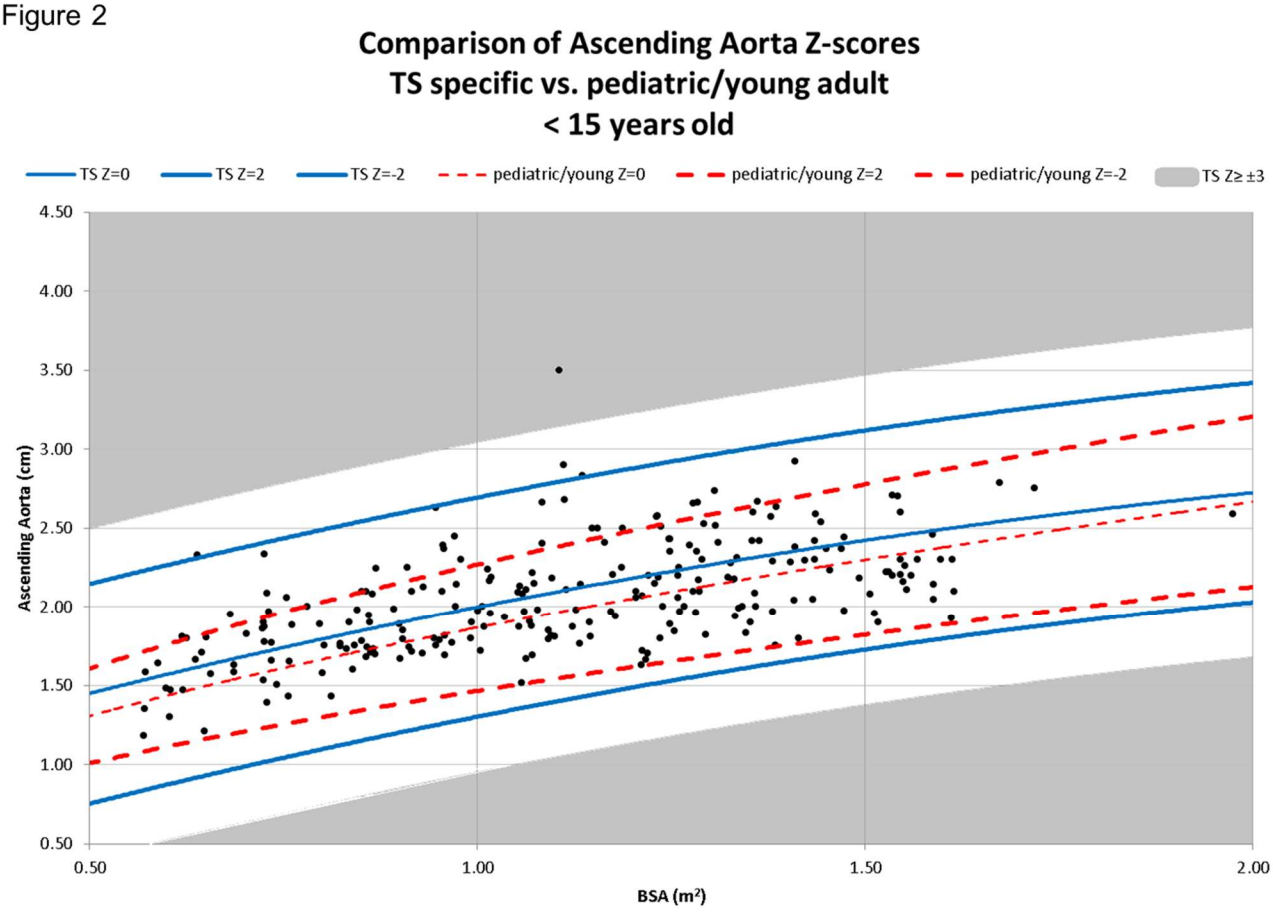


Figure 2: Comparison of TS-specific ascending aorta Z-scores to pediatric/young adult based Z-scores, by plotting of AAO dimensions (cm) versus BSA (m^2) for 246 subjects <15years.
The lines for Z-scores generated corresponding to 0 and ±2 are overlaid on the plot (blue continuous lines = TS-specific and red dashed lines = pediatric/young adult based based). Gray shading represents TS specific Z-score lines > ±3. (TS, Turner syndrome; AAO, ascending aorta).

Turner syndrome-specific Z-scores in subjects <15 years are significantly lower than Z-scores calculated according to pediatric/young adult based formula. The mean Z-scores for 246 measurements for subjects <15 years of age are −0.09 ± 0.89 (Turner syndrome specific) versus 0.54 ± 1.40 (pediatric/young adult based) is significantly different ($p < 0.0001$, Wilcoxon signed rank).

A majority, 95.5%, of the Turner syndrome-specific Z-scores were lower than pediatric/young adult based Z-scores (delta range 0–3.5), while only 4.5% had Turner syndrome-specific Z-scores higher than pediatric/young adult based Z-scores (delta range 0 to −0.15). A scatter plot of Turner syndrome specific Z-scores versus pediatric/young adult based Z-scores shows this trend, where most data points fall below the diagonal line, indicating Turner-specific Z-scores are lower than the pediatric/young adult based Z-scores (Figure 3).

Figure 3

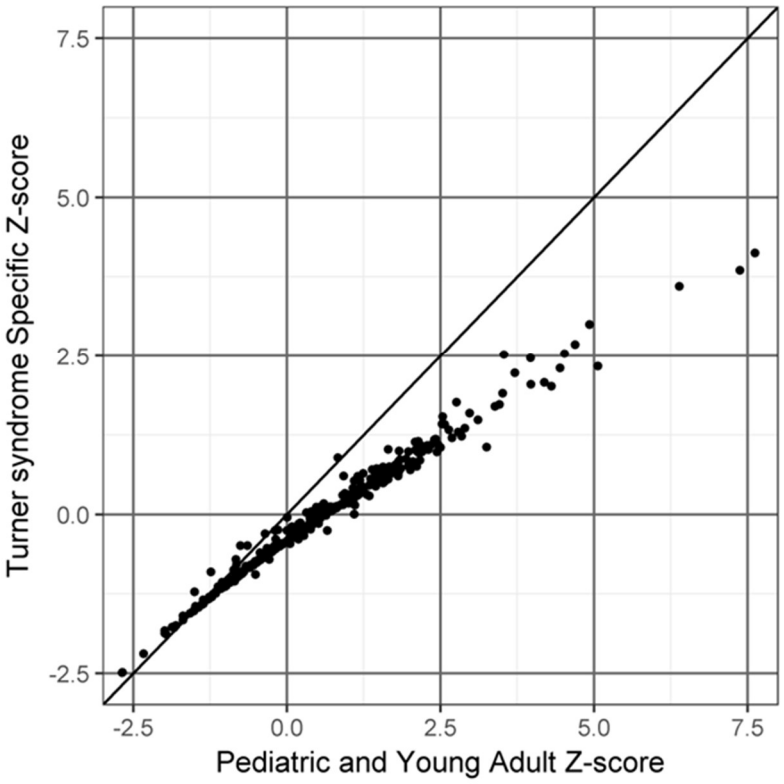


Figure 3: Pediatric and young adult vs. Turner syndrome-specific Z-scores.
Pediatric and young adult based Z-scores plotted against Turner syndrome-specific Z-scores shows that TS specific Z-scores trend lower as the Z-scores becomes higher

However, as BSA increases the difference between the Turner syndrome-specific Z-scores and pediatric/young adult based Z-scores approaches zero (Figure 4a). Similarly, as BMI increases, differences approach zero (Figure 4d), whereas regardless of age pediatric and young adult-based Z-scores are consistently higher than Turner-specific Z-scores (Figure 4b). Importantly, for subjects,

<15 years, who have a BSA similar to those who are ≥15 years (mean BSA 1.62m^2 ± 0.25) there seems to be good agreement between Turner syndrome-specific and pediatric and young adult based Z-scores, although there is a small sample size (Figure 4a and Supplementary Figure S4). A larger AAO absolute diameter correlates with a markedly positive difference between the two Z-scores particularly for values >3 cm (Figure 4c, $R^2 = 0.29$, $p < 0.0001$).

Discussion

This study has two principal findings: (i) Ascending ASI in children under 15 years is likely to be a poor predictor of risk because many children have values >2.5 cm/m^2 and (ii) Turner syndrome-specific Z-scores are significantly lower than Z-scores calculated using a pediatric/young adult reference population. Matura et al. (2007) suggested that an ASI >2.5 cm/ m^2 is predictive of aortic dissection. Carlson et al. (2012) subsequently confirmed this observation in nine cases of adults with type A aortic dissection where eight of nine had ascending ASI >2.5 cm/m^2. In the present study among 246 subjects ≥15 years, none of whom had a history of aortic dissection only 1.4% had an ascending ASI >2.5 cm/m^2. Thus, over the age of 15 years, ascending ASI >2.5 cm/m^2 can alert clinicians to those at risk for aortic dissection without classifying a large percentage of the population as high risk. The average BSA in subjects' ≥15 years was 1.62m^2. Thus, an average size adult woman with Turner whose aorta is >2.5 cm/m^2 is predicted to have absolute ascending aortic diameter of >4 cm. ASI calculation in individuals who are short-statured and obese or those who weigh very little relative to their height should be made with caution. Thus, an absolute value of >4 cm maybe better predictor of risk in those with very low or high BMIs. Our results suggest that a child with Turner syndrome under 15 years is 10 times more likely than older healthy individuals to have an ascending ASI >2.5 cm/m2, increasing the risk of obtaining false positive values. We found that as children approach the age of 15 years their aortic growth slows relative to body size, as indicated by

Figure 4

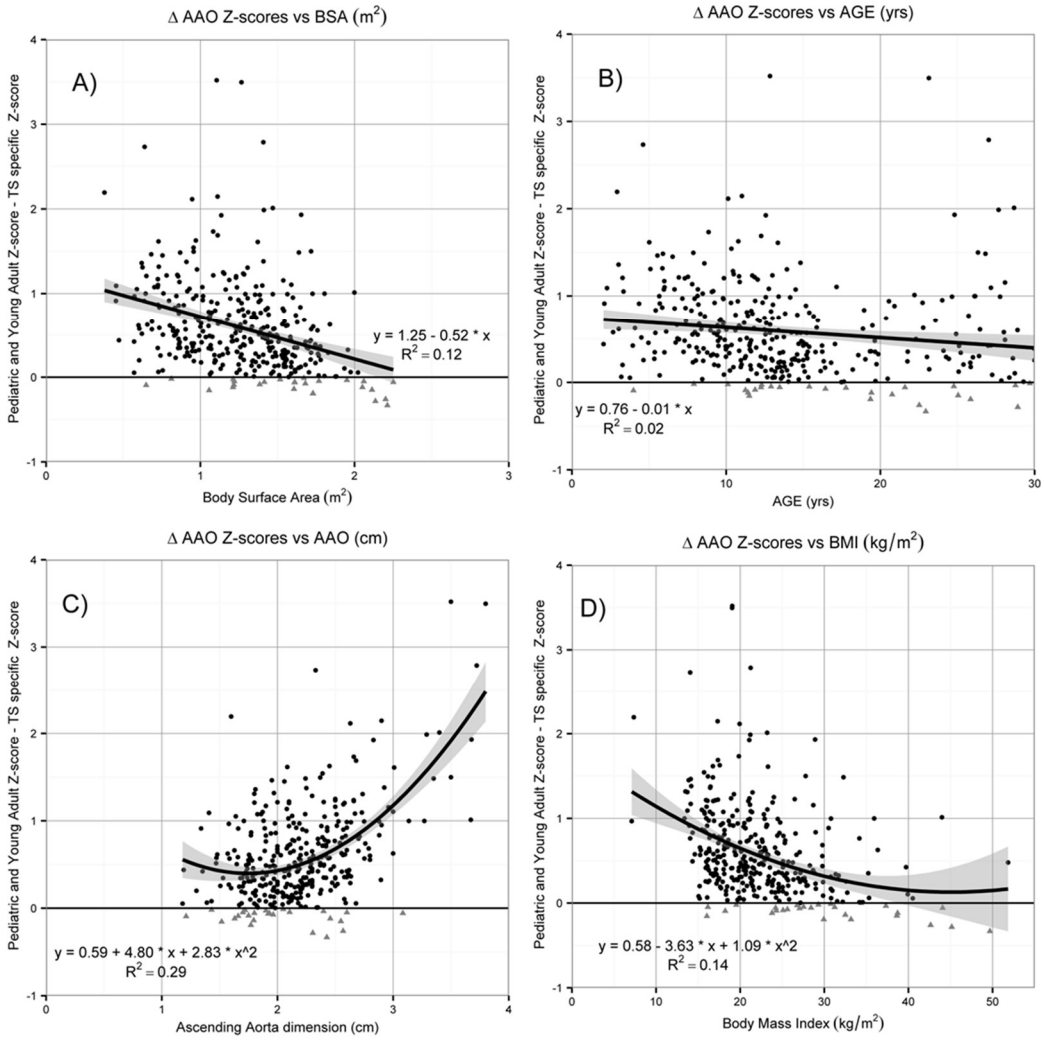


Figure 4: Plots of the difference between pediatric and young adult based Z-scores and Turner syndrome-specific Z-scores versus BSA, age, AAO dimensions, and BMI. The linear or quadratic regression lines were plotted with shading representing the 95%CI, along with equations and R2. Gray triangles are cases where the difference between TS-specific Z scores is higher than pediatric and young adult based Z-scores. (AAO, ascending aorta; BSA, body surface area; BMI, body mass index; TS, Turner syndrome; CI, confidence interval)

decelerating ascending ASI. On the other hand, as healthy children under the age of 15 years grow, Z-scores remain stable. Therefore, ASI is only useful after 15 years where ASI becomes independent of age. For those older than 15 years of age, employing Z-scores may be reasonable, but there are no studies to date that demonstrate this.

Lopez et al. (2008) determined that Turner syndrome alone is associated with small increases in ascending aorta aortic diameter independently of other factors such as bicuspid aortic

valve and aortic stenosis that separately increase aortic size. It is important to note that for an adult sized individual with Turner syndrome, an increase of a Turner syndrome-specific Z-score from 2 to 2.5 represents an increase in the ascending aortic diameter of 1.37mm, which is approximately the axial resolution of standard echocardiography probes. Thus, a slightly larger aorta is likely to be a benign characteristic of the Turner syndrome phenotype and the higher Z-score estimates for healthy girls with Turner syndrome associated with general population-based reference could result in stigmatization, counterproductive restriction from sports participation and/or inappropriate medical treatment. We found that the difference between Turner syndrome-specific Z-scores and those derived from a general reference population diverged significantly at the largest absolute ascending aorta diameters (Figure 4c). Thus, it appears that the variability of Z-score estimates increases more in general population-based values as vessels reach extremely large diameters, as others have shown (Ronai et al., 2016). It has been reported that among those with adult BSA, there is no difference in Z-scores when the two formulas are compared and could be used interchangeably (Prakash, Gen, & Milewicz, 2017) as suggested in (Figure 4a). On the other hand, for subjects at least to the age of 30 years, Turner specific Z-scores are significantly lower when compared to Z-scores based on the pediatric/young adult reference population (Figure 4B and Supplementary Figure S5).

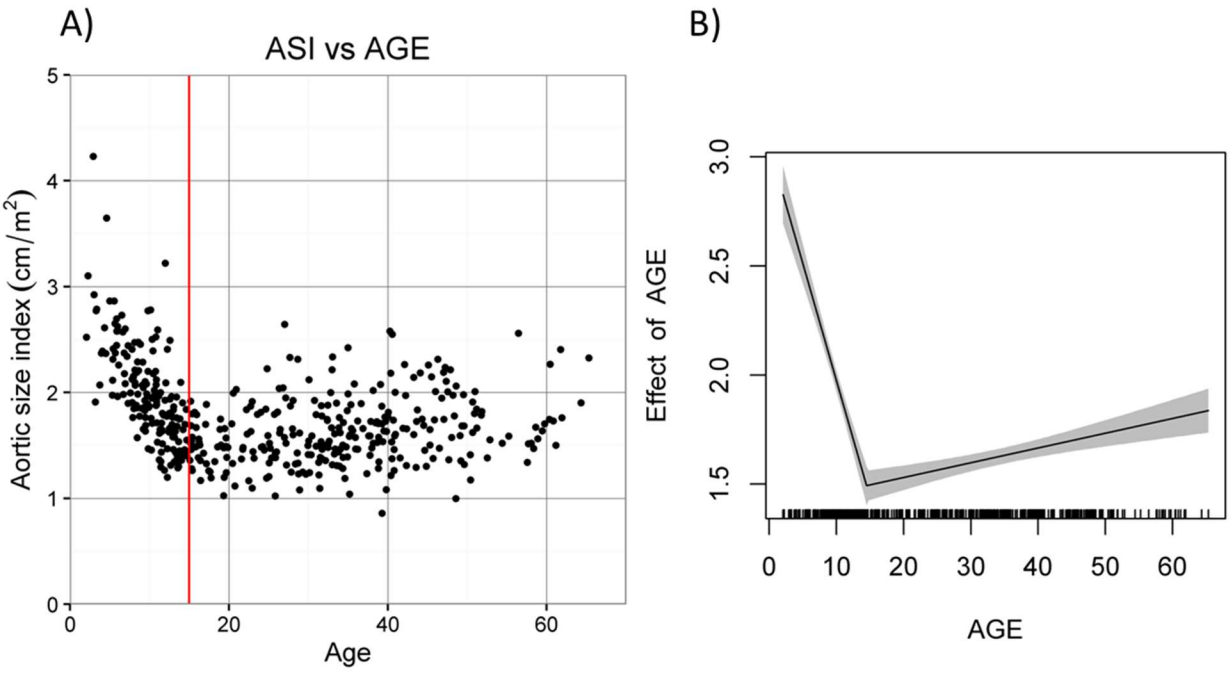
Since ascending ASI is age-dependent for those under 15 years, we believe that Turner syndrome-specific Z-scores may be helpful for decision-making. However, it is important to note that aortic dissection is rare under the age of 15 and no study has demonstrated Z-scores to be predictive of aortic dissection. For clinicians who must make decisions regarding medical/surgical therapies or sports participation, it may be reasonable to extrapolate from the adult experience. Thus, a Turner syndrome-specific Z-score >4 (Z-score calculator http://www.parameterz.com/refs/quezada-ajmg-2015) may serve as an indication of risk for aortic dissection because a Z-score value of four corresponds to an ascending ASI of >2.5 cm/m^2 in an adult with average body size (BSA ~1.6 m^2).

4.1 | Limitations

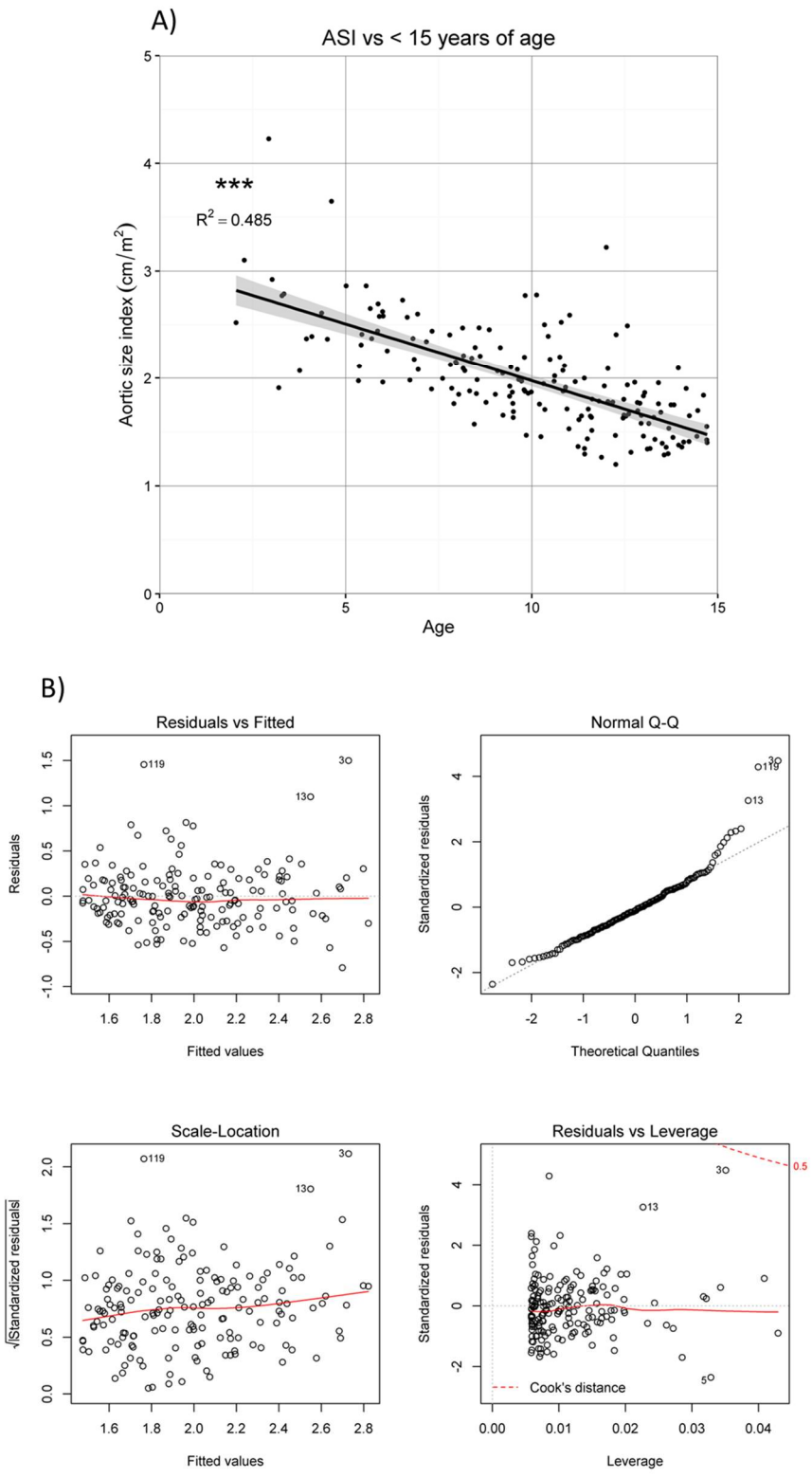
Aortic dissection under the age of 15 years is rare (Carlson & Silberbach, 2007) and no studies at any age have demonstrated that Z-score measurements are predictive. More follow-up is needed to determine if Turner syndrome-specific Z-scores improve predictive value compared to Z-scores based on a pediatric/young adult reference population.

Hopefully, longitudinal data will become available as the Turner syndrome Healthy Heart study that formed the basis of the present study, proceeds. To advance this research we would urge girls and women with Turner syndrome to join the Turner syndrome research registry (TSRR, http://www.turnersyndrome.org/ts-registry-love). Despite the limitation of available prospective data, the high false positive rate of the ASI for children under the age of 15 years makes Z-score assessment of aortic size in children a valuable tool for clinicians endeavoring to make difficult management decisions. Although, we found that many subjects under the age of 15 years and very few >15 years had ASI >2.5 cm/m^2 it is possible that longitudinal study may demonstrate a predictive value of the ASI in the young despite its low sensitivity.

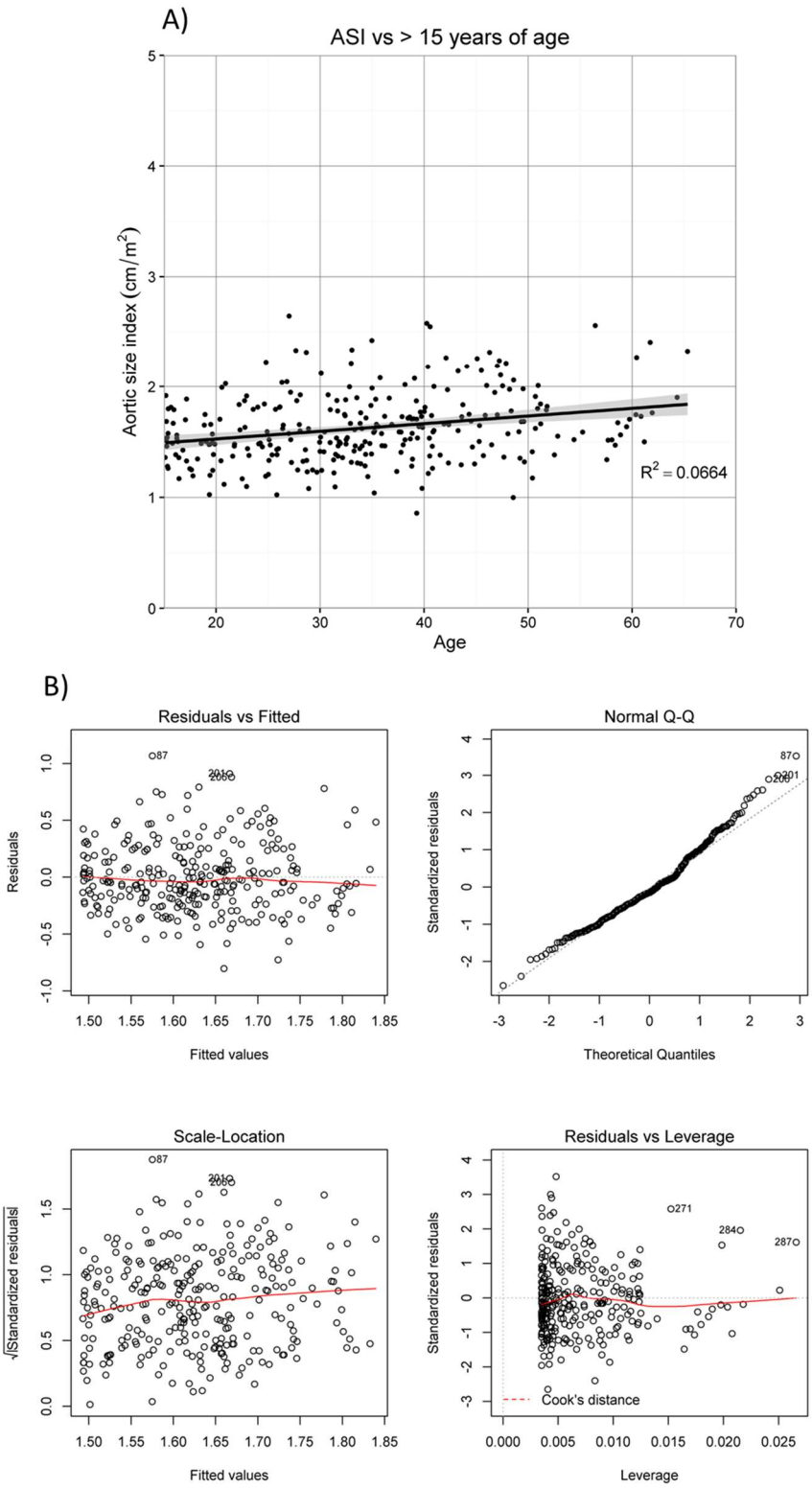
In summary, ascending ASI in those under 15 years of age is significantly larger than those ≥15 years, increasing the likelihood for overestimation of the risk for aortic dissection. Therefore, for those <15 years, Z-score estimates of aortic size adjusted for body size are preferable. Furthermore, Turner syndrome-specific Z-score estimates in those who have BSAs <1.62m^2 are lower than Z-scores based on the general pediatric/young adult reference population. We encourage the use of Turner syndrome-specific Z-scores because diagnosis of aortic enlargement using general population-based Z-score formulas may result in increased stigmatization, inappropriate restriction from sports, exclusion from other heart healthy activities, and increasing the risk of unneeded medical or operative treatments.



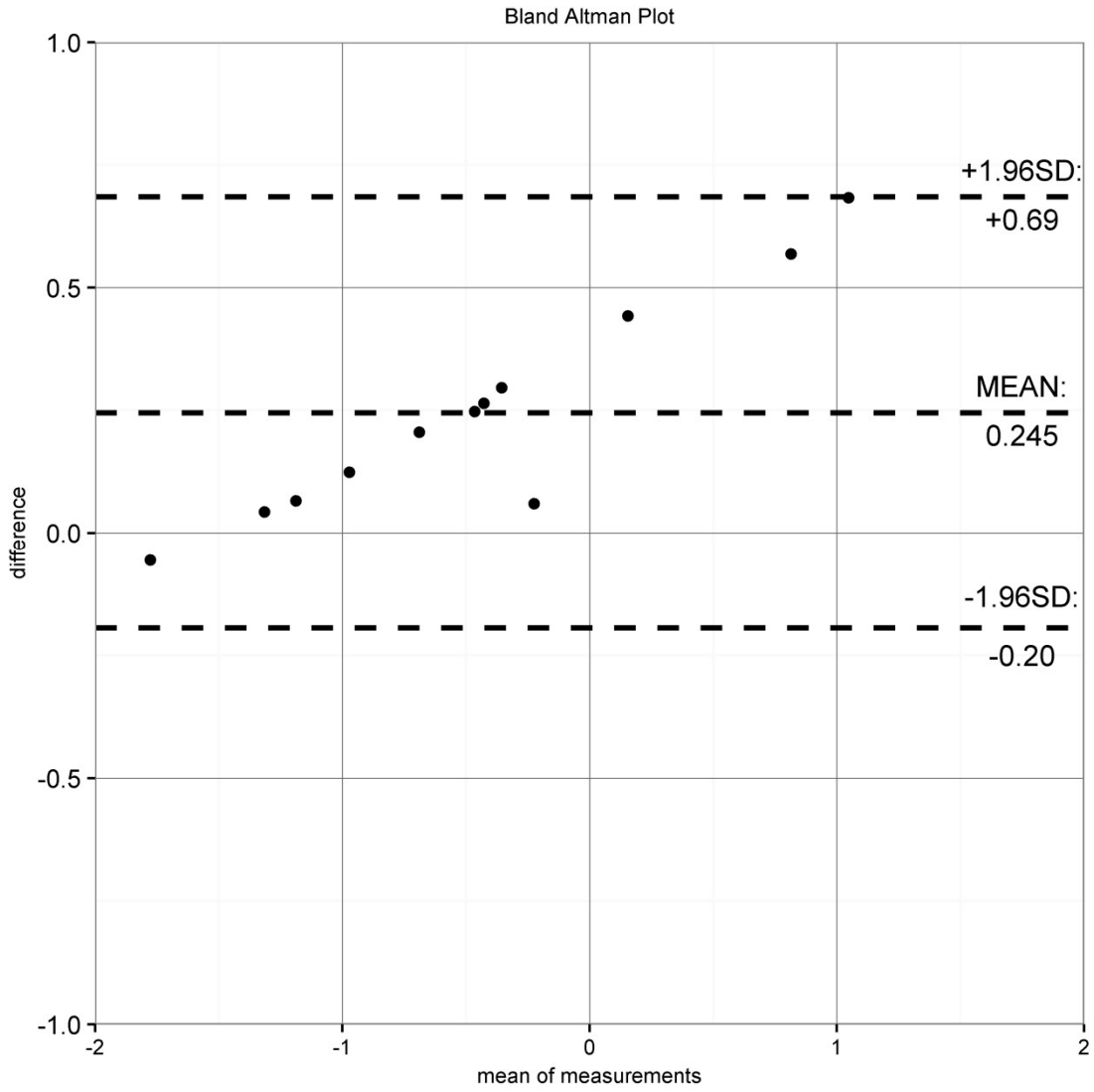
Supplementary Figure S1. Estimated breakpoint of ascending ASI and age relationship (A) Regression line, gray bars represent 5th and 95th confidence intervals (B) (Segmented package in R)



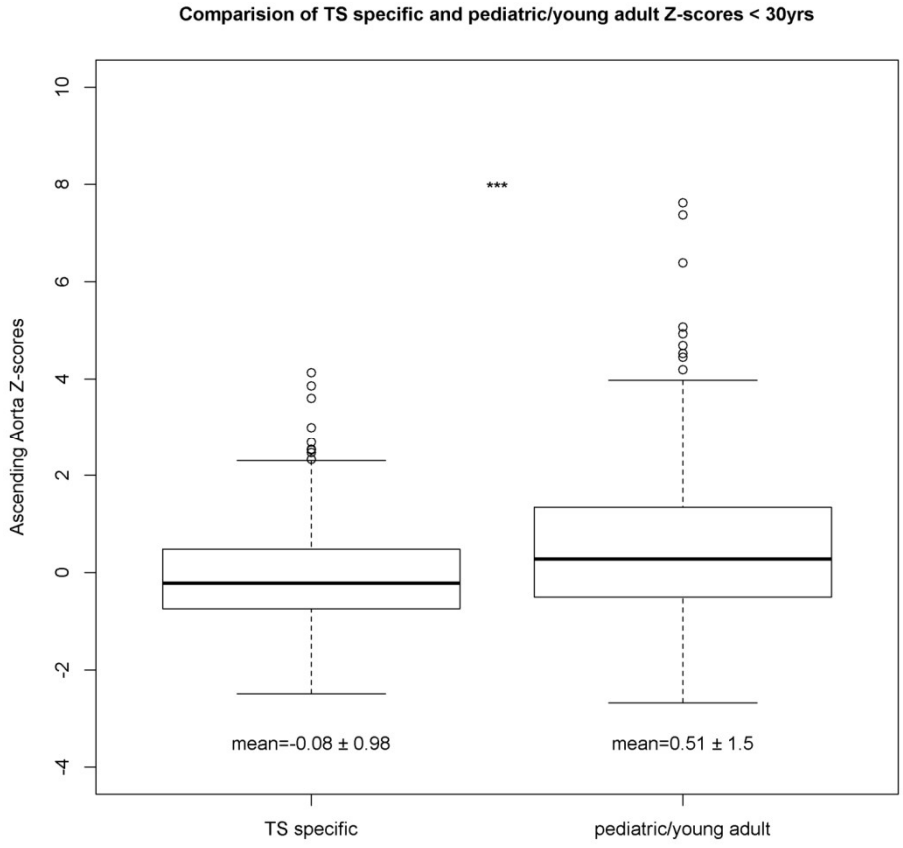
Supplementary Figure S2. Linear regression model for ascending ASI (outcome) and < 15 years of age (predictor) (A) Additional QC plots for the model (B).



Supplementary Figure S3. Linear regression model for ascending ASI (outcome) and > 15 years of age (predictor) (A) Additional QC plots for the model (B).



Supplementary Figure S4: Bland-Altman plot of difference in ascending aorta Z-scores (general population based minus Turner-syndrome specific) against the mean of the two measurements, for those with an adult like BSA (mean BSA 1.62m^2 ± 0.25).



Supplementary Figure S5: Comparison of TS-specific and pediatric/young adult Z-scores for all individuals < 30yrs old. (*** p<0.0001, Mann-Whitney). Means and standard deviation are reported.

References

Carlson, M., Airhart, N., Lopez, L., & Silberbach, M. (2012). Moderate aortic enlargement and bicuspid aortic valve are associated with aortic dissection in turner syndrome/clinical perspective. Circulation, 126(18), 2220–2226.

Carlson, M., & Silberbach, M. (2007). Dissection of the aorta in Turner syndrome: Two cases and review of 85 cases in the literature. Journal of Medical Genetics, 44(12), 745–749.

Colan, S. D. (2009). Normal echocardiographic values for cardiovascular structures. In W. W. Lai, M. S. Cohen, & L. L. Merten, (Eds.), Echocardiography in pediatric and congenital heart disease (pp. 765–785). West Susse UK: Wiley-Blackwell.

Davies, R. R., Gallo, A., Coady, M. A., Tellides, G., Botta, D. M., Burke, B., Elefteriades, J. A. (2006). Novel measurement of relative aortic size predicts rupture of thoracic aortic aneurysms. Annals of Thoracic Surgery, 81(1), 169–177.

Gravholt, C. H., Andersen, N. H., Conway, G. S., Dekkers, O. M., Geffner, M. E., Klein, K. O. International Turner Syndrome Consensus, Group. (2017). Clinical practice guidelines for the care of girls and women with Turner syndrome: Proceedings from the 2016 Cincinnati International Turner Syndrome Meeting. European Journal of Endocrinology, 177(3), G1–G70.

Haycock, G. B., Schwartz, G. J.,&Wisotsky, D. H. (1978). Geometric method for measuring body surface area: A height-weight formula validated in infants, children, and adults. Journal of Pediatrics, 93(1), 62–66.

Lippe, B. M., & Kogut, M. D. (1972). Aortic rupture in gonadal dysgenesis. Journal of Pediatrics, 80(5), 895–896.

Lopez, L., Arheart, K. L., Colan, S. D., Stein, N. S., Lopez-Mitnik, G., Lin, A. E.,. Silberbach, M. (2008). Turner syndrome is an independent risk factor for aortic dilation in the young. Pediatrics, 121(6), e1622–e1627.

Lopez, L., Colan, S. D., Frommelt, P. C., Ensing, G. J., Kendall, K., Younoszai, A. K.,. Geva, T. (2010). Recommendations for quantification methods during the performance of a pediatric echocardiogram: A report from the Pediatric Measurements Writing Group of the American Society of Echocardiography Pediatric and Congenital Heart Disease Council. Journal of the American Society of Echocardiography, 23(5), 465–495.

Matura, L. A., Ho, V. B., Rosing, D. R., & Bondy, C. A. (2007). Aortic dilatation and dissection in Turner syndrome. Circulation, 116(15), 1663–1670.

Pettersen, M. D., Du, W., Skeens, M. E., & Humes, R. A. (2008). Regression equations for calculation of z scores of cardiac structures in a large cohort of healthy infants, children, and adolescents: An echocardiographic study. Journal of the American Society of Echocardiography, 21(8), 922–934.

Prakash, S., Gen, T. A. C. R. I., & Milewicz, D. (2017). Turner syndrome specific and general population Z-scores are equivalent for most adults with Turner syndrome. American Journal of Medical Genetics Part A, 173(4), 1094–1096.

Quezada, E., Lapidus, J., Shaughnessy, R., Chen, Z., & Silberbach, M. (2015). Aortic dimensions in Turner syndrome. American Journal of Medical Genetics Part A, 167A, 2527–2532.

Ronai, C., Hamaoka-Okamoto, A., Baker, A. L., de Ferranti, S. D., Colan, S. D., Newburger, J. W., & Friedman, K. G. (2016). Coronary artery aneurysm measurement and z score variability in kawasaki disease. Journal of the American Society of Echocardiography, 29(2), 150–157.

Sluymans, T., & Colan, S. D., (2009). Structural measurements and adjustment for growth. In M. S. Cohen, T. Geva, & L. L. Mertens, (Eds.), Echocardiography in pediatric and congenital heart disease. West Sussex, UK: Wiley-Blackwell.

Sluysmans, T., & Colan, S. D. (2005). Theoretical and empirical derivation of cardiovascular allometric relationships in children. Journal of Applied Physiology, 99(2), 445–457.

Chapter II:

TIMP3 and *TIMP1* are risk genes for bicuspid aortic valve and aortopathy in Turner syndrome

Holly Corbitt[1, 2], Shaine A. Morris[3], Claus H. Gravholt[4, 5], Kristian H. Mortensen[6], Rebecca Tippner-Hedges[1], GenTAC Registry Investigators, Michael Silberbach[7, ¶,*], Cheryl L. Maslen[1, 2, ¶,*]

1.Knight Cardiovascular Institute, Oregon Health & Science University, Portland, Oregon
2.Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon
3. Department of Pediatrics, Division of Pediatric Cardiology, Baylor College of Medicine, Houston, Texas
4. Department of Endocrinology and Internal Medicine and Medical Research Laboratories, Aarhus University Hospital, Aarhus, Denmark
5. Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark;
6.Cardiorespiratory Unit, Great Ormond Street Hospital for Children, London, United Kingdom
7. Department of Pediatrics, Division of Pediatric Cardiology, Oregon Health & Science University, Portland, Oregon

*Corresponding authors

Email: maslenc@ohsu.edu (CLM), silberbm@ohsu.edu (MS)

¶These authors contributed equally to this work.

Published:
PLoS Genetics
https://doi.org/10.1371/journal.pgen.1007692

Abstract

Turner syndrome is caused by complete or partial loss of the second sex chromosome, occurring in ~1 in 2,000 female births. There is a greatly increased incidence of aortopathy of unknown etiology, including bicuspid aortic valve (BAV), thoracic aortic aneurysms, aortic dissection and rupture. We performed whole exome sequencing on 188 Turner syndrome participants from the National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Related Conditions (GenTAC). A gene-based burden test, the optimal sequence kernel association test (SKAT-O), was used to evaluate the data with BAV and aortic dimension z-scores as covariates. Genes on chromosome Xp were analyzed for the potential to contribute to aortopathy when hemizygous. Exome analysis revealed that *TIMP3* was associated with indices of aortopathy at exome-wide significance ($p = 2.27 \times 10^{-7}$), which was replicated in a separate cohort. The analysis of Xp genes revealed that *TIMP1*, which is a functionally redundant paralogue of *TIMP3*, was hemizygous in >50% of our discovery cohort and that having only one copy of *TIMP1* increased the odds of having aortopathy (OR = 9.76, 95% CI = 1.91–178.80, p = 0.029). The combinatorial effect of a single copy of *TIMP1* and *TIMP3* risk alleles further increased the risk for aortopathy (OR = 12.86, 95% CI = 2.57–99.39, p = 0.004). The products of genes encoding tissue inhibitors of matrix metalloproteinases (TIMPs) are involved in development of the aortic valve and protect tissue integrity of the aorta. We propose that the combination of X chromosome *TIMP1* hemizygosity and variants of its autosomal paralogue *TIMP3* significantly increases the risk of aortopathy in Turner syndrome.

Author Summary

BAV is the most frequent congenital heart defect, occurring in about 1–2% of the population with 70% of cases occurring in males. BAV increases risk for thoracic aortic aneurysm (TAA) and early death. Approximately 30% of individuals with Turner syndrome have BAV/TAA, making this an important population for the study of this disease. Given that individuals with Turner syndrome are

missing a complete or partial second sex chromosome, it is presumed that X chromosome genes are involved in causing the defect. This is consistent with the bias towards occurrence in euploid males. However, not everyone with Turner syndrome has a BAV, so we hypothesized that autosomal genes may also play a role. Using whole exome sequencing we have shown that deleterious variation in *TIMP3* is associated with BAV and indices of TAA. We further found that there is a synergistic interaction between loss of the X chromosome gene, *TIMP1*, and deleterious variation in *TIMP3* that significantly increases that risk. TIMP1 and TIMP3 play roles in aortic valve morphogenesis and in stabilizing the aortic wall, loss of which leads to TAA. Hence our findings have implications for understanding the cause of BAV/TAA in all populations and as a potential therapeutic target.

Introduction

Turner syndrome is the most common sex chromosome aneuploidy, where ~50% have a complete monosomy X and ~48% have either a partial loss, rearrangement, or mosaicism of a second X chromosome.[1] The remaining ~2% have a partial or mosaic Y chromosome. Although Turner syndrome can be compatible with life, less than 1% of Turner syndrome fetuses survive.[2] The majority of prenatal deaths are due to cardiovascular defects.[3] Live born females with Turner syndrome share a constellation of phenotypes including primary ovarian insufficiency, short stature, lymphedema, webbed neck, skeletal deformities, neurocognitive disability, and a high incidence of congenital cardiovascular malformations. In particular, they are at a greatly increased risk for having left heart obstructions including hypoplastic left heart syndrome, BAV, coarctation of the aorta, and TAA.[4] Heart defects are the major cause of premature death. The degree to which a second sex chromosome is retained is the primary determinant of the morbidity and mortality in Turner syndrome, an observation that strongly implicates X chromosomal genetics in the pathology of acquired and congenital cardiovascular disease.[5, 6] In depth studies have shown that BAV, coarctation of the aorta, and risk for aneurysm are linked to the short arm of the X chromosome (Xp).[7, 8]

BAV is a congenital malformation where the aortic valve is comprised of two leaflets as opposed to the normal three leaflet configuration. BAV is associated with lifelong heart disease including valve calcification, stenosis, aortic endocarditis, and thoracic aortic dilation (TAD) that has a high risk of progression to aneurysm, dissection and rupture, and premature death. It is the most common congenital heart malformation occurring in about 2% of the general population where it is predominantly found in males, which comprise about 70% of all BAV cases.[9] However, despite the prevalence in the population, little is known about the etiology of BAV. There is clearly a genetic component as 10–40% of BAV is familial.[10] BAV and aortic aneurysm are thought to have a common genetic etiology.[11] Mutations in *NOTCH1*[12], *GATA5*[13], and *NKX2.5*[14] have been identified as the causative factor in some families with inherited BAV, but the majority of cases remain unexplained. The sex bias in euploid BAV indicates that having two X chromosomes may be protective. In Turner syndrome the incidence of BAV is increased by at least 50-fold over that seen in the euploid population.[15] This suggests that the lack of a second X chromosome predisposes both males and Turner syndrome females to have BAV and TAA, a condition known as BAV aortopathy.

Although there is a paucity of information about the etiology for BAV, a great deal is known about the pathogenic events underlying TAA and dissections associated with BAV. Numerous studies have shown significantly increased expression of matrix metalloproteinases (MMPs) and decreased expression of TIMPs in aneurysmal tissue.[16] This is significant because the role of MMPs is to degrade extracellular matrix (ECM); an activity that is inhibited by TIMPs. It is thought that in aneurysms the ECM in the aortic wall becomes degraded by MMPs, which weakens the aorta allowing it to succumb to hemodynamic stress thereby enlarging the diameter and thinning the aortic wall. In particular, increased expression of MMP2 and MMP9, which degrade the collagen and elastin components of the aortic wall, and a decrease in TIMP1, which inhibits MMP2 and MMP9 activity, have been implicated in the pathogenesis of aortic aneurysms.[16] In addition, an increased

MMP9/TIMP1 ratio has been shown to be elevated in chronic aortic dissection, demonstrating a persistent role for ECM degradation.[17]

Deficiency of the second sex chromosome contributes to aortopathy in Turner syndrome, but its loss is not sufficient to cause disease since ~50% of women with Turner syndrome have a normal aortic valve and aortic dimensions. We hypothesized that autosomal genetic variation sensitized by sex chromosome deficiency causes aortopathy in Turner syndrome. To address this hypothesis we used whole exome sequencing to identify autosomal genetic variation associated with BAV and TAD in Turner syndrome. We used TAD as an indicator of aneurysm formation. This study of a discovery cohort of 188 and a replication cohort of 53 individuals with Turner syndrome identified an exome-wide significant association between *TIMP3* (MIM: 188826) and BAV/TAD. Furthermore, investigation of the *TIMP3* paralog, *TIMP1* (MIM: 305370), revealed that having more than one copy of the Xp chromosome gene *TIMP1* was protective against BAV/TAD. Combinatorial analysis shows a synergistic effect between having a single copy of *TIMP1* plus the *TIMP3* risk allele and the occurrence of BAV/TAD. Knowledge of a direct link between TIMP family-gene expression and aortopathy points the way to the development of novel biomarkers for disease progression and therapies to combat catastrophic aortic dissection and rupture in Turner syndrome.

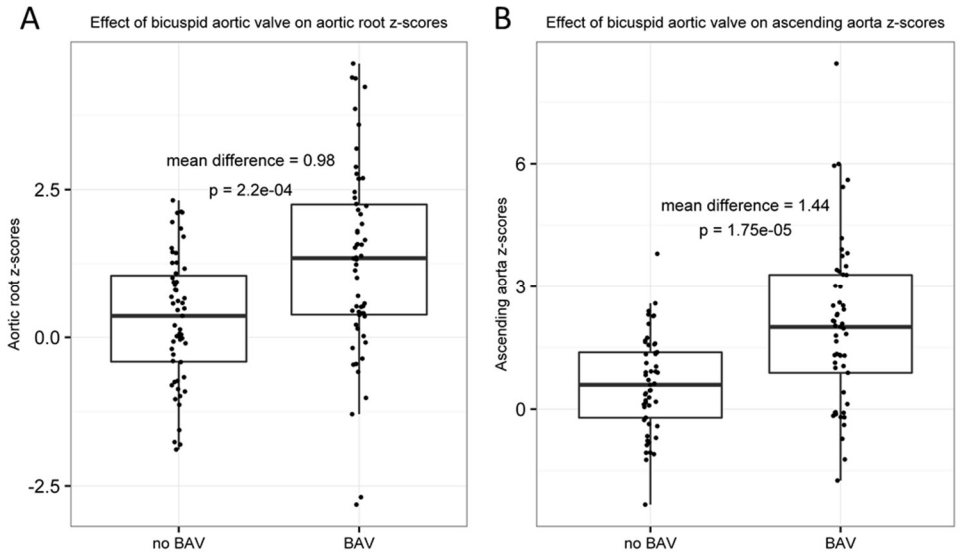


Fig 1. Aortic z-scores are associated with BAV in the Turner syndrome discovery cohort. Box plot of **A)** AR z-scores and **B)** AAO z-scores for individuals with and without a BAV, where both were significantly associated with the presence of a BAV, p = 0.0002 and p<0.0001, respectively.

Results

Correlation of Aortic Enlargement with BAV

The presence of BAV was associated with a higher aortic root (AR) z-score (mean AR z-score in BAV 1.29 ±1.59, versus no BAV 0.31 ± 1.08, p = 0.0002, mean difference = 0.98; Fig 1A). BAV was also associated with a significantly higher ascending aorta (AAO) z-score (mean AAO z-score in BAV 2.04 ± 1.99, versus no BAV 0.61 ± 1.18, p<0.0001, mean difference = 1.44; Fig 1B).

Association of *TIMP3* with BAV and Aortic Dilation

SKAT-O analysis revealed that variants in *TIMP3* on chromosome 22 achieved exome-wide significance for association with BAV and TAD. *TIMP3* was associated with the occurrence of BAV when it was used as the sole dichotomous phenotype ($p = 1.58 \times 10^{-6}$; Fig 2A), with the significance level increasing by an order of magnitude when BAV and AR z-scores were evaluated as covariates ($p = 2.27 \times 10^{-7}$; Fig 2B). This demonstrates a *TIMP3*-driven association between BAV and aortic enlargement in Turner syndrome. The quantile-quantile plots showed that there was no departure from observed vs. expected p-values (S1 Fig). Targeted exome sequencing of *TIMP3* in a replication cohort also showed a significant association of *TIMP3*variants with BAV and AR z-scores using SKAT-O (p = 0.038; Table 1).

Table 1. Replication Cohort Sequencing Results

Chromosome	Position	ID	REF	ALT	MAF cases	MAF controls
22	33253280	rs9862	T	C	50.00%	50.00%
22	33253292	rs11547635	C	T	14.29%	6.40%
22	33255244	rs149161075	C	T	3.57%	0.00%
Gene	**P value**	**N variants in Test**				
TIMP3	0.03734*	3				

Variants identified in Danish cohort by Sanger sequencing of *TIMP3* exons.
ID; rs identifier from dbSNP
MAF; minor allele frequency
P value; calculated by SKAT-O gene-based association test
*; reaches the significance threshold of <0.05
N; number

https://doi.org/10.1371/journal.pgen.1007692.t001

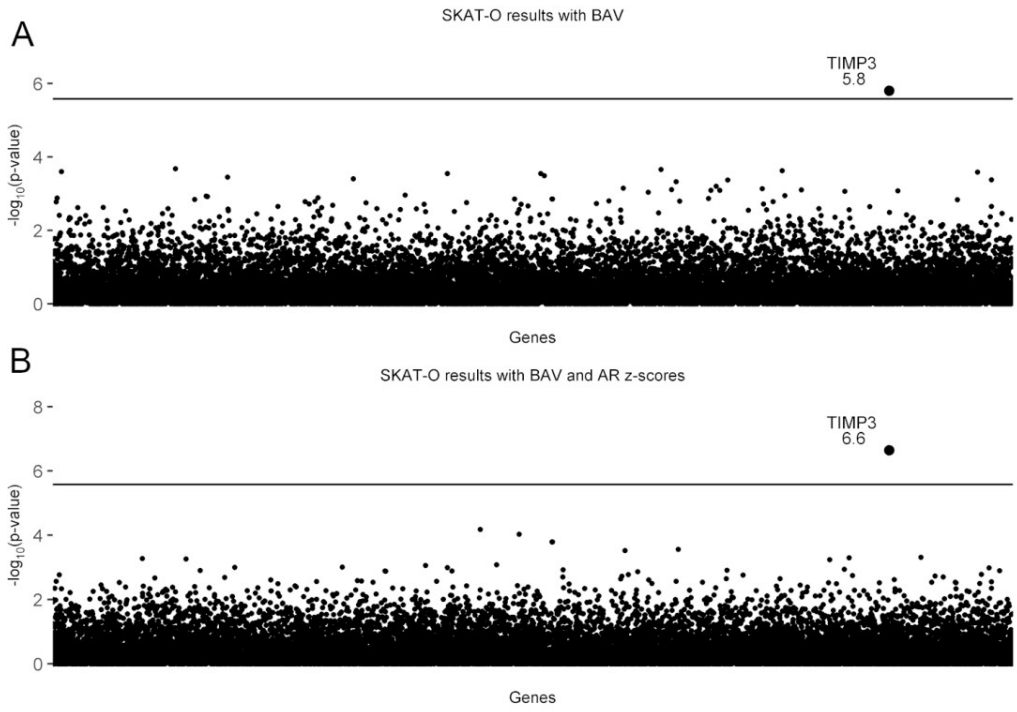


Fig 2. SKAT-O analysis shows that *TIMP3* variants are associated with BAV/TAD.
Manhattan plots showing the exome-wide significant finding that *TIMP3* variants are associated with BAV, and with AR enlargement as an indicator of TAD. The horizontal line is the threshold for exome-wide significance (based on testing 19,392 genes, the exome-wide significance p-value = 2.578×10^{-6}). *TIMP3* is the only gene that exceeds exome-wide significance. It is notable that no other genes approach the significance line. **A**) Shows the association with BAV as the sole predictor (p = 1.58×10^{-6}). **B**) Shows the association results for BAV and aortic root (AR) z-scores as covariates (p = 2.27×10^{-7}). The significance level for *TIMP3* increases nearly 10-fold when AR z-scores were added.

TIMP3 rs11547635 is the Major Autosomal Risk Allele

There were a total of four variants identified in *TIMP3* in the discovery cohort (Table 2). Of the four variants, rs11547635 was determined to be the SNP predominantly driving the association based on the increased allele frequency in cases compared to controls (p = 0.001, chi-squared) and evidence that the variant is deleterious based on the CADD score of 16.67. This is above the recommended deleterious significance cutoff of 15, which indicates that is in the top 5% of all damaging variants in the human genome. On the gene level, *TIMP3* has a GDI PHRED score of 0.449, placing in the top 10% of genes intolerant of mutations. The lead driving SNP encodes a synonymous C>T transition at p.Ser87 in exon 3. Another SNP, rs9862, which is a synonymous variant at p.His83 is always present along with the p.Ser87 variant in the BAV cases in this study. Importantly, these variants,

which have been studied in various types of cancer are associated with reduced TIMP3 plasma levels.[18–20] In combination the two variants disrupt two core ETS1 binding consensus sequences and prevent ETS1 binding, which is thought to be the basis of the reduction in expression.[20] Our discovery that known deleterious variants in *TIMP3* are significantly associated with BAV and TAD of the aortic root in Turner syndrome fits well with the known role for TIMPs in protection against aortopathy. Nearly 25% of our Turner syndrome cohort carry these SNPs, making them a significant risk genotype. The two additional *TIMP3* variants, rs149161075 and rs369072080, are rare and occur only in cases in this study.

rsID	Protein Change	CADD Score	Expected Allele Frequency (Alleles/ Total)	Observed Frequency in Cases (Cases/ Total)	Observed Allele Frequency in Cases (Alleles/Total)	Observed Frequency in Controls (Controls/ Total)	Observed Allele Frequency in Controls (Alleles/Total)
rs9862	p.His83 =	2.597	49.1% (32790/66734)	64.8% (57/88)	43.8% (77/176)	71.0% (71/100)	50.5% (101/200)
rs11547635	p.Ser87 =	16.67	7.1% (4735/66738)	23.8% (21/88)	11.9% (21/176)	6.0% (6/100)	3.5% (6/200)
rs149161075	p.Phe172 =	5.801	0.3% (203/66736)	2.3% (2/88)	1.1% (2/176)	0% (0/100)	0% (0/200)
rs369072080	p.Gly173 =	0.002	0% (0/66738)	1.1% (1/88)	0.6% (1/176)	0% (0/100)	0% (0/200)

All of the *TIMP3* variants identified through whole exome sequencing of subjects in our Turner discovery syndrome cohort are listed. The dbSNP rs identifier is listed, along with the consequence of the change, ExAC expected allele frequencies (European non-Finnish), CADD score, the number of subjects that had a BAV (case) or a normal valve (control), and the allele frequency of each variant, reported as percentages.

https://doi.org/10.1371/journal.pgen.1007692.t002

Table 2. Summary of the *TIMP3* Variants Associated with BAV/TAD
All of the *TIMP3* variants identified through whole exome sequencing of subjects in our Turner syndrome cohort are listed. The dbSNP rs identifier is listed, along with the consequence of the change, ExAC expected allele frequencies (European non-Finnish), CADD score, the number of subjects that had a BAV (case) or a normal valve (control), and the allele frequency of each variant, reported as percentages.

TIMP1 Copy Number Influences Aortopathy Risk

Analysis of all of the genes on Xp identified *TIMP1* as the top gene meeting our aortopathy criteria, which includes the potential for escape from X-inactivation, no Y chromosome or autosome homologues, and expression in the aorta. The list of all of the genes that met the criteria is shown in Table 3, ranked according to the likelihood that they could contribute to aortopathy. The list of all Xp genes and their characteristics can be found in S1 Table. *TIMP1*polymorphically escapes X inactivation,[21] has partial functional redundancy with *TIMP3*[22], and is highly expressed in the aorta with nearly 10-fold higher expression than

any of the other genes (GTExPortal). In addition, it is the only Xp gene that meets these criteria and has a known role in aortic valve development.[23] *TIMP1* is also the only gene on Xp with a known association with aortic aneurysms in both humans and mouse models. *Timp1* mouse models are susceptible to the development of aortic aneurysms[24, 25] and TIMP1 is known to be reduced in TAA in humans.[16, 26] Additionally, overexpression of Timp1 prevents aneurysm degradation and rupture in a rat model.[27] We therefore hypothesized that reduced copy number of *TIMP1* in Turner syndrome increases the risk for BAV/TAD. Using BAV as the only variable the analysis revealed that subjects with only one copy of *TIMP1* have a 4.50 increased odds of having a BAV than those who have greater than one copy (p = 0.0009, 95% CI = 1.9–11.8, Fig 3A). When BAV with TAD was studied as the outcome, having only one copy of *TIMP1* increased these odds substantially (OR = 9.76, p = 0.029, CI = 1.91–178.80, Fig 3B).

Table 3. Xp Genes that Meet Criteria for Being Involved in BAV/TAD

Gene	ensembl	cdsStart—cdsEnd*	X inactivation status	Pseudogene/ Y homolog	Aorta Expression (GTEx median RPKM)	Valve Development	Rank
TIMP1	**ENSG00000102265**	**X:47442814–47446090**	**Variable**	**No**	**423.6**	**Yes[26]**	**1**
RBM3	ENSG00000102317	X:48433568–48435474	Variable	No	48.96	unknown	2
UBA1	ENSG00000130985	X:47058201–47074328	Active	No	41.44	unknown	3
INE2	ENSG00000281371	X:15805712–15805712	Active	No	12.36	unknown	4
AP1S2	ENSG00000182287	X:15845447–15870647	Mostly Active	No	8.701	unknown	5
GEMIN8	ENSG00000046647	X:14027031–14039597	Mostly Active	No	6.69	unknown	6
CA5B	ENSG00000169239	X:15768146–15800787	Variable	No	5.198	unknown	7
NHS	ENSG00000188158	X:17653686–17750584	Mostly Active	No	4.852	unknown	8
TRAPPC2	ENSG00000196459	X:13732525–13738082	Mostly Active	No	4.289	unknown	9
CA5BP1	ENSG00000186312	X:15721474–15721474	Active	No	3.813	unknown	10
CTPS2	ENSG00000047230	X:16608915–16721025	Mostly Active	No	3.593	unknown	11
INE1	ENSG00000224975	X:47065254–47065254	Active	No	2.09	unknown	12
TCEANC	ENSG00000176896	X:13680627–13681683	Mostly Active	No	0.4852	unknown	13
PPEF1	ENSG00000086717	X:18725899–18845605	Variable	No	0.05353	unknown	14
GRPR	ENSG00000126010	X:16142076–16170768	Mostly Active	No	0.02957	unknown	15
FAM9C	ENSG00000187268	X:13056559–13061908	Active	No	0.005627	unknown	16

*Position on the X chromosome; cdsStart, start of coding sequence; cdsEnd, end of coding sequence, hg19

https://doi.org/10.1371/journal.pgen.1007692.t003

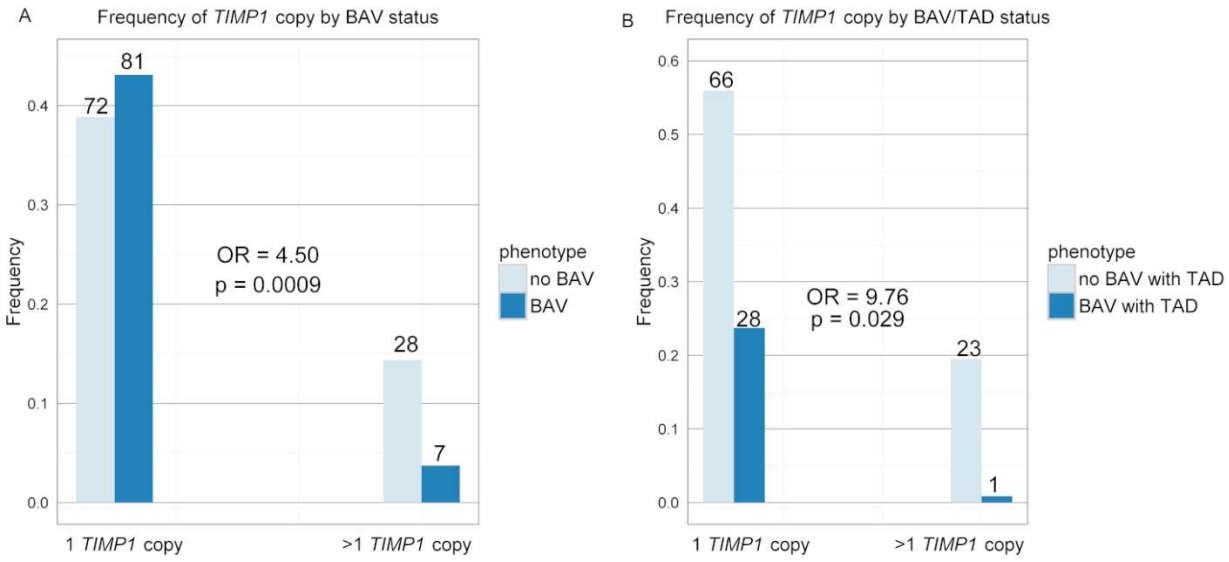


Fig 3. *TIMP1* copy number is associated with the risk of BAV and BAV with TAD.
Bar graph showing the frequency of BAV with or without TAD when only one copy of *TIMP1* is present compared to the frequency when there is more than one copy of *TIMP1*. **A**) The dark blue bars represent individuals with a diagnosis of BAV and the light blue bars represent individuals without a BAV. **B**) The dark blue bars represent individuals with a diagnosis of BAV with TAD and the light blue bars represent individuals without a BAV with TAD. For both a logistic regression model was run, with *TIMP1* copy number as the categorical predictor and the phenotype as the response variable. Odds ratio and p-value were calculated and showed that having a single copy of *TIMP1*significantly increased the odds of having a BAV and a BAV with TAD.

TIMP3 and *TIMP1*-associated risk is specific to the aorta

To determine the specificity of the association between *TIMP3* rs11547635 and *TIMP1* copy number for having BAV/TAD or other phenotypic features, we compared cases with or without rs11547635. Height, weight, blood pressure, body surface area, the presence of webbed neck, broad chest, primary ovarian insufficiency, hypertension, or lymphedema occurred with equal frequency in subjects with or without the rs11547635 SNP (Table 4). On the other hand BAV, BAV with TAD, coarctation of the aorta, and any aortic disease occurred with significantly higher frequency in the group with rs11547635, indicating that it is specifically associated with aortopathy. *TIMP1* copy number associations were similar but also included systolic blood pressure, lymphedema and webbed neck (Table 5).

Table 4. Attributes of Turner syndrome subjects with or without TIMP3 rs11547635

Phenotype—Continuous	N	Mean with rs11547635	SD	Mean without rs11547635	SD	P value		
Height (cm)	175	140.65	23.60	142.67	15.66	0.58		
Weight (kg)	168	56.07	21.10	55.15	20.30	0.84		
BSA (m^2)	165	1.42	0.39	1.42	0.32	1		
BP, systolic	161	116.35	15.23	116.71	17.13	0.93		
BP, diastolic	161	68.39	11.97	72.20	10.59	0.12		
Phenotype—Categorical	**N**	**Affected with rs11547635**	**Unaffected with rs11547635**	**Affected without rs11547635**	**Unaffected without rs11547635**	**Odds Ratio**	**95% CI**	**P value**
Lymphedema	188	4	23	40	121	0.53	0.2–1.6	0.37
Broad chest	188	7	20	51	110	0.76	0.3–1.9	0.71
Webbed neck	188	16	11	69	92	1.94	0.8–4.4	0.17
POI[t]	154	12	79	10	53	0.81	0.3–2.0	0.82
Hypertension	178	11	16	60	91	1.04	0.5–2.4	0.92
Any Dissection	188	1	26	5	156	1.20	0.1–10.7	1.000[F]
Coarctation	188	12	15	33	128	3.10	1.3–7.3	0.014*
BAV	188	21	6	66	95	5.04	1.9–13.2	0.0008**
TAD[#]	118	9	9	29	71	2.45	0.9–6.8	0.140
BAV with TAD[#]	118	9	9	20	80	4.00	1.4–11.4	0.010[F]
BAV without TAD[#]	118	7	9	31	87	0.46	0.2–1.3	0.23[F]
Any aortic risk factor[~]	188	22	5	79	82	4.57	1.6–12.7	0.004**

Categorical p-value: chi-squared with yates correction.

[F] Fishers exact test, Chi-square is calculated only if all expected cell frequencies are greater than or equal to 5

Continuous p-value: t-Test: Two-Sample Assuming Equal Variances

N, total number of subjects with data available for the category

[t] POI, primary ovarian insufficiency. Excluded individuals under the age of 13 years.

[#] Aortic root or ascending aorta (z-score > 1.9)

[~] Aortic risk factors include BAV, coarctation, dilated aortic root or ascending aorta (z-score > 1.9)

[$] For this group, the sample size was small and there were no affected individuals, so an accurate odds ratio could not be calculated.

*Significance level <0.05

**Significance level <0.005

https://doi.org/10.1371/journal.pgen.1007692.t004

Table 5. Attributes of Turner syndrome subjects with 1 or >1 copy of TIMP1

Phenotype—Continuous	N	Mean with 1 *TIMP1* copy	SD	Mean with >1 *TIMP1* copy	SD		P value	
Height (cm)	175	142.58	15.55	141.53	22.85		0.75	
Weight (Kg)	168	55.79	20.12	53.22	22.13		0.52	
BSA (m^2)	165	1.43	0.32	1.40	0.40		0.58	
BP, systolic	161	118.19	16.66	109.69	16.65		0.014*	
BP, diastolic	161	72.30	10.25	68.76	13.33		0.11	
Phenotype—Categorical	**N**	**Affected with 1 *TIMP1* copy**	**Controls with 1 *TIMP1* copy**	**Affected with >1 *TIMP1* copy**	**Controls with >1 *TIMP1* copy**	**Odds Ratio**	**95% CI**	**P value**
Lymphedema	188	43	110	1	34	13.29	1.8–100.2	0.003**
Broad chest	188	48	105	10	25	1.14	0.5–2.6	0.92
Webbed neck	188	77	76	8	27	3.42	1.5–8.0	0.006*
POI[t]	154	78	49	13	14	1.71	0.7–4.0	0.29
Hypertension	178	61	84	10	23	1.76	0.8–4.0	0.23
Any Dissection[$]	188	6	147	0	35	-	-	-
Coarctation	188	44	109	1	34	13.72	1.8–103.4	0.003**
BAV	188	81	72	7	28	4.50	1.9–11.8	0.0009**
TAD[#]	118	34	60	4	20	2.83	0.9–9.0	0.11
BAV with TAD[#]	118	28	66	1	23	9.76	1.3–75.8	0.019*
BAV without TAD[#]	118	27	67	4	20	2.01	0.6–6.5	0.34
Any aortic disease[~]	188	90	63	11	24	3.12	1.4–6.8	0.006**

Categorical p-value: chi-squared with yates correction.

[F] Fishers exact test, Chi-square is calculated only if all expected cell frequencies are greater than or equal to

5Continuous p-value: t-Test: Two-Sample Assuming Equal Variances

N, total number of subjects with data available for the category

[t] POI, primary ovarian insufficiency. Excluded individuals under the age of 13 years.

[#] Aortic root or ascending aorta (z-score > 1.9)

[~] Aortic risk factors include BAV, coarctation, dilated aortic root or ascending aorta (z-score > 1.9)

[$] For this group, the sample size was small and there were no affected individuals, so an accurate odds ratio could not be calculated.

*Significance level <0.05

**Significance level <0.005

https://doi.org/10.1371/journal.pgen.1007692.t005

Synergistic Effect of TIMP1 and TIMP3 Variation in Aortopathy

We investigated the combinatorial effect of *TIMP1* and *TIMP3* variation on the outcome of BAV alone, and BAV with TAD. This analysis shows that the combination of having only one copy of *TIMP1* and being a carrier of *TIMP3* rs11547635 specifically increases the odds for having a BAV by nearly twenty-fold (OR = 18.00, 95% CI = 5.19–74.89, p<0.001) and also for having a BAV with TAD (OR = 12.86, 95% CI = 2.57–99.39, p = 0.004) compared to the group with no rs11547635 and >1 *TIMP1* (Table 6).

Table 6. Combinatorial effects of TIMP1 copy number and TIMP3 SNP rs11547635 on the outcome of BAV or BAV with TAD

Groups, outcome of BAV	N	N affected	% affected	OR	95% CI	P value
no rs11547635 & >1 *TIMP1* copy	33	6	18.18%	1.00**	-	-
no rs11547635 & only 1 *TIMP1* copy	128	60	46.88%	3.97	1.63–11.22	0.005
yes rs11547635 & >1 *TIMP1* copy*	2	1	50.00%	-	-	-
yes rs11547635 & only 1 *TIMP1* copy	24	20	80.00%	18.00	5.19–74.89	<0.001
Groups, outcome of BAV with TAD	**N**	**N affected**	**% affected**	**OR**	**95% CI**	**P value**
no rs11547635 & >1 *TIMP1* copy	22	2	9.09%	1.00**	-	-
no rs11547635 & only 1 *TIMP1* copy	78	20	25.64%	3.45	0.89–22.80	0.115
yes rs11547635 & >1 *TIMP1* copy*	2	0	0.00%	-	-	-
yes rs11547635 & only 1 *TIMP1* copy	16	9	56.25%	12.86	2.57–99.39	0.004

Logistic regression model for the combinatorial effect of have both rs11547635 and one copy of *TIMP1* on the outcome of having a BAV with TAD.
BAV and TAD are defined as having a BAV and at least one z-score $\geq$ 1.9 (rounded to the 10th decimal place); those without an AR or AAO measurement were excluded.
*For this group, the sample size was too small, so an accurate odds ratio could not be calculated.
N; total number of subjects.
N affected; number of subjects with BAV or BAV and TAD.
** Reference group

https://doi.org/10.1371/journal.pgen.1007692.t006

Discussion

Turner syndrome, like all genetic syndromes, is characterized by a primary inherent defect that sensitizes downstream modifier genes to breach a pathologic threshold. Thus, a single triggering event is capable of unleashing a myriad of phenotypic variations. Consistent with this disease model, we found that in Turner syndrome hemizygosity of *TIMP1* due to lack of a complete second X chromosome is associated with genetic variation of its paralogue, *TIMP3* on chromosome 22, synergistically heightening the risk for BAV and TAD, which is the first sign of aneurysm formation. Given the detailed understanding of the fundamental role of MMPs in thoracic aortic disease, the results of this study have clear biological relevance. In the euploid population there is a significant reduction in TIMP1 and TIMP3 expression in BAV-associated TAA and a highly significant increase in MMP2 and MMP9, which are both regulated by TIMP1 and TIMP3.[16] This results in a considerable MMP/TIMP imbalance in aneurysms compared to control aortas.

We propose that hemizygosity for *TIMP1* is the X chromosome basis for increased susceptibility for BAV and aortopathy in Turner syndrome. This coupled with a SNP-driven decrease in TIMP3

expression synergistically increases risk for both BAV and BAV with TAD. This is consistent with our hypothesis that a gene or genes on Xp interact with autosomal variants that are benign unless expressed on a genetically sensitized background such as that in Turner syndrome. The inherent decrease in TIMP1 in Turner syndrome subjects missing a complete second copy of the X chromosome sensitizes those individuals to decreased TIMP3 expression. In addition, a global methylation profile for Turner syndrome found that the Turner syndrome X chromosome has a unique methylation pattern when compared to the X chromosome of euploid males.[11] Notably, *TIMP1* tends to be hypermethylated in Turner syndrome,[28] which suggests that the expression level may be decreased even beyond the reduction in copy number.

Importantly, *TIMP1* and *TIMP3* have functional redundancy in the aorta. Both exercise inhibitory control over MMP2 and MMP9, which are the two MMPs associated with degradation of the aortic wall. We propose that decreased TIMP1 expression due to a reduction in copy number sensitizes the aorta to MMP-induced damage, but protection is conferred by the expression of TIMP3. Decreased expression of both negates that protection making the aortic wall vulnerable to degradation which can lead to TAD and aneurysm. In addition, TIMPs 1 and 3 are expressed in the aortic valve, where they play a role in valve remodeling,[23] which is a critical activity in the development of the tricuspid aortic valve. This fundamental link between BAV pathogenesis and downstream TAD provides a previously unrecognized mechanism for the heightened risk for aortopathy in Turner syndrome. In a study of 18 women with TS and aortic dissection, 6 cases were available for biochemical analysis, and that study showed a skewed ratio of collagen I to collagen III (normally 30:70%) with 60% collagen I and only 30% collagen III,[29] which could well be the end result of an altered MMP/TIMP activity.

As with all studies of this nature there are some limitations and caveats. The exome sequencing was done on DNA isolated from peripheral blood, so the molecular karyotypes reflect the chromosome composition in that tissue. It is possible that the karyotype in other tissues such as the developing heart may differ, particularly with respect to mosaicism. In addition, our analyses did not include

potential effects of the autosomal rearrangements found in some of the study subjects. These were genetically heterogeneous and often in single individuals, so it is unlikely that they would significantly affect the results of this study. Another limitation is that this study did not assess any potential influence of maternal genetic effects, nor did we assess the parent-of-origin of the retained X chromosome.

There is no clear explanation for the strikingly higher prevalence of aortopathy in euploid men compared to women. And, the larger questions regarding the role of the sex chromosome genes in the differential susceptibility to common diseases has received little attention. Bellott and colleagues proposed that dosage differences between X chromosome genes and homologous ancestral genes retained on the Y chromosome may account for phenotypic differences between men and women.[30] Our data supports another model where expressed genes that escape inactivation on the second X chromosome and that are also absent from the Y chromosome (like *TIMP1*) play a role in the frequently observed sex bias in disease.

In conclusion, we propose that aortopathy in Turner syndrome results from an inherent dysregulation of the TIMP/MMP ratio. This imbalance increases risk for both congenital cardiovascular defects and later onset aortic disease. Beyond Turner syndrome, the lack of a second copy of *TIMP1* in euploid males may also explain the increased risk for BAV/TAD compared to euploid females. The findings of this study represent a significant advance in the understanding of the mechanisms underlying aortopathy in Turner syndrome.

Materials and Methods

Ethical Approval

The Turner syndrome cohort was accessed from the National Registry of Genetically-Triggered Thoracic Aortic Aneurysms and Related Conditions (GenTAC).[31] GenTAC study subject recruitment was approved by the institutional review board for each member of the GenTAC investigative team, and informed consent for participation in associated research studies was

obtained for each study subject. The project was approved by the Oregon Health & Science University institutional review board. The Danish cohort was approved by the Central Denmark Region Ethical Scientific Committee (#2012-500-12) and registered at ClinicalTrials.gov (#NCT01678274).

Study Populations

GenTAC spent a decade recruiting study subjects with conditions related to thoracic aortic aneurysms, including collection of biospecimens, rigorous evaluation and documentation of clinical data, and collection of follow-up data for longitudinal studies. The majority of subjects enrolled in GenTAC had aorta imaging studies that provide information on aortic dimensions and evaluation of aortic valve status. All images, such as echocardiograms, CT and MRT studies were collected clinically, but transferred to the GenTAC imaging core (ICORE) for re-evaluation by a single cardiac imaging expert for consistency of measurements and interpretation.[32] The discovery cohort for this study was composed of Turner syndrome study subjects of Northern European (non-Finnish) descent. Inclusion criteria included a diagnosis of Turner syndrome, self-reported race as white, ethnicity as non-Hispanic, evaluation for a diagnosis of BAV, and availability of aortic dimension measurements and body morphometrics. The diagnosis of BAV was based on clinical images and interpretations. An additional 53 study subjects from a prospective study in Denmark were used as an independent replication cohort.[33, 34]

Phenotyping

For the purposes of this study we defined aortopathy (cases) as those having a BAV with or without TAD. In keeping with clinical norms for Turner syndrome, a thoracic aortic dimension z-score ≥ 1.9 was used as the definition of TAD as an indicator of aneurysm formation. All study subjects were confirmed for a diagnosis of Turner syndrome based on either clinical karyotype or exome sequence-based karyotyping. Subjects were phenotyped for presence of a BAV or a normal aortic

valve. Our final Turner syndrome discovery cohort was composed of 88 cases (Turner syndrome with BAV) and 100 controls (Turner syndrome with no BAV). Within this cohort, 113 subjects had aortic root (AR) dimensions and 106 subjects had ascending aorta (AAO) dimensions. For the replication cohort 14 had a BAV and 39 had a normal aortic valve. For all subjects AAO and AR diameters were converted into z-scores using methodology that was specifically developed for children and adults with Turner syndrome to correct for the altered longitudinal growth in Turner syndrome.[35] Briefly, the regression equations and coefficients were used to calculate expected aortic dimensions based on body surface area (BSA, Haycock formula) for each individual in the study with a measurement (Eqs 1&2). The z-scores were calculated by comparing expected aortic dimensions to actual aortic dimensions and incorporating the mean squared error (MSE; Eq 3).[35] Expected aortic dimension data points and lines were generated for each z-score.

*Equation*s:

1) Aortic Root equation: $(\text{expected})^2 = (1.035 + (0.589 * \text{BSA}) + (-0.129 * \text{BSA}^2))^2$

2) Ascending Aorta equation: $(\text{expected})^2 = (0.942 + (0.593 * \text{BSA}) + (-0.122 * \text{BSA}^2))^2$

3) Z-score equation: = (√ (actual dimension (cm) - √ (expected dimension (cm)) / (√ (MSE))

The BSA (m^2) vs. AR or AAO (cm) for BAV cases (triangle) and BAV controls (square) were plotted. Overlaid on the same plot are the polynomial trend lines for z = 0, z = 1, z = -1, z = 2, z = -2, z = 3, z = -3 (S2 Fig).

Whole Exome Sequencing, Quality Control and Data Cleaning

In total, 215 genomic DNA samples isolated from peripheral blood were submitted for exome sequencing and the exome capture kit Roche Nimblegen SeqCap EZ was used to prepare the sequencing libraries. Whole exome sequencing (WES) was performed by the NHLBI Resequencing & Genotyping Service at the University of Washington (D. Nickerson, US Federal Government contract number HHSN268201100037C). In summary, 16 samples failed post-sequencing QC and 199 samples passed post-sequencing QC. The average read depth for the targeted exome was 71X,

with 86% of the target regions covered at greater than 20X. Reads were mapped to the hg19 UCSC genome build using the Burrows-Wheeler aligner, version 0.7.10. Variants were called using the GATK best practices pipeline, where in the 199 samples, 195,034 variants were called. BAM files and VCF files were transferred to the Maslen lab for evaluation.

Data cleaning and filtering was performed using PLINK v1.90b3g[36], which 1) removed any variants with less than 99% genotyping rate, where 6,815 variants were removed; 2) removed individuals with more than 5% missing genotypes, where no individuals were removed; 3) excluded markers that fail the Hardy-Weinberg equilibrium test using a threshold of $1.0x10^{-6}$, where 2,334 variants were removed. A principal components analysis (PCA) was performed using the R package SNPRelate to calculate the eigenvectors (EVs) for each subject.[37] Data were prepared for PCA analysis by taking common SNPs (MAF >5%) and pruning out SNPs in linkage disequilibrium with an $r^2 > 0.2$, stepping along five SNPs at a time within 50kb windows. We plotted EV1 vs EV2 to look for population outliers (S3A Fig). Population outliers were removed and the analysis was repeated a total of four times until no more outliers remained (S3B Fig). In total, 11 subjects were detected at EV1 < -0.3 and EV2 > 0.3 and were removed from the dataset. Additionally, we use the first three eigenvectors as covariates in most downstream analysis. The final dataset contained 185,885 variants across 188 subjects, providing a total genotyping rate of 0.998084.

Gene-Based Statistical Analyses

To enhance the probability of identifying an exome-wide significant signal a gene-based burden test, the optimal sequence kernel association test (SKAT-O), was used to evaluate the data.[38] This analysis clusters variants into genes for a gene by phenotype analysis, which improves signal strength for exome data from smaller cohorts as it reduces the multiple testing burden. This state-of-the-art approach is particularly useful for studies of rare disorders such as Turner syndrome. The 185,885 variants which passed QC from the WES pipeline were assigned to their respective genes using hg19_refGene. Variants were allowed to be in more than one gene since the test compares

gene burden in the same gene, not between different genes. All analyses included the first three principal component eigenvalues as covariates to adjust for any underlying population structure. First, SKAT-O was used to test for an association with the dichotomous BAV status. Second, SKAT-O was used to test for an association with BAV and aortic diameter z-scores as a proxy for TAD evaluated as a continuous variable. For each analysis, a quantile-quantile (Q-Q) plot was generated to look for departure of the observed p-values from the expected p-values.

Variant Annotation and Validation

Combined Annotation Dependent Depletion (CADD) scores were used as a tool for scoring the deleteriousness of the genetic variants identified in exome sequencing data. PHRED-scaled CADD scores integrate multiple annotations into a single metric that outperforms other commonly used algorithms of this type. A CADD score ≥20 indicates that a variant is among the top 1% most deleterious variants in the human genome. We used the recommended cutoff score of ≥15 as our threshold for considering a variant to be likely deleterious. The allele frequency of each variant was queried in the Exome Aggregation Consortium (ExAC) database of exome data from over 60,000 unrelated individuals, from which we used the European non-Finnish population.[39] All variants with alleles that were overrepresented in cases were validated by Sanger sequencing. For the replication cohort, we performed targeted Sanger sequencing of all *TIMP3* exons and followed the same SKAT-O association test as described above.

Second Sex Chromosome Status Determination

X and Y chromosome information from the WES data was used to assess the presence of any second sex chromosome. X and Y SNP plots were generated for each study subject and compared to control reference plots to define the second sex chromosome status for each individual.[40] Alternate allele frequencies from the exome variant calls were used to create SNP plots. Briefly, the alternate allele frequencies were calculated for all variants on the X chromosome and sorted by

position for each subject. In R, scatter plots were generated and evaluated for the presence of a second X chromosome. This was repeated for the Y chromosome and the Integrative Genome Viewer (IGV) was used to confirm the presence of Y chromosome reads.[40] Reference plots of a control female with 46,XX karyotype, a control female with 45,X karyotype, and a control male with 46,XY karyotype were generated (S4A Fig). We then generated X and Y chromosome plots for each subject in this study.

In these plots, the X-axis is sorted by position on the X chromosome and the Y-axis is the alternate (ALT) allele frequency. As expected for a 46,XX karyotype, some SNPs are homozygous for the ALT allele (1.0), homozygous for the reference (REF) allele (0.0), or heterozygous for the ALT/REF allele (0.5). In contrast, a 45,X karyotype only has SNPs that are homozygous for the ALT allele, or homozygous for the REF allele because only one copy is present. The 46,XY karyotype looks similar to the 45,X plot, but has SNPs heterozygous for the ALT/REF allele clustered in the captured pseudoautosomal (PAR) region. The presence of any Y chromosome material was confirmed using IGV. Available clinical karyotypes were compared to molecular karyotypes generated from the SNP data and basic second sex chromosome status groups were created to categorize the study subjects. While the majority of subjects were true monosomy 45,X, examples of other karyotypes included Xp deletions, Xq deletions, Xq isochromosomes, and X chromosome rings for their second X chromosome (S4B Fig); mosaicism for the second X chromosome, either 45,X/46,XX or 45,X/47,XXX, or mosaicism for Y chromosome material (S4C Fig), although we were unable to quantify the Y mosaicism level based on the plots.

While most plots were straight forward in their interpretation, some were more complicated. In those cases, the clinical karyotype was relied upon. To assess the mosaicism observed in a large number of subjects, a model was created to predict the percent 45,X mosaicism based on alternate allele frequencies (S5 Fig). The equations from each model were used, where y is the percent 45,X mosaicism and x is the alternate allele frequency. The average of the upper and lower predicted values was used as the final estimate of 45,X mosaicism. The molecular karyotypes and

estimated *TIMP1* copy number based on the percentage of cells with a second X chromosome are shown in Table 7.

Table 7. Karyotypes of the TS cohort

Karyotype	N	*TIMP1* copy number
45,X	109	1.0
45,X/46,XY	15	1.0
46,X,i(Xq)	13	1.0
45,X/46,X,i(Xq)	7	1.0
45,X[50%]/47,XXX[50%]*	4	2.0
45,X [50%]/46,X,ring(X)[50%]	2	1.5
46,X,ring(X), small	2	1.0
45X/46,X,i(Xq)/47,XXX*	1	2.0
46,X,ring(Xp11.1q28)[11%]/46,XX[89%]*	1	1.9
45,X[30%]/46,XX[70%]	1	1.7
45,X[30%]/46,XX[70%]	1	1.7
45,X[32%]/46,XX[68%]	1	1.7
45,X[35%]/46,XX[65%]	1	1.7
45,X[41%]/46,XX[59%]	1	1.6
45,X/46,X,del(Xq21.1)*	1	1.5
45,X[55%]/46,XX[45%]	1	1.5
45,X[65%]/46,XX[35%]	1	1.4
45,X[63%]/46,X,del(Xq11.23)[37%]	1	1.4
45,X[64%]/46,X,del(Xq22q24)[36%]	1	1.4
45,X[72%]/46,XX[28%]	1	1.3
45,X[75%]/46,XX[25%]	1	1.3
45,X[74%]/46,del(Xq13.1)[26%]	1	1.3
45,X[75%]/46,X,ring(X)[25%]	1	1.3
45, X[80%]/46,X,ring(X)[20%]	1	1.2
45,X[82%]/46,XX[18%]	1	1.2
45,X[84%]/46,XX[16%]	1	1.2
45,X[85%]/46,XX[15%]	1	1.2
45,X[85%]/46,XX[16%]	1	1.2
45,X[81%]/46,X,psuidic(Xq21)[19%]*	1	1.2
45,X[82%]/46,del(Xq22)[18%]*	1	1.2
45,X[82%]/46,X,del(Xp22.3p11.4)[18%]	1	1.2
45,X[83%]/46,X,ring(X)[17%]	1	1.2
45,X[85%]/46X,ring(X)[15%]	1	1.2
45,X[88%]/46,XX[12%]	1	1.1
45,X[88%]/46,X,del(Xq13.1)[12%]	1	1.1
45,X,add(15)(p11.2)*	1	1.0
45,X/46,X,+mar*	1	1.0
45,X[20%]/46,X,i(Xq)[80%]*	1	1.0
45,X[82%]/46,X,del(Xp)[18%]	1	1.0
45,X [86%]/ 46,X +mar [13%]*	1	1.0

*Clinical karyotype

N; number of subjects

https://doi.org/10.1371/journal.pgen.1007692.t007

Genes on Xp were evaluated to identify candidates likely to contribute to aortopathy. We hypothesized that an aortopathy gene would be found on Xp, would escape X inactivation in euploid females,[41–43] would be expressed in the aortic wall, and would not be a pseudogene, or have a Y homologue.

Statistical Analysis

To calculate the magnitude of the association between BAV status and aortic z-score, a linear regression model was fit where BAV was the predictor and aortic z-score was the response variable. This was performed separately for both AR z-score and AAO z-score. The mean differences and 95% confidence intervals were generated to accompany p-values. Boxplots for each AR and AAO z-scores were plotted against BAV status.

To investigate if the *TIMP3* paralog *TIMP1* was associated with BAV, a general logistic regression model was performed where *TIMP1* copy number was the categorical predictor, 1 copy and >1 copy of *TIMP1* were the variables, and BAV status or BAV with TAD was the response variable with no BAV serving as the reference. Odds ratios and 95% confidence intervals were generated to accompany p-values.

To investigate the combination of the *TIMP3* variant rs11547635 and *TIMP1* as risk factors for the presence of a BAV or BAV with TAD, four groups were formed: 1) no *TIMP3* rs11547635 and >1 copy of *TIMP1*, 2) with *TIMP3* rs11547635 and >1 copy of *TIMP1*, 3) no *TIMP3* rs11547635 and only 1 copy of *TIMP1*, and 4) with *TIMP3* rs11547635 and only 1 copy of *TIMP1*. Separate general logistic regression models were created to compare these four groups in order to determine their associations with BAV, or the combination of BAV and TAD. Odds ratios and 95% confidence intervals were generated to accompany p-values.

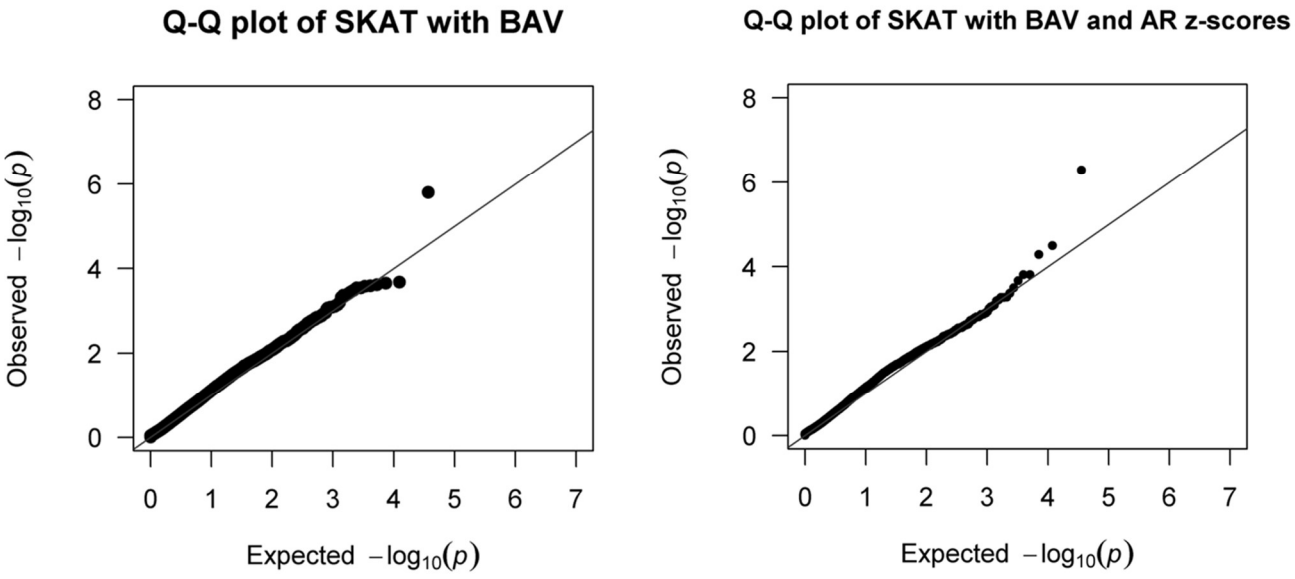
Other physical attributes of Turner syndrome were studied to determine if any were also associated with the *TIMP3* rs11547635 risk allele. Continuous variables (height, weight, body surface area, systolic blood pressure, and diastolic blood pressure) were analyzed using a Student's two-sample t-

test, where the means of those with or without *TIMP3* rs11547635 were compared. Categorical variables (lymphedema, broad chest, webbed neck, primary ovarian insufficiency, hypertension, coarctation of the aorta, bicuspid aortic valve, and any aortic risk factor) were analyzed using a Chi-squared test with Yate's correction or Fisher's exact test as appropriate. The same analysis was done using *TIMP1* copy number as the variable.

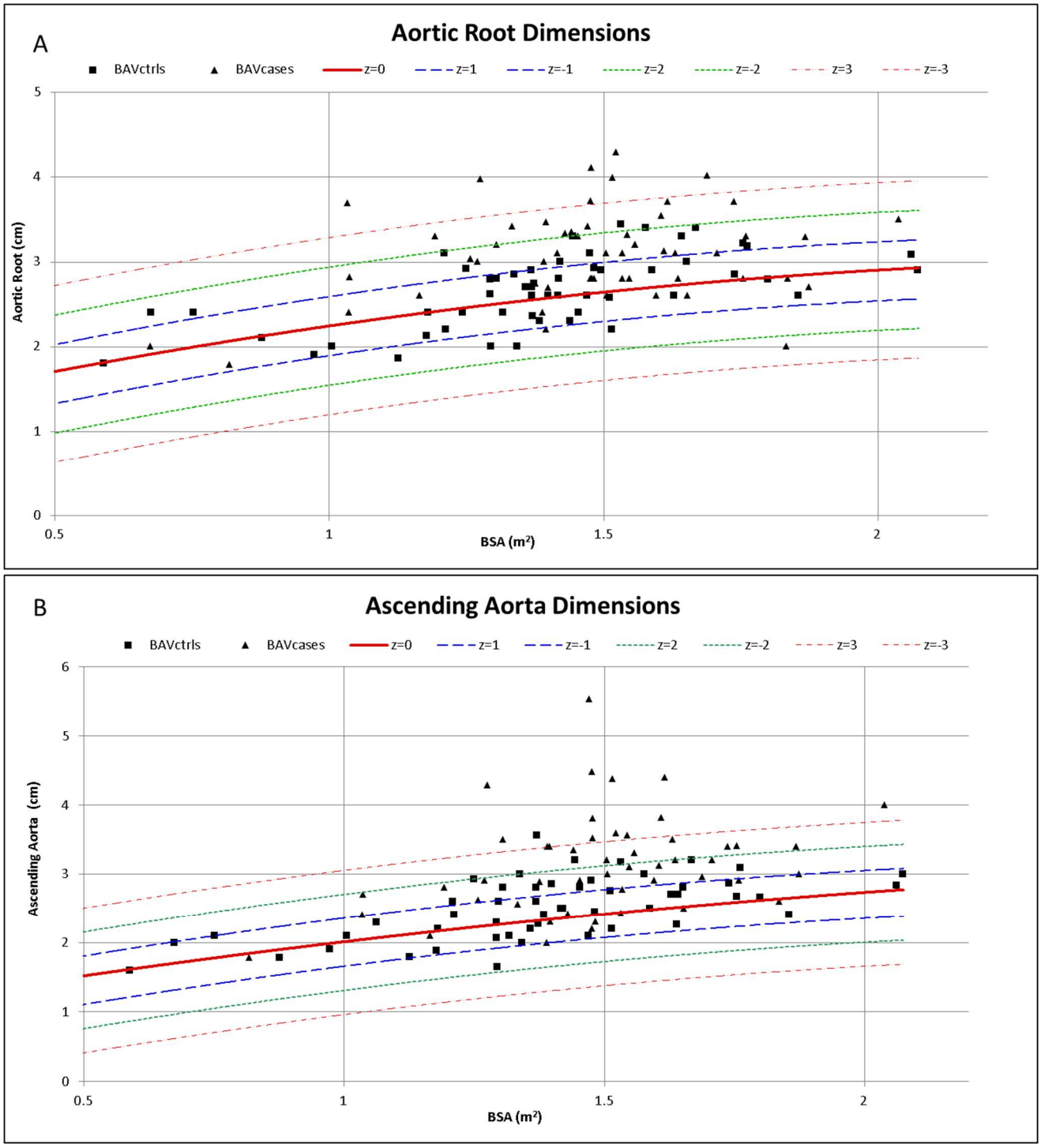
Acknowledgments

Thanks to Drs. Brian O'Roak and Deidre Krupp for input on bioinformatics analyses, Mr. Jacob Gutierrez for technical assistance, and Drs. Lucia Carbone (OHSU), Debbie Nickerson (UW), and Elaine Ostrander (NIH/NHGRI) for helpful discussions regarding the manuscript. We also thank the study subjects and their families for their participation, and the Tuner Syndrome Society of the US for their participation. We thank the study subjects and their families and the Turner Syndrome Society of the US and the Turnerforeningen I Danmark for their participation. The GenTAC investigators include F. Asch, J. Bavaria, P. Desvigne-Nickens, R. Devereux, H. Dietz, K. Eagle, J. Habashi, K. Holmes, B. Kroner, J. Kushner, S. LeMaire, C. Maslen, D. Milewicz, K. Milewski, S. Morris, S. Prakash, R. Pyeritz, W. Ravekes, M. Roman, R. Shohet, M. Silberbach, H. Song, E. Tolunay, H. Tseng, and J. Weinsaft.

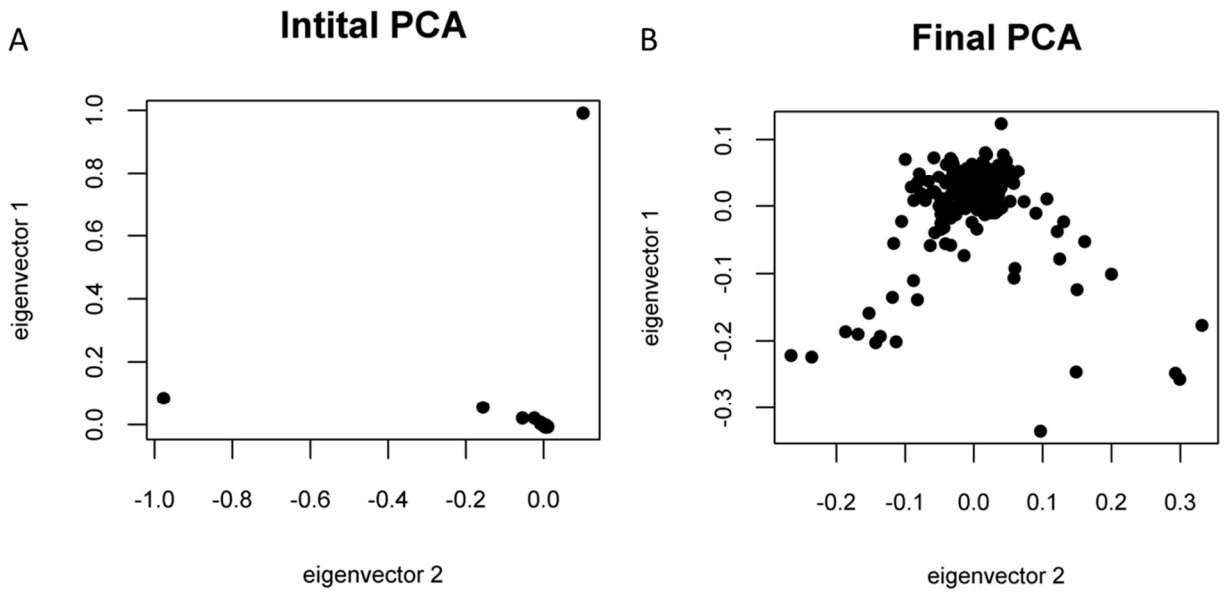
Supporting Information



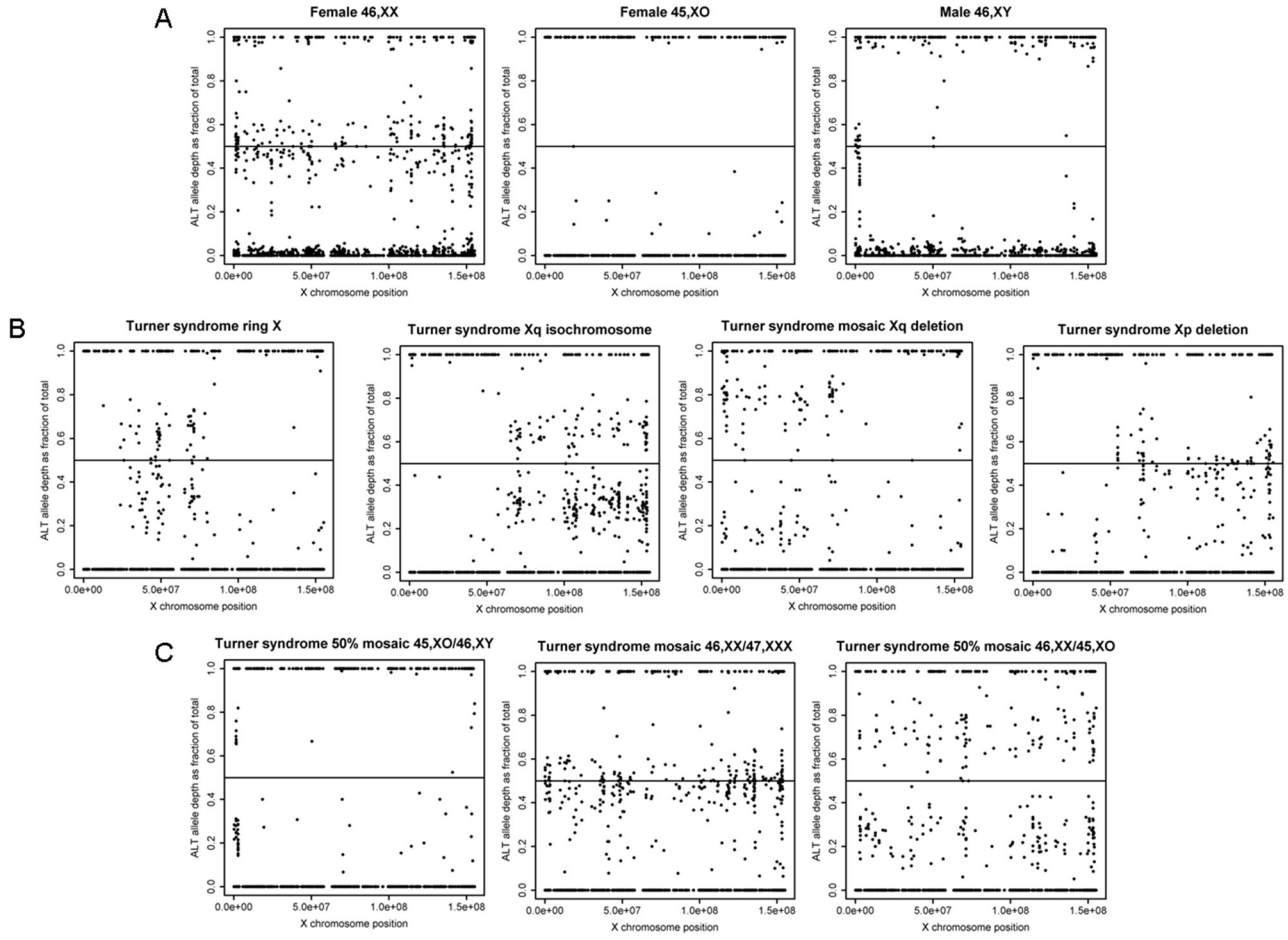
S1 Fig. Quantile-Quantile (Q-Q) Plots for the SKAT-O Analyses. Q-Q plots for SKAT-O analysis of BAV and BAV with AR Z-scores, which shows no significant deviation from the normal distribution.



S2 Fig. Calculation of Aortic Z-Scores. Plot of BSA (m2) vs. AR or AAO (cm) for BAV cases 10 (triangles) and BAV controls (squares) and polynomial trend 11 lines for expected aorta dimensions for each z-score. A) Plot for aortic root dimensions. B) Plot for ascending aorta dimensions.

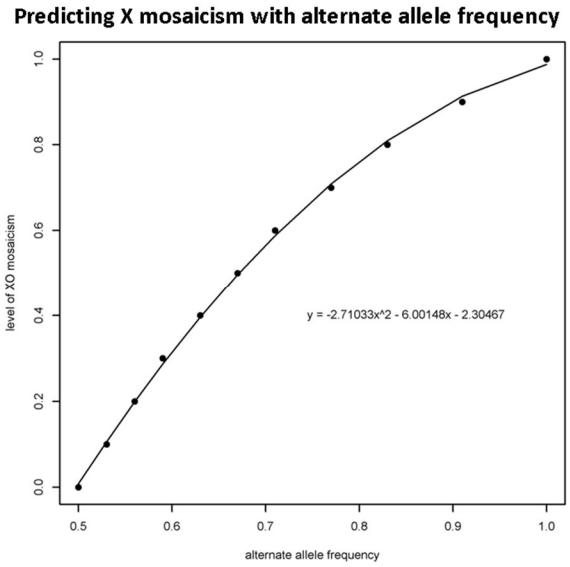


S3 Fig. Principle Component Analysis. Principal Component Analysis (PCA) plots of WES samples. A) PCA analysis on all 199 subjects in the study, where eigenvector 1 is plotted against eigenvector 2. B) Final PCA plot after samples were removed due to being population outliers. A total of 11 subjects were removed and a total of 188 subjects remained in the study.

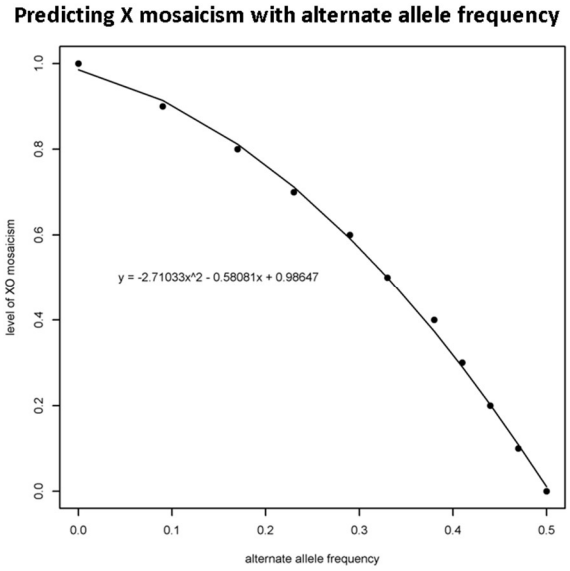


S4 Fig. X and Y SNP Plots for Second Sex Chromosome Status Determination. X chromosome SNP plots examples. A) Plots of a known controls representing 45,X, 46,XX, and 46,XY. B) Examples plots of Turner syndrome subjects with ring X, iso Xq, mosaic Xq deletion, and Xp deletion for their second X chromosome. C) Example plots of mosaic 45,X/46,XY, 45,X/47,XXX, and 45,X/46,XX.

45,X mosaicism	expected alt/total alleles	lower alt allele frq
1.00	0alt / 10	0.00
0.90	1alt / 11	0.09
0.80	2alt / 12	0.17
0.70	3alt / 13	0.23
0.60	4alt / 14	0.29
0.50	5alt / 15	0.33
0.40	6alt / 16	0.38
0.30	7alt / 17	0.41
0.20	8alt / 18	0.44
0.10	9alt / 19	0.47
0.00	10alt / 20	0.50



45,X mosaicism	expected alt/total alleles	upper alt allele frq
1.00	10alt / 10	1.00
0.90	10alt / 11	0.91
0.80	10alt / 12	0.83
0.70	10alt / 13	0.77
0.60	10alt / 14	0.71
0.50	10alt / 15	0.67
0.40	10alt / 16	0.63
0.30	10alt / 17	0.59
0.20	10alt / 18	0.56
0.10	10alt / 19	0.53
0.00	10alt / 20	0.50



S5 Fig. Modeling of Percent Mosaicism from Alternate Allele Counts. A model to predict percent mosaicism from alternate allele counts. The table on the left is the expected alternate allele counts and allele frequencies (frq) for each level of X mosaicism. The plots on the right are the corresponding fitted model using this data, where the line is the fitted trend line with its equation. These equations were used to estimate X mosaicism from observed alternate allele frequencies.

https://doi.org/10.1371/journal.pgen.1007692.s006

S1 Table. Genes on Xp ranked according to the potential to be involved in aortopathy.

This list includes all genes on Xp and an analysis of their potential to be involved in the aortopathy phenotype. The decision involved X-inactivation status, whether or not the gene was a pseudogene or a Y-chromosome homolog, and the level of expression in the aorta. The genes are ranked by likelihood for being involved in the pathogenesis of aortopathy based on these criteria.

1. Prakash S, Guo D, Maslen CL, Silberbach M, Milewicz D, Bondy CA, et al. Single-nucleotide polymorphism array genotyping is equivalent to metaphase cytogenetics for diagnosis of Turner syndrome. Genet Med. 2014;16(1):53–9. pmid:23743550
2. Gravholt CH, Juul S, Naeraa RW, Hansen J. Prenatal and postnatal prevalence of Turner's syndrome: a registry study. BMJ. 1996;312(7022):16–21. pmid:8555850
3. Barr M Jr., Oman-Ganes L. Turner syndrome morphology and morphometrics: Cardiac hypoplasia as a cause of midgestation death. Teratology. 2002;66(2):65–72. pmid:12210009
4. Sybert VP. Cardiovascular malformations and complications in Turner syndrome. Pediatrics. 1998;101(1):E11. pmid:9417175
5. Palmer CG, Reichmann A. Chromosomal and clinical findings in 110 females with Turner syndrome. Hum Genet. 1976;35(1):35–49. pmid:1002163
6. Zinn AR, Page DC, Fisher EM. Turner syndrome: the case of the missing sex chromosome. Trends Genet. 1993;9(3):90–3. pmid:8488568
7. Bondy C, Bakalov VK, Cheng C, Olivieri L, Rosing DR, Arai AE. Bicuspid aortic valve and aortic coarctation are linked to deletion of the X chromosome short arm in Turner syndrome. J Med Genet. 2013;50(10):662–5. pmid:23825392
8. Prakash SK, Bondy CA, Maslen CL, Silberbach M, Lin AE, Perrone L, et al. Autosomal and X chromosome structural variants are associated with congenital heart defects in Turner syndrome: The NHLBI GenTAC registry. Am J Med Genet A. 2016;170(12):3157–64. pmid:27604636
9. Wang L, Ming Wang L, Chen W, Chen X. Bicuspid Aortic Valve: A Review of its Genetics and Clinical Significance. J Heart Valve Dis. 2016;25(5):568–73. pmid:28238238
10. Huntington K, Hunter AG, Chan KL. A prospective study to assess the frequency of familial clustering of congenital bicuspid aortic valve. J Am Coll Cardiol. 1997;30(7):1809–12. pmid:9385911
11. Silberbach M. Bicuspid aortic valve and thoracic aortic aneurysm: toward a unified theory. J Am Coll Cardiol. 2009;53(24):2296–7. pmid:19520255
12. Garg V, Muth AN, Ransom JF, Schluterman MK, Barnes R, King IN, et al. Mutations in NOTCH1 cause aortic valve disease. Nature. 2005;437(7056):270–4. pmid:16025100
13. Shi LM, Tao JW, Qiu XB, Wang J, Yuan F, Xu L, et al. GATA5 loss-of-function mutations associated with congenital bicuspid aortic valve. Int J Mol Med. 2014;33(5):1219–26. pmid:24638895
14. Qu XK, Qiu XB, Yuan F, Wang J, Zhao CM, Liu XY, et al. A novel NKX2.5 loss-of-function mutation associated with congenital bicuspid aortic valve. Am J Cardiol. 2014;114(12):1891–5. pmid:25438918
15. Sybert VP, McCauley E. Turner's syndrome. N Engl J Med. 2004;351(12):1227–38. pmid:15371580
16. Rabkin SW. Differential expression of MMP-2, MMP-9 and TIMP proteins in thoracic aortic aneurysm—comparison with and without bicuspid aortic valve: a meta-analysis. Vasa. 2014;43(6):433–42. pmid:25339161
17. Zhang X, Wu D, Choi JC, Minard CG, Hou X, Coselli JS, et al. Matrix metalloproteinase levels in chronic thoracic aortic dissection. J Surg Res. 2014;189(2):348–58. pmid:24746253
18. Jackson HW, Defamie V, Waterhouse P, Khokha R. TIMPs: versatile extracellular regulators in cancer. Nat Rev Cancer. 2017;17(1):38–53. pmid:27932800
19. Bashash M, Shah A, Hislop G, Treml M, Bretherick K, Janoo-Gilani R, et al. Genetic polymorphisms at TIMP3 are associated with survival of adenocarcinoma of the gastroesophageal junction. PLoS One. 2013;8(3):e59157. pmid:23527119
20. Su CW, Huang YW, Chen MK, Su SC, Yang SF, Lin CW. Polymorphisms and Plasma Levels of Tissue Inhibitor of Metalloproteinase-3: Impact on Genetic Susceptibility and Clinical Outcome of Oral Cancer. Medicine. 2015;94(46):e2092. pmid:26579821
21. Anderson CL, Brown CJ. Polymorphic X-chromosome inactivation of the human TIMP1 gene. Am J Hum Genet. 1999;65(3):699–708. pmid:10441576
22. Baker AH, Edwards DR, Murphy G. Metalloproteinase inhibitors: biological actions and therapeutic opportunities. J Cell Sci. 2002;115(Pt 19):3719–27. pmid:12235282
23. Dreger SA, Taylor PM, Allen SP, Yacoub MH. Profile and localization of matrix metalloproteinases (MMPs) and their tissue inhibitors (TIMPs) in human heart valves. J Heart Valve Dis. 2002;11(6):875–80; discussion 80. pmid:12479292
24. Ikonomidis JS, Gibson WC, Butler JE, McClister DM, Sweterlitsch SE, Thompson RP, et al. Effects of deletion of the tissue inhibitor of matrix metalloproteinases-1 gene on the progression of murine thoracic aortic aneurysms. Circulation. 2004;110(11 Suppl 1):II268–73. pmid:15364874

25. Silence J, Collen D, Lijnen HR. Reduced atherosclerotic plaque but enhanced aneurysm formation in mice with inactivation of the tissue inhibitor of metalloproteinase-1 (TIMP-1) gene. Circulation Research. 2002;90(8):897–903. pmid:11988491
26. Koullias GJ, Ravichandran P, Korkolis DP, Rimm DL, Elefteriades JA. Increased tissue microarray matrix metalloproteinase expression favors proteolysis in thoracic aortic aneurysms and dissections. Annals of Thoracic Surgery. 2004;78(6):2106–10; discussion 10–1. pmid:15561045
27. Allaire E, Forough R, Clowes M, Starcher B, Clowes AW. Local overexpression of TIMP-1 prevents aortic aneurysm degeneration and rupture in a rat model. Journal of Clinical Investigation. 1998;102(7):1413–20. pmid:9769334
28. Trolle C, Nielsen MM, Skakkebaek A, Lamy P, Vang S, Hedegaard J, et al. Widespread DNA hypomethylation and differential gene expression in Turner syndrome. Sci Rep. 2016;6:34220. pmid:27687697
29. Gravholt CH, Landin-Wilhelmsen K, Stochholm K, Hjerrild BE, Ledet T, Djurhuus CB, et al. Clinical and epidemiological description of aortic dissection in Turner's syndrome. Cardiology in the Young. 2006;16(5):430–6. pmid:16984695
30. Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho TJ, et al. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. Nature. 2014;508(7497):494–9. pmid:24759411
31. Weinsaft JW, Devereux RB, Preiss LR, Feher A, Roman MJ, Basson CT, et al. Aortic Dissection in Patients With Genetically Mediated Aneurysms: Incidence and Predictors in the GenTAC Registry. J Am Coll Cardiol. 2016;67(23):2744–54. pmid:27282895
32. Asch FM, Yuriditsky E, Prakash SK, Roman MJ, Weinsaft JW, Weissman G, et al. The Need for Standardized Methods for Measuring the Aorta: Multimodality Core Lab Experience From the GenTAC Registry. JACC Cardiovasc Imaging. 2016;9(3):219–26. pmid:26897684
33. Mortensen KH, Erlandsen M, Andersen NH, Gravholt CH. Prediction of aortic dilation in Turner syndrome—the use of serial cardiovascular magnetic resonance. J Cardiovasc Magn Reson. 2013;15:47. pmid:23742092
34. Mortensen KH, Hjerrild BE, Stochholm K, Andersen NH, Sorensen KE, Lundorf E, et al. Dilation of the ascending aorta in Turner syndrome—a prospective cardiovascular magnetic resonance study. J Cardiovasc Magn Reson. 2011;13:24. pmid:21527014
35. Quezada E, Lapidus J, Shaughnessy R, Chen Z, Silberbach M. Aortic dimensions in Turner syndrome. Am J Med Genet A. 2015;167A(11):2527–32. pmid:26118429
36. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7. pmid:25722852
37. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics. 2012;28(24):3326–8. pmid:23060615
38. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82–93. pmid:21737059
39. Song W, Gardner SA, Hovhannisyan H, Natalizio A, Weymouth KS, Chen W, et al. Exploring the landscape of pathogenic genetic variation in the ExAC population database: insights of relevance to variant classification. Genet Med. 2016;18(8):850–4. pmid:26681313
40. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178–92. pmid:22517427
41. Balaton BP, Cotton AM, Brown CJ. Derivation of consensus inactivation status for X-linked genes from genome-wide studies. Biol Sex Differ. 2015;6:35. pmid:26719789
42. Esposito T, Gianfrancesco F, Ciccodicola A, D'Esposito M, Nagaraja R, Mazzarella R, et al. Escape from X inactivation of two new genes associated with DXS6974E and DXS7020E. Genomics. 1997;43(2):183–90. pmid:9244435
43. Peeters SB, Cotton AM, Brown CJ. Variable escape from X-chromosome inactivation: identifying factors that tip the scales towards expression. Bioessays. 2014;36(8):746–56. pmid:24913292

Chapter IIb:

In silico and functional studies of TIMPs and MMPs - unpublished data

Introduction:

Our whole exome sequencing study on a discovery cohort of 188 Turner syndrome (TS) individuals and a replication cohort of 53 individuals exhibited that *TIMP3* and *TIMP1* are risk genes for aortopathy in the TS population. Specifically, the combinatorial effect of deleterious variants in *TIMP3* and the loss of *TIMP1* copy number synergistically heightened the risk for BAV and associated TAD, which is the first sign of aneurysm formation.

Next, we aim to pinpoint the mechanism in which TS individuals enter the aortopathy disease cycle, in which we hypothesize is through the TIMP/MMP imbalance. To this aim, we will investigate if *TIMP1* copy number is correlated to TIMP1 protein levels, the functional consequences of the *TIMP3* risk alleles, and the MMP activity levels.

Finally, we contend that the euploid male population with BAV and associated aortopathy (non-syndromic BAV aortopathy) might enter the disease cycle in the same manner based on only having one copy of *TIMP*1.

Do TIMP1 protein levels correlate with *TIMP1* copy number?

Introduction:

As previously mentioned, Turner syndrome (TS) is the result of a complete or partial loss of the second sex chromosome. Specifically, ~50% have a complete monosomy X, ~30% either have a partial deletion or rearrangement, ~20% are mosaic for a second sex chromosome, and < 2.5 % has some Y chromosome material. It is hypothesized that the degree to which the second sex chromosome is retained leads to varying symptoms of TS. Additionally, deletion of the Xp arm of the X chromosome has been linked to risk for aortopathy in TS. Therefore, we assessed this second sex

chromosome status for the patients in our WES cohort and then looked at candidate X chromosome genes that could account for the increased risk of aortic disease in TS. The criteria we used for ranking the candidate X chromosome genes were that they must be on the Xp arm, escape X inactivation, be expressed in the heart, and must not have a pseudogene on the Y chromosome. Based on these criteria, *TIMP1* was our top candidate gene.

In euploid females, it has been shown that *TIMP1* has a cell to cell and sample to sample variability in escaping inactivation on the second X chromosome[1,2]. Notably, TIMP1 is a secreted protein and therefore presents a unique compensatory mechanism, in that those TS patients that are mosaic for cells that have two *TIMP1* copies could compensate for cells that only have one *TIMP1* copy. Based on the level of mosaicism and second sex chromosome status, each patient was assigned their *TIMP1* copy number. For example, a patient that had 50% 46,XX and 50% 45,X would be assigned a *TIMP1* copy number of 1.5. Using this data we showed that *TIMP1* copy number was a significant risk factor for BAV and TAD in the Tuner syndrome population, where having more than one copy was protective. To further support this model, we investigated if our assigned *TIMP1* copy number did indeed correlate with TIMP1 protein levels.

Hypothesis:

TIMP1 protein levels will be decreased in those with only one copy of *TIMP1*; and quantification of those levels will correlate with *TIMP1* copy number.

Results:

We investigated if *TIMP1* copy number was correlated to the actual secreted TIMP1 protein level. This would support our hypothesis that an inherent deficiency in an X chromosome gene is the root for the increase risk of aortic disease in in the Turner syndrome population and also the observed sex bias of aortic disease in the euploid male population. To this end, we collected plasma and

saliva from study subjects during the annual Turner Syndrome Society of the US (TSSUS) patient conference. And an ELISA assay was used to measure TIMP1 protein levels in the plasma of Turner syndrome patients with known *TIMP1* copy number.

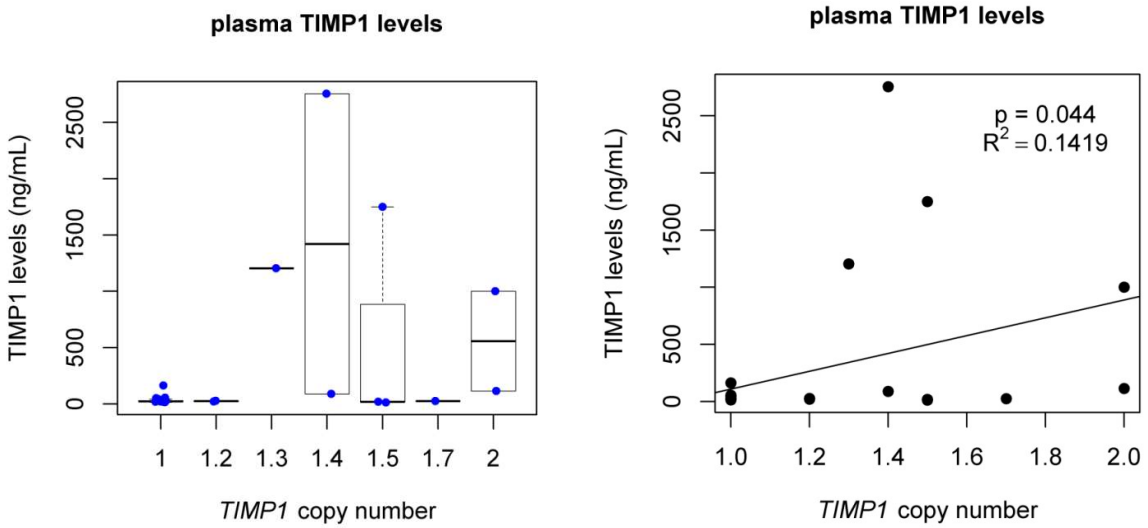


Figure 7: TIMP1 protein levels in the plasma of Turner syndrome patients with variable *TIMP1* copy numbers. We show a positive correlation between increased protein level and increased TIMP1 copy number. P-value was calculated using a linear regression model where TIMP1 copy number was the predictor variable and TIMP1 protein level was the outcome.

We show that *TIMP1* copy number is significantly correlated with TIMP1 protein levels in plasma (Figure 7, p=0.044). Additionally, we observed that the samples with only one copy of *TIMP1* are tightly clustered with little variability, whereas the samples with more than 1 copy have more variability.

This also presents a hypothesis for the observed sex bias in BAV associated aortopathy. It is known that euploid males are predisposed to have BAV and associated aortopathy, compared to euploid females, which indicates that having two X chromosomes is protective. We further support and refine this hypothesis by locating the potential causal gene, *TIMP1*, leading to a dosage effect and the consequential TIMP/MMP imbalance which feeds into the aortopathy disease cycle.

What are the possible consequences of rs9862 and rs11547635 *TIMP3* SNPs?

Introduction:

On the gene level, *TIMP3* has a gene damage index PHRED score of 0.449, placing in the top 10% of genes intolerant of mutations[3]. Our lead SNP rs11547635 in *TIMP3* has a CADD score of 16.49, which is above the suggested deleterious significance cutoff of 15. It is a synonymous C>T transition at Ser87 in exon 3. The variant rs11547635, is in linkage disequilibrium with a second variant rs9862 which is also a synonymous variant caused by a T>C transition at His83 in exon 3.

While synonymous variants do not alter the amino acid content of their proteins, they have been shown to affect gene expression, protein folding efficiency, etc. In fact, over 50 human diseases have now been associated with synonymous variants[4]. A large amount of bioinformatic and experimental methods have been developed to gain insight into the effects of synonymous variants.

Functional predictions for *TIMP3* rs9862 and rs11547635 were done using computational tools to analyze gene and protein sequences to assess potential consequences of SNPs. PROMO was used for the prediction of putative transcription binding sites[5,6], ESEfinder 3.0 was used to identify potential disruption of exonic splicing enhancers (ESEs)[7,8], and The Sequence Manipulation Suite was used to evaluate codon usage[9]. Experimental methods such as splicing assays, qPCR, binding assays, site-directed mutagenesis, and western blots were used to test hypothesis and predictions from the bioinformatic analyses.

Hypothesis:

The *TIMP3* variants rs9862 and rs11547635 are synonymous, located within the body of an exon, possess a palindromic sequence, and alters the preferred codon. Therefore we hypothesize that through one of these mechanisms the variants will impact transcription or translation, which will in turn cause a decrease in TIMP3 protein expression.

Results:

One of the primary mechanisms in which a synonymous variant impacts the gene it resides in is at the transcriptional level. Our two variants of interest, rs9862 and rs11547635, are separated by 11bp and possess a palindromic sequence, making it a good candidate site for protein binding. The region of these variants is predicted to have three transcription factor binding sites, two of which are disrupted when the rs9862 and rs11547635 variants are present (STAT4 and EST1) (Figure 1). EST1 is known to bind as a homodimer to a palindromic sequence of ACGGAAG/CTTCCGA, which is in the region between rs9862 and rs11547635 and encompasses rs11547635 (underlined below).

Wildtype: agtacatcca**T**acggaagcttc**C**gagagtctct & Double mutant: agtacatcca**C**acggaagcttc**T**gagagtctct

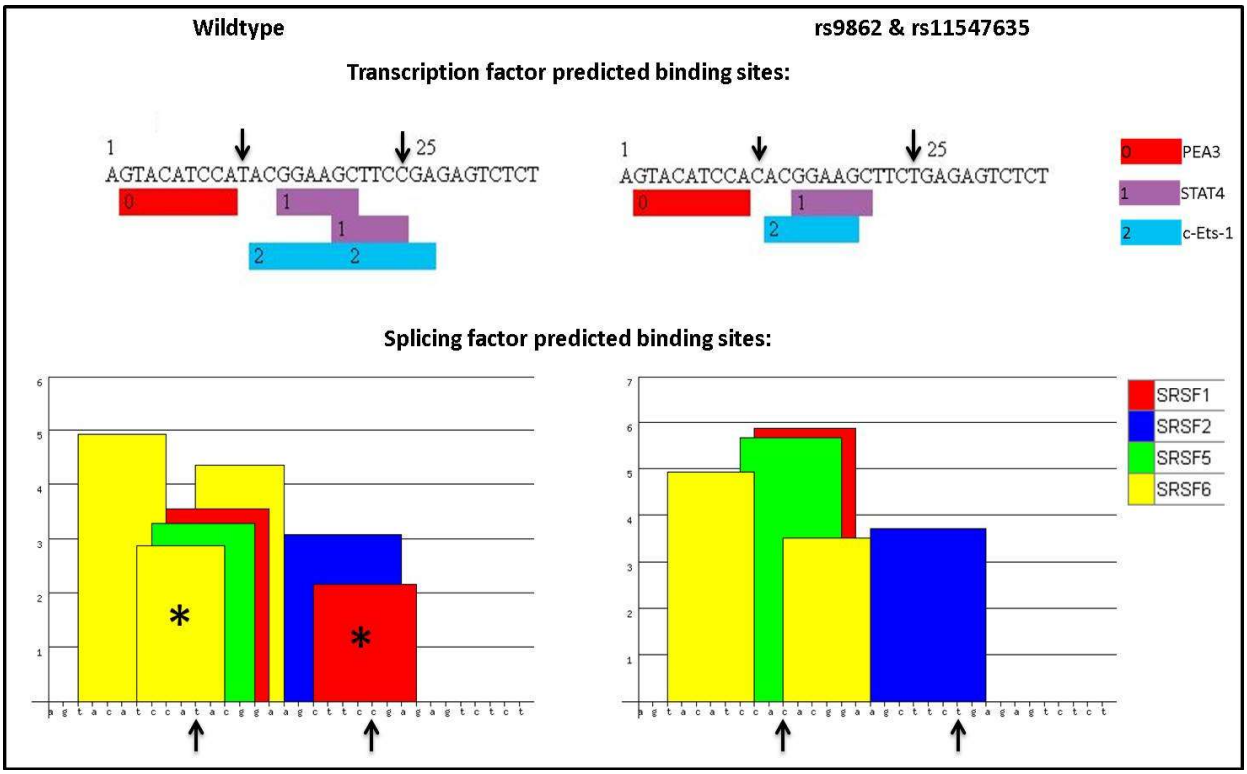


Figure 1: The potential function effects of the TIMP3 variants rs9862 and rs11547635. Top) PROMO prediction for putative transcription binding sites shows the loss of STAT4 and Ets-1 binding sites when rs9862 and rs11547635 are present, arrows point to alleles. Bottom) ESEfinder predicts the loss of SRSF6 and SRSF1splicing factors binding site when rs9862 or rs11547635 is present, respectively, represented by the stars.

Further supporting the *in silico* analysis, a gel-shift assay showed differential binding of ETS1 between the wildtype sequence and the double mutant sequence[10]. Another study showed that *TIMP3* mRNA levels were dramatically decreased when ETS1 was inhibited[11]. These studies

demonstrate that ETS1 not only binds to the core consensus sequence (underlined above), but that this binding could be critical for *TIMP3* mRNA expression.

In the aneurysmal state of the aorta, the ECM is highly fragmented and sparse, leading to the release of TGFß1 and cytokines and to the apoptosis of VSMCs. An immune response is elicited to the damaged tissue by T-lymphocytes and macrophages which have been shown to express MMPs and TIMPs. Therefore, we performed *TIMP3* qPCR on M1-macrophages of patients representing the different genotypes of rs9862 and rs11547635 to investigate if there were changes in *TIMP3* mRNA expression, which would support our hypothesis of impacts at the transcriptional level. We observed a step-wise decrease in *TIMP3* mRNA expression in a dosage dependent manner, where the homozygous double mutant was statically significantly different from the homozygous wildtype (Figure 2, p = 0.0129).

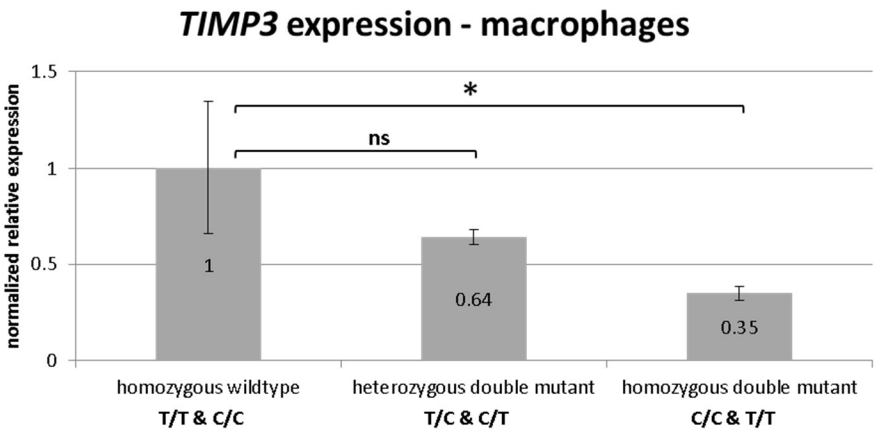


Figure 2: qPCR relative expression levels of *TIMP3*. Differentiated M1 macrophages from TS patients were used representing different alleles of rs9862 and rs11547635. *TIMP3* expression levels were normalized to the housekeeping gene 18s and displayed in the plot as relative to the control (homozygous wildtype). A student's t-test was performed on the mean of three technical replicates compared to the homozygous wildtype. Normalize relative expression and ±SEM are reported.

A second mechanism of potential impacts of synonymous variants is through the binding of splicing factors or changes to the core splicing motifs. In fact, the bulk of the information for splicing is thought to be contained within the exon body through binding sites for exonic splicing enhancers and

repressors[12]. Additionally, alternative splicing is regulated differentially depending on the tissue and developmental stage. While the "code" for splicing events such as alternative splicing, splicing enhancement, or repression, is still being determined, algorithms have been developed to scan sequences for potential binding sites. Our sequence comprised of the two variants eliminates the exon splicing enhancer/repressors SRSF6 and SRSF1 binding sites when rs9862 or rs11547635 is present, respectively.

To test the binding or the disruption of binding of SRSF6 and SRSF1, which would support our hypothesis of impacts at the splicing level, a gel-shift experiment was designed. Briefly, labeled RNA probes with or without the variants of interest were incubated with recombinant SRSF6 or SRSF1 and run on a gel. No binding was observed to either the wildtype or mutant RNA probes. Since the binding of splicing factors to their consensus sequence is dependent on multiple experimental conditions, troubleshooting is recommended. Specifically, salt concentrations, the phosphorylation state of SR proteins, and/or other co-factors that are required for binding to occur, could all impact proper binding. This will be followed up on in future work in the lab.

To indirectly test effects of the SR factors, we also investigated if the *TIMP3* rs11547635 allele resulted in proper splicing of exon 3 using a mini-gene assay, which is a splicing reporter vector. There is well documented variation in splicing depending on the tissue type or developmental stage, in which different splice "codes" and splicing factors are used[13]. Therefore, the mini-gene assay was performed in two different cell types: human embryonic kidney cells (HEK293s) and vascular smooth muscles cells (VSMCs). We saw no evidence for altered splicing through examination of the expected band size and confirmed this by Sanger sequencing (Figure 3). Although the mini-gene assay is a quick and efficient way to test for the effects of variants on splicing, there are many pitfalls since the genomic context of the intro/exon boundaries is changed. Hence, we also investigated splicing in an endogenous genomic context and relevant cell type, with regards to aortic disease,

using six patient derived M1 macrophages. After collecting RNA from the cells and making cDNA, primers were used to amplify *TIMP3* cDNA. Confirming the mini-gene assay, we saw no changes in splicing between mutant forms of *TIMP3* and the wildtype alleles (Figure 4). Additionally, we can conclude it is not cell type specific, although it is possible that it could still be developmental stage dependent.

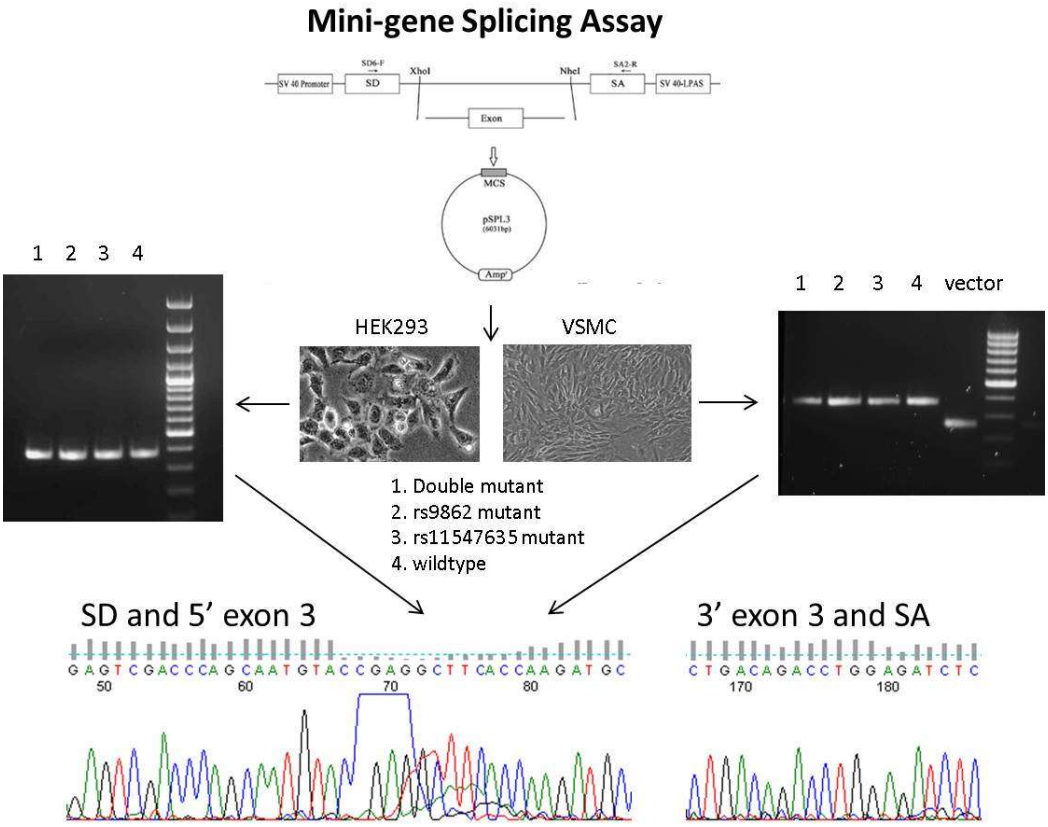


Figure 3: Mini-gene assay of *TIMP3* rs11547635 variant and wildtype allele, showing that it is properly spliced and the band is the expected size of 376bp.

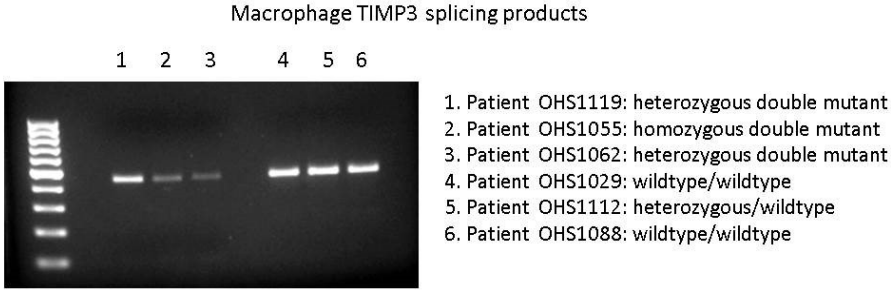


Figure 4: *TIMP3* cDNA was amplified from patients with and without the risk variants rs9862 and rs11547635. All samples had the correct and expected size band of 455bp.

The third mechanism we investigated *in silico* is the effect synonymous mutations can have on codon usage. Codon usage biases are well recognized in its ability to effect gene expression through RNA processing, translation, folding, and post-translational modifications (PTM). In extreme cases, it has been shown experientially that the use of particular codons can increase or decrease the expression of a transgene by 1000 fold[14]. The human *TIMP3* gene shows extreme codon bias for serine residues. The 4 serine codons used across the *TIMP3* mRNA are UCC, AGC, AGU, and UCG at frequencies of 50%, 36%, 7%, and 7%, respectively. The wildtype codon UCC is the "preferred codon". SNP rs11547635 results in a codon switch from the wildtype UCC to UCU, which is not represented in *TIMP3* serine codon usage (Figure 5). We hypothesize this could impact translation speed, efficiency, protein stability, or downstream PTMs.

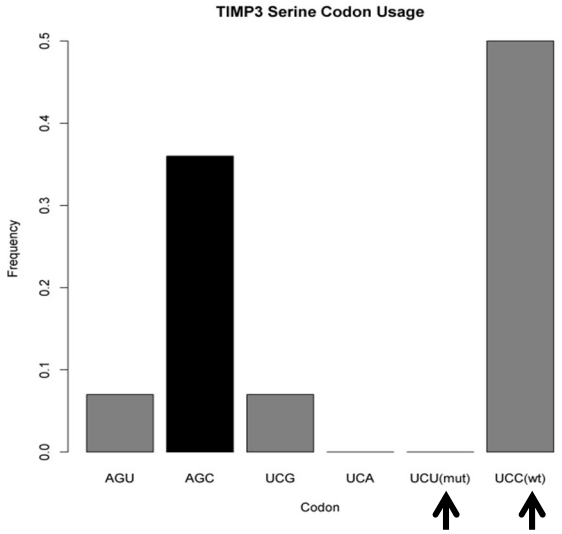


Figure 5: SNP rs11547635 results in a codon switch from the wildtype UCC to UCU, which is not represented in the wildtype serine codon usage in the *TIMP3* gene

To test this hypothesis, site-directed mutagenesis was performed on the stock DDK-tagged *TIMP3* cDNA plasmid in which three *TIMP3* plasmids were made, representing genotypes observed in our samples:

SDM1) rs9862, rs11547635
SDM2) rs9862, wildtype
WT) wildtype, wildtype

Each plasmid was Sanger sequenced verified before transfection into HEK293 cells. After stable transfection by growing under selection pressure, the cells were collected and protein was extracted.

A western blot was performed to investigate TIMP3 protein expression levels and any PTM-TIMP3 changes. We observed three bands representing different forms of expressed TIMP3 protein: monomeric TIMP3 at 24kDa, Glycosylated TIMP3 at 27kDa, and a TIMP3 dimer at 48kDa (Figure 6).

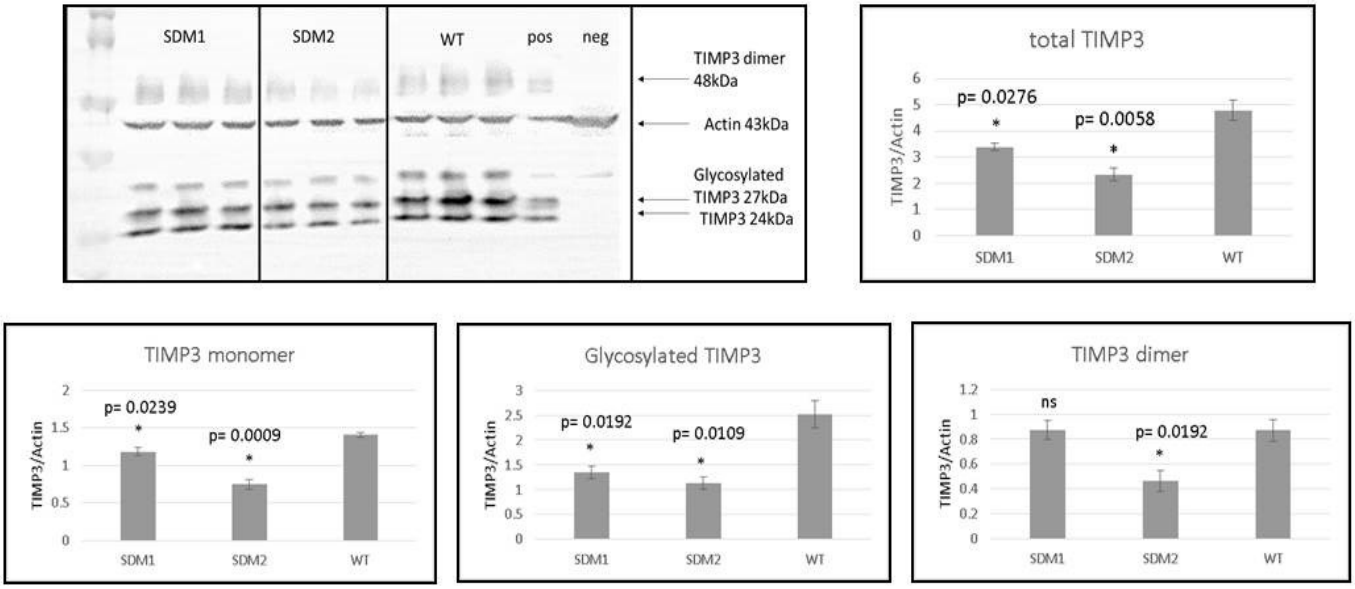


Figure 6: Decreased protein expression of TIMP3 in mutant forms TIMP3 compared to the wildtype. A western blot of TIMP3 from the cell pellets of HEK293 transfected with flag-tagged *TIMP3* cDNA plasmids representing different mutational states of interest. SDM1: rs9862, rs11547635; SDM2: rs9862, wt WT; wt, wt. Three TIMP3 bands were present and actin was used to normalize across samples. Error bars represent the standard error of the mean. P-values calculated using two tailed student's t-test.

Monomeric and glycosylated TIMP3 are significantly decreased in both mutant forms of *TIMP3*, when compared to the wildtype *TIMP3*. Importantly, TIMP3 is a secreted protein and must undergo glycosylation in the ER lumen to become properly folded, secreted, and able to bind the ECM, thus the glycosylated TIMP3 form is the biologically relevant TIMP3 of interest[15]. Highlighting the importance of glycosylated TIMP3, a recent study showed that engineering TIMP3 molecules to have extra site specific glycosylation resulted in higher protein expression by extending its half-life, improved its ECM binding capabilities, and overall activity[16]. Our variants of interest, rs9862 and rs11547635, are synonymous SNPs and do not change the amino acid composition of the protein and are unlikely to affect actual glycosylated sites.

Although the mechanism in which the *TIMP3* risk alleles affect TIMP3 expression still need to be refined. The work presented here supports the premise that they lead to decreased expression of TIMP3 because both the mRNA expression and protein levels are decreased in the mutants compared to the wildtype. We propose that this is on the transcription level and is propagated and potentially further compounded on the translational level (Figure 7).

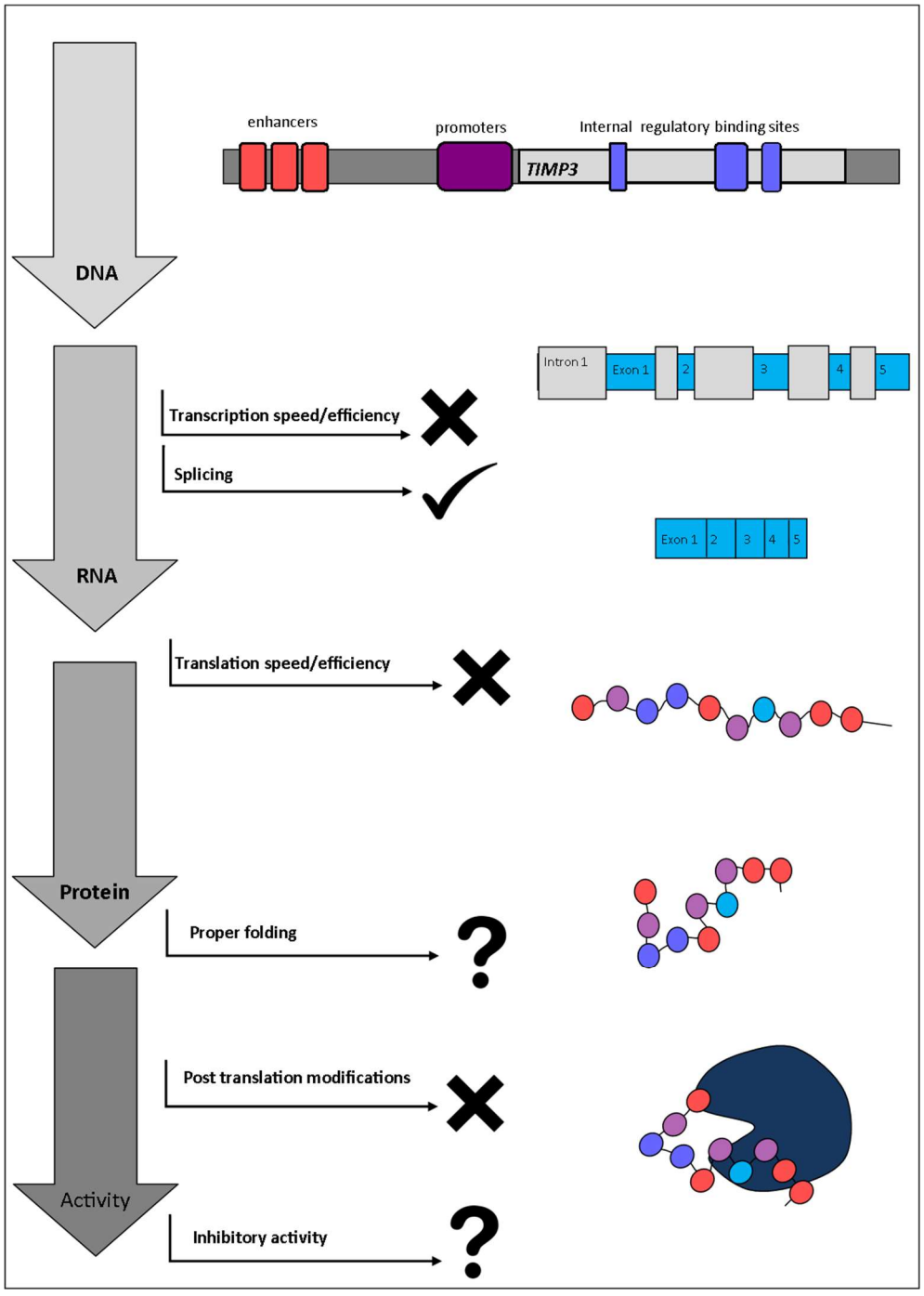


Figure 7: Model of the possible mechanisms of a synonymous variant's effects on the gene in which they reside in. We tested the effects of TIMP3 risk alleles rs9862 and rs11547635. Effects on the transcriptional level were seen by differential *TIMP3* mRNA expression through qPCR, while proper splicing was confirmed through a mini-gene and cDNA analysis. Differential TIMP3 protein levels and amount of glycosylated protein were also observed using a western blot. Proper folding and quantification of inhibitory activity still need to be tested for.

Do MMP protein levels, in the saliva and/or plasma of patients, correlate with aortic disease and *TIMP3* status? And do these levels correlate to actual activity of the enzyme?

Introduction:

There are 23 MMPs in humans and a large body of work has showed their importance in the progression of aortopathy. Particularly of interest to this study, a meta-analysis showed that MMP2 and MMP9 are significantly increased in BAV associated thoracic aortic aneurysms (TAA)[17]. It is also known that MMP9 is the most abundant protease produced by human aneurysm tissue[18]. Further supporting these findings, MMP9 knockout mice had attenuated TAA formation, after TAA induction using $CaCl_2$[18]. Therefore, we focus on the abundance and activity level of MMP9 in the following experiments.

While measuring MMP protein levels could prove to be a useful tool to assess the progressiveness of aortopathy, their activity level is also an important aspect in their contribution to a disease model. MMPs are secreted from the cell in their pro-MMP form (inactive) and then activated extracellularly by the cleavage of the pro domain by an array of proteases and chemicals. A majority of the time, TIMPs irreversibly bind to the active form of MMPs in a 1:1 stoichiometry, which inhibits MMP activity (there are a few pro-MMP/TIMP interactions). The antibodies used in the MMP ELISAs do not distinguish between the two forms. Therefore, a zymogram assay was also performed to investigate the ratio of MMP9 to pro-MMP9 and their respective activity levels.

Hypothesis:

MMP protein levels will be higher in TS patients with aortopathy and the *TIMP3* risk variants compared to those with healthy aortas and no *TIMP3* risk variants. Furthermore, we hypothesize that these total MMP protein levels will correlate with their enzymatic activity.

Results:

As mentioned above, we collected plasma and saliva from study subjects during the annual Turner Syndrome Society of the US (TSSUS) patient conference. And an ELISA assay was used to measure total MMP9 protein levels in the plasma of Turner syndrome patients with or without aortopathy, known *TIMP1* copy number, and with or without the TIMP3 risk alleles.

Case (*TIMP3* rs9862 heterozygous, rs11547635 heterozygous, 1 copy of *TIMP1*, BAV w/ TAA)
Vs.
Control (*TIMP3* rs9862 wildtype, rs11547635 wildtype, 1 copy of *TIMP1*, no BAV and no TAA)

The TS case with the TIMP3 risk alleles and aortopathy had significantly higher MMP9 total protein levels in their plasma than a TS individual, without the *TIMP3* risk allele and no aortopathy (control) (p=0.0014, Figure 8).

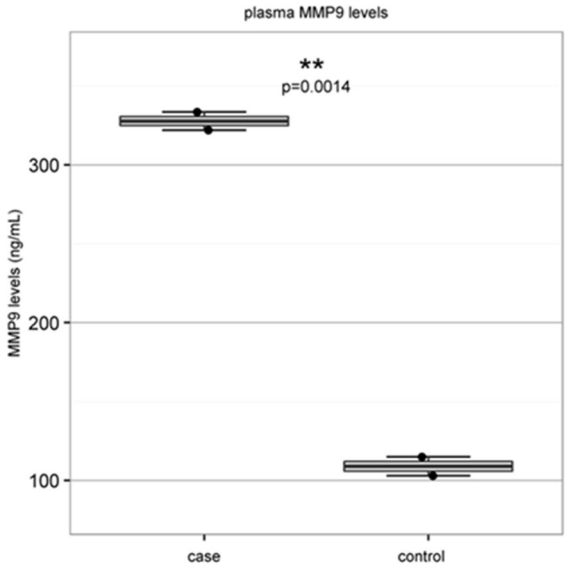


Figure 8: MMP9 protein levels in the plasma of Turner syndrome patients. There is a significant increase in circulating total MMP9 protein levels in the cases compared to the control. Samples were run in duplicate and the p-value calculated using a students' two tailed t-test.

To measure actual MMP9 activity, a zymogram was run on saliva from the same samples. A particular advantage of this system is that the dimer, the proenzyme, and active forms of MMP9, can be separated on the basis of molecular weight and their levels can be detected. Although the pro and dimer forms are normally inactive, the denaturing conditions in the first step unfold the pro (inhibitory) domain of the MMPs that when re-folded in the renaturing step allows the enzyme to stay

active[19]. As mentioned earlier, MMPs interact with their inhibitors, TIMPs, to regulate their activity. Importantly, the sample preparation method dissociates this noncovalent MMP-TIMP complex, allowing us to be able to measure total MMP9 activity[20]. A gelatin substrate gel was used as the substrate for MMP9. The saliva of patients was used to investigate the MMP9 activity levels, with each sample was run in triplicate.

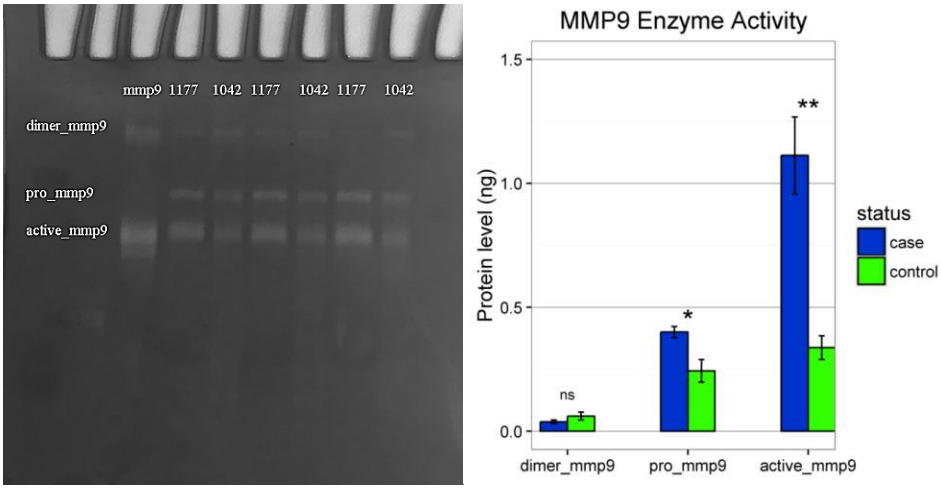


Figure 9: Zymogram showing MMP9 activity levels in the saliva of a TS case vs a TS control. The bar plot shows the significant differences in pro-MMP9 and mature MMP9 levels, where they are increased in the case compared to the control samples.

The TS case with aortic disease and the *TIMP3* risk alleles had significantly higher MMP9 enzymatic activity in their saliva, when compared to a control with a healthy aorta and wildtype TIMP3 alleles. Specifically, the pro-MMP9 and the monomeric active MMP9 forms were higher (Figure 9). Importantly, the total MMP9 protein level in saliva seems correlated with the total protein level measure in the plasma via an ELISA. The major function of TIMPs is to inhibit their MMP counterparts to maintain homeostasis of the ECM[21].

Conclusion:

The work presented here is a thorough investigation of the genetic basis of TS associated aortopathy (specifically, BAV and TAA); from an unbiased, case-control, exome wide scan of a discovery cohort of 188 TS patients and replication cohort of 58 TS patients to the functional analysis

of those genetic components. In our model of TS associated aortopathy, we propose that the inherent deficiency of *TIMP1* represents the sex chromosome sensitizing genetic factor and that an additional autosomal genetic hit in *TIMP3* causes an imbalance to the required TIMP/MMP protein homeostasis (Figure 10). We propose that this triggering event is the genetic basis of why some TS individuals reach the aneurysmal disease threshold and other do not. Additionally, both TIMP3 and TIMP1 play a role in aortic valve development, highlighting a previously unrecognized mechanism between BAV pathogenesis and downstream TAD and provide an explanation for the heightened risk for BAV associated aortopathy in TS.

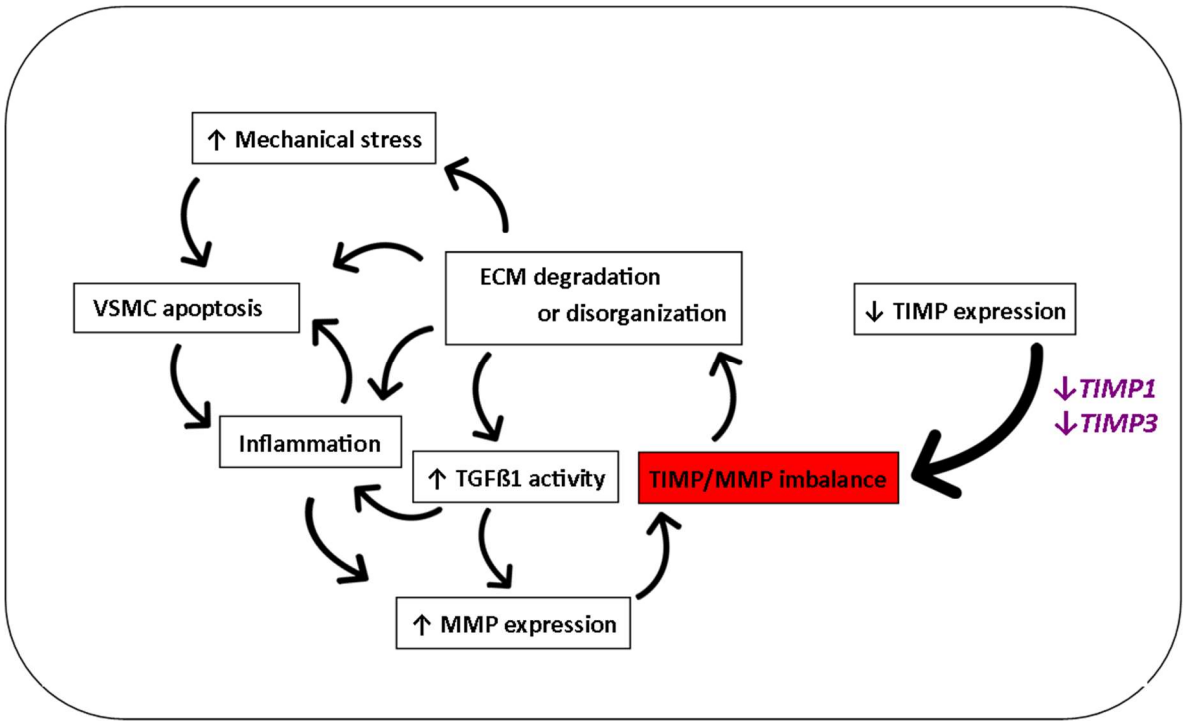


Figure 10: Feedback cycle of the pathogenic events in TAA. We propose that in Turner syndrome (TS) the disease cycle is entered through decreased TIMP1 and TIMP3 expression leading to an imbalance in the TIMP/MMP ratio.

When the equilibrium between TIMPs and MMPs becomes imbalanced in favor of more MMPs and less TIMPs the ECM can begin to breakdown leading to a pathological disease progression (Figure 11). We hypothesize that this incites a pro-inflammatory cell invasion of T-lymphocytes and macrophages, which are thought to be the source of MMPs in diseased tissue, further spiking the

imbalance of TIMPs and MMPs[22]. Through this work we add to the current understanding of the genetic and molecular etiology of TS associated aortopathy and fill a gap that was in the field. Our work also presents a hypothesis for the observed sex bias in BAV associated aortopathy. It is known that euploid males are predisposed to have BAV and associated aortopathy, compared to euploid females, which indicates that having two X chromosomes is protective. We propose that hemizygosity for TIMP1 in euploid males increases risk for BAV and later onset aortic disease. We further support and refine this hypothesis by locating the potential causal gene leading to a dosage effect and to the TIMP/MMP imbalance which feeds into the aortopathy disease cycle.

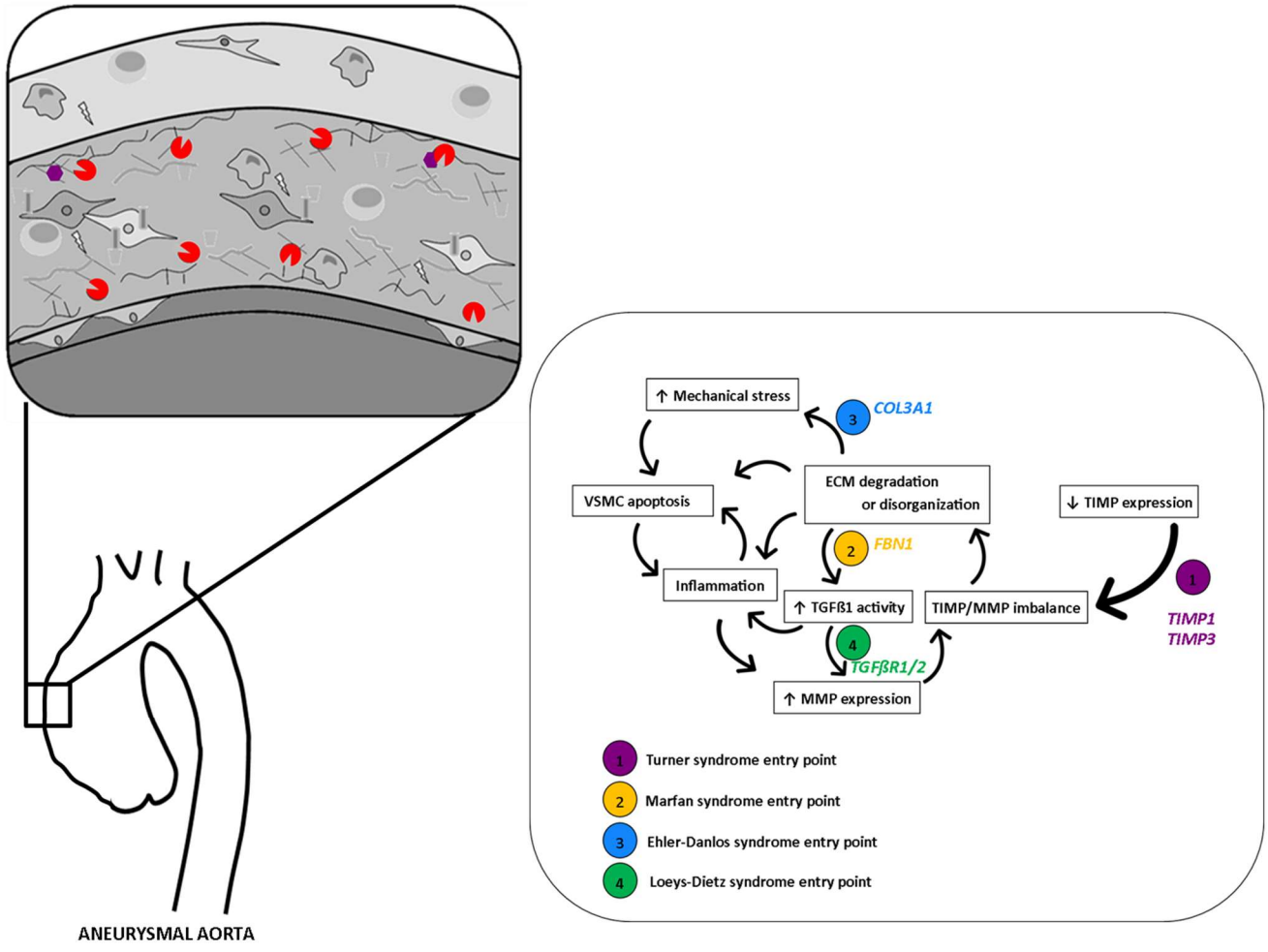


Figure 11: Final model of the genetic and molecular etiology of TS associated aortopathy adding to the current knowledge in the field of TAA progression.

Methods:

Computational tools

PROMO was used for the prediction of putative transcription binding sites, ESEfinder 3.0 was used to identify potential disruption of exonic splicing factors, and The Sequence Manipulation Suite was used to evaluate codon usage.

Gel-shift assay of splicing factors

Biotin labeled RNA probes of 30bp in length, with or without the variants of interest, were ordered for Sigma. Using the LightShift Chemiluminescent RNA EMSA Kit (Thermo #20158) the probes were incubated with recombinant SRSF6 or SRSF1 and run on a 6% DNA Retardation gel (Invitrogen #EC6365BOX). No binding was observed to either the wildtype or mutant RNA probes. Salt concentrations, the phosphorylation state of SR proteins, and or other co-factors that are required for binding could all impact proper binding and should be followed up on in future work in the lab.

Mini-gene assay

The mini-gene assay was performed using the pSPL3 vector (from Dr. Hansen Center for Genomic Medicine, University of Copenhagen), in which three TIMP3 variants and wildtype sequences were cloned in. Briefly, primers were designed with restriction enzyme sites and were designed to include flanking introns of ~170bp surrounding the exon 3 of *TIMP3*. These PCR products were cloned into the pSPL3 vector and transformed into XL10-Gold Ultra competent cells. Clones were sequenced to confirm the mutant and wildtype allele presence. The purified vector was transfected into HEK293T and VSMCs (ATCC #CRL1999) cells using Lipofectamine 3000 (Thermo Fisher), RNA was purified (Qiagen RNeasy kit), and cDNA was generated using the New England Bio Labs MuLV RT system. The splicing products were amplified with PCR using the SD6-Forward primer

(TCTGAGTCACCTGGACAACC) and SA2-Reverse primer (ATCTCAGTGGTATTTGTGAGC). The splicing products were run on a 2% agarose gel for 1 ½ hours at 70 volts.

Patient derived macrophages

Patient peripheral blood mononuclear cells (PBMCs) were isolated by density centrifugation of whole blood using SepMate tubes and Lymphoprep reagent (StemCell Technologies). Monocyte-derived macrophages (MDMs) were matured from these PBMCs by incubating with 50 ng/mL GM-CSF for 7days (R&D Systems).

Quantitative real-time PCR

(Done by a research associate in the lab, Rebecca)

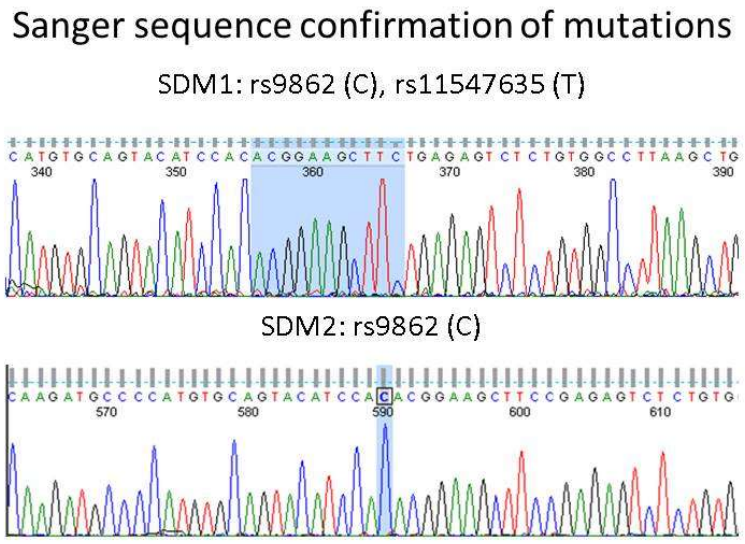
Total RNA was isolated from MDMs using TRIzol reagent (Invitrogen). A total of 500 ng of RNA was reverse transcribed into cDNA using M-MuLV reverse transcriptase. Quantitative real-time PCR was performed using ABI StepOnePlus and TaqMan reagents (Applied Biosystems). TIMP3 expression levels were quantified using TaqMan Gene Expression Assay Hs00165949_m1. Each reaction was performed in triplicate using standard reaction conditions. The TIMP3 mRNA levels were normalized to those of the housekeeping gene, 18s. Calculations were performed via a comparative cycle threshold method. A students' two tailed t-test was performed on the mean delta Ct of each group.

Endogenous TIMP3 splicing

RNA was collected from the patient derived MDMs using TRIzol Reagent (Thermo #15596026) and cDNA was synthesized using MuLV RT. TIMP3 cDNA (splicing products) were amplified with PCR using the TIMP3-Forward primer (CTCGGGCTCATCGTGCTC) and TIMP3-Reverse primer (AAGCAAGGCAGGTAGTAGCA). The products were run on a 2% agarose gel for 1 ½ hours at 70 volts.

Site-directed mutagenesis

Construction of the mutant TIMP3 plasmids (SDM1 and SDM2) were performed on Timp3 ORF, Myc-DDK-tagged, in pCMV6 vector (Origene, #RC202600) using the QuikChange Multi Site-Directed Mutagenesis Kit (Agilent). Antisense primers containing the nucleotide change of interest were designed and obtained from Sigma. Plasmids for transfection were prepared with PureYield Plasmid Miniprep System (Promega) and were Sanger sequenced to confirm the mutations and to ensure that no other mutations had been introduced (Supplementary Figure 1).



Supplemental Figure 1: Sanger sequence chromatogram confirming the mutations were introduced into the TIMP3 cDNA plasmid using site-directed mutagenesis.

TIMP3 plasmids were transfected into HEK293 cells using Lipofectamine 3000 (Thermo Fisher). Cells were treated with geneticin (G418) and selectively pressured for 1 week before cell pellet collection.

Western blot

(Done by a research associate in the lab, Rebecca)

The total protein was collected from the cell pellet of the SDM clones and the concentration was determined using Bio-Rad DC Protein Assay. Western blot was performed using the NuPAGE Novex

system (Thermo Fisher) using the primary antibodies: Anti-FLAG tag antibody (TA50011-100, Origene) for DDK-tagged TIMP3 and Anti-Beta-Actin antibody (A2228, Sigma). A positive control HEK293T lysate transfected with the RC202600 plasmid and a negative control of HEK293T lysate of the empty plasmid was purchased from Origene (#LY400129). GeneSys image acquisition system and Gene Tools software were used for imaging and band quantification (Syngene). Samples were run in triplicate and p-values were calculated using a students' two tailed t-test.

Plasma collection and ELISA assays

Samples were collected from existing GenTAC registry participants (OHS), where whole blood was process using the Sepmate protocol where plasma was isolated, cryopreserved, and stored. An ELISA assay (Sigma #RAB0466) was used to measure TIMP1 protein levels and an ELISA assay (Sigma #RAB0372) was run to investigate total MMP9 plasma levels in the plasma of TS individuals. The colormertric intesity was measured using the Synergy H1 Hybrid reader (BioTek). The mean absorbance was calulated for each set of duplicate standards, controls, and samples. The standard curve and sample data was ploted in R. An ELISA for MMP2 was also conducted, but the protein levels were too low to detect.

Zymography

A Zymography (Novex system, Thermo) was run under denaturing (SDS) but non-reducing conditions through a polyacrylamide gel containing gelatin for 60V for 20 min and 125V for 120 min. The resolved proteins were renatured by the exchange of the SDS with a nonionic detergent and the gel is incubated at 37°C overnight. The gel was stained with Coomassie Blue and proteolytic activities were detected as clear bands against a dark background of un-degraded gelatin. Human recombinant MMP9 was used as a positive control and to calculate actual protein levels. ImageJ was used to quantify bands and the means of three replicates were used to calculate statistical significance using a t-test.

1. Tukianinen et al. (2017) Landscape of X chromosome inactivation across human tissues. Nature volume 550, pages 244–248
2. Catherine L. Anderson and Carolyn J. Brown. Polymorphic X-Chromosome Inactivation of the Human TIMP1 Gene. (1999) Am. J. Hum. Genet. 65:699–708, 1999.
3. Itan Y. et al. The human gene damage index as a gene-level approach to prioritizing exome variants. PNAS. 2015;
4. Chen, R. et al (2010) Non-synonymous and synonymous coding SNPs show similar likelihood and effect size of human disease association. PLoS ONE. 5, e13574.
5. Xavier Messeguer, Ruth Escudero, Domènec Farré, Oscar Nuñez, Javier Martínez, M.Mar Albà. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics, 18, 2, 333-334, 2002.
6. Domènec Farré, Romà Roset, Mario Huerta, José E. Adsuara, Llorenç Roselló, M.Mar Albà, Xavier Messeguer. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic Acids Res, 31, 13, 3651-3653, 2003.
7. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. 2003;31(13):3568-71. Epub 2003/06/26
8. Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet. 2006;15(16):2490-508. Epub 2006/07/11. doi: 10.1093/hmg/ddl171.
9. Stothard P. et al. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques; 2000;28(6):1102, 4. Epub 2000/06/27. PubMed PMID: 10868275.
10. Bashash M. et al. Genetic Polymorphisms at TIMP3 Are Associated with Survival of Adenocarcinoma of the Gastroesophageal Junction. PLoS One; 8(3), e59157. 2013. doi:[10.1371/journal.pone.0059157]
11. Kessler CA1, Schroeder JK, Brar AK, Handwerger S. Transcription factor ETS1 is critical for human uterine decidualization. Mol Hum Reprod. 12(2):71-6. Epub 2006 Feb 2. DOI: 10.1093/molehr/gal008
12. Wang Z, Burge CB. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA. 2008;14(5):802-13.
13. Gene Yeo, Dirk Holste, Gabriel Kreiman, and Christopher B Burge. Variation in alternative splicing across human tissues. Genome Biology 2004, Volume 5, Issue 10, Article R74
14. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression.Trends Biotechnol. 2004 Jul; 22(7):346-53.
15. Wei-Hsuan Yu‡, Shuan-su C. Yu, Qi Meng§, Keith Brew and J. Frederick Woessner Jr. TIMP-3 Binds to Sulfated Glycosaminoglycans of the Extracellular Matrix. The Journal of Biological Chemistry. (2000) 275, 31226-31232.
16. Chintalgattu V, Greenberg J, Singh S, et al. Utility of Glycosylated TIMP3 molecules: Inhibition of MMPs and TACE to improve cardiac function in rat myocardial infarct model. Pharmacol Res Perspect. 2018;e00442.
17. Vasa. 2014 Nov;43(6):433-42. doi: 10.1024/0301-1526/a000390. Differential expression of MMP-2, MMP-9 and TIMP proteins in thoracic aortic aneurysm - comparison with and without bicuspid aortic valve: a meta-analysis. Rabkin SW1.
18. Ikonomidis JS et al. Effects of deletion of the matrix metalloproteinase 9 gene on development of murine thoracic aortic aneurysms. Circulation. 2005 Aug 30;112 (9 Suppl):I242-8.
19. Frankowski H, Gu YH, Heo JH, Milner R, Del Zoppo GJ. Use of gel zymography to examine matrix metalloproteinase (gelatinase) expression in brain tissue or in primary glial cultures. *Methods Mol Biol*. 2012;814:221-33.
20. Stetler-Stevenson WG Tissue inhibitors of metalloproteinases in cell signaling: metalloproteinase-independent biological activities.. *Sci Signal. 2008 Jul 8; 1(27):re6.*
21. Gomez D. E., Alonso D. F., Yoshiji H., Thorgeirsson U. P. Tissue inhibitors of metalloproteinases: structure, regulation and biological functions. (1997) Eur. J. Cell Biol.74:111–122.
22. Burger D., Rezzonico R., Li J. M., Modoux C., Pierce R. A., Welgus H. G., Dayer J. M. Imbalance between interstitial collagenase and tissue inhibitor of metalloproteinases 1 in synoviocytes and fibroblasts upon direct contact with stimulated T lymphocytes: Involvement of membrane-associated cytokines (1998) Arthritis Rheum. 41:1748–1759.

Chapter IIc: CHADL is associated with ascending aorta Z-scores (TAA)

Although AAO Z-scores are highly correlated with BAV status they are not always. For a variety of other reasons, one can have a dilated aorta in the absence of a BAV. Therefore, just AAO Z-scores were used in the SKAT-O analysis as a continuous variable to investigate if any genes were associated.

Results:

The results of this SKAT-O analysis revealed that variants in *CHADL* achieved exome-wide significance for association with TAA alone (using the ascending aorta as a proxy). *CHADL* was associated with TAA when AAO Z-scores were used as the sole continuous phenotype (p-value = $1.14x10^{-6}$; Figure 1). The Q-Q plots showed that there was a slight departure from observed vs. expected p-values, but the lambda value was <1 and therefore not due to population stratification (Supplementary Figure 1). Table 1 is a list of the variants within *CHADL* along with their CADD scores, the consequence of the nucleotide change, ExAC allele frequencies, their dbSNP rs identifiers, and the subjects BAV status. All 9 SNPs are found only in those with a BAV and are rare variants found at very low frequencies in the ExAC database (<1%). Additionally, all but two variants had CADD scores > 10, predicting that most of the variants are of large effect size.

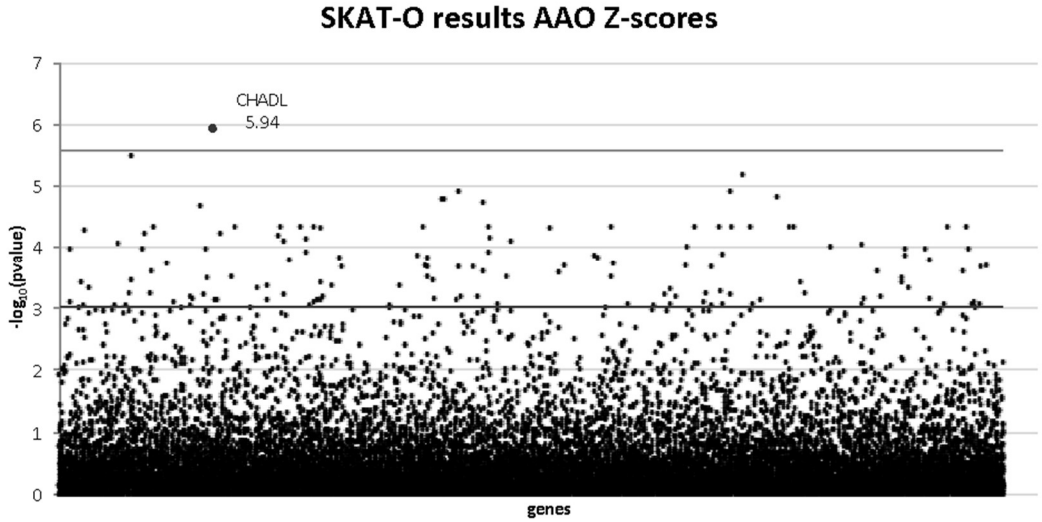


Figure 1: Manhattan plot for the ascending aorta (AAO) Z-scores showing exome-wide significance for *CHADL* ($1.14x10^{-6}$).

Of these nine rare variants, three occur within two subjects; therefore we investigated if these two samples were cryptically related or dublicates to ensure we were not seeing a spurious association. Comparing all individuals with a *CHADL* variant, identity by decent was performed to calculate the PI_HAT parameter. We found that no comparison was >0.1 and therefore no cryptic relatedness was identified (Supplementary Table 1). Hence, the three variants, which have essentially the same allele frequencies in the ExAC database, are probably historically the result of a single genetic event and remain in linkage disequilibrium. Therefore, we investigated if these variants were instead an ancestral haplotype using HAPCOMPASS. We could not phase all three variants, due to the read lengths of WES and the lack of informative variants, but we found that rs9619954 and rs9619955 are indeed in the same haplotype block (Supplementary Table 1). Given their very similar allele frequencies in the ExAC database, we can infer that the third variant, rs6002293, is also in this haplotype. Therefore, Q710R, D721N, rs6002293 (5'UTR) is an ancestral haplotype, which we will refer to as *CHADL*+3.

We can confirm *CHADL* association is not due to population differences by looking at PCA plot with samples that have a *CHADL* mutation. We see that subjects are scattered throughout the plot and therefore are not associated due to population structure (Supplementary Figure 2).

GENE	rsID	CHANGE	CHANGE	EXAC	Case	Ctrl	CADD
CHADL	rs77419992	SPLICE	Donor (1bp from site) K754	0.002125	1	0	11.3
CHADL	22:41631213	MISS	T733I	NA	1	0	14.2
CHADL	rs9619954	MISS	D721N	0.00646	2	0	21.4
CHADL	rs9619955	MISS	Q710R	0.00709	2	0	15.05
CHADL	rs201020621	SYN	R696	0.000612	1	0	13.87
CHADL	22:41633192	SYN/SPLICE	Donor (13bp from site) S628	NA	1	0	3.861
CHADL	22:41633605	MISS	A491T	NA	1	0	13.08
CHADL	rs8135399	SYN	R26	0.00538	1	0	10.27
CHADL	rs6002293	5'UTR	5'UTR (34 bp from TSS)	0.00669	2	0	na

Table 2: CHADL variants identified through whole exome sequencing of subjects in our TS cohort. Their dbSNP rs identifier is listed for those in the database, along with their consequence, ExAC allele frequencies (European non-Finish), their CADD score, and if the subjects had a BAV (cases). The CHADL+3 variants are highlighted.

A mini-gene assay to assess potential splicing mutations was performed for rs77419992 and 22:41633192. We show that 100% of the *CHADL* rs77419992 products are incorrectly spliced. The C to T mutation in the 3' exon 5 boundary has two products: 1) that has all of exon 5 but also has part of intron 5 and 2) is completely missing exon 5 (Figure 2). Variant 22:41633192 has normal splicing.

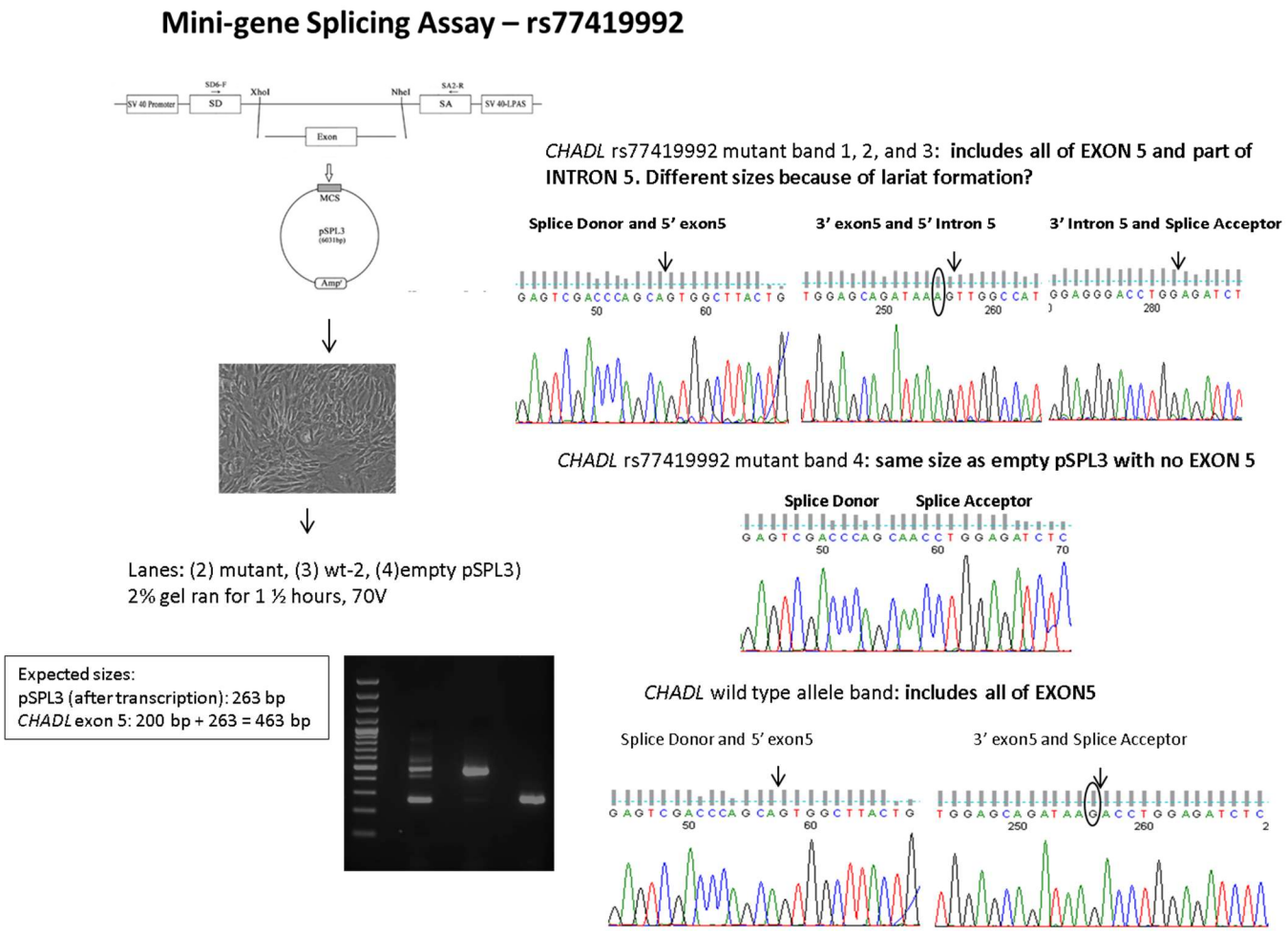


Figure 2: Minigene assay of *CHADL* rs77419992 variant and wildtype allele, showing that 100% of the rs77419992 variant is miss-spliced.

Methods:

(See TIMP3 & TIMP1 paper for more detailed methods concerning sample collection, sequencing, and phenotyping.)

AAO Z-scores were available for 106 samples, where 53 are BAV cases and 53 are BAV controls. For the variants identified in *CHADL*, all were verified using Sanger sequencing.

For the Q-Q plot: The observed median value of the chi-squared statistic for the null markers divided by the expected median value of the chi-squared statistic (approximately 0.456 for 1 df tests) is the "inflation factor," lambda. If lambda is less than or equal to 1, no adjustment is necessary.

HAPCOMPASS was used to phase variants of interest in close proximity to obtain haplotype information[1]. For subjects in the study that shared multiple rare variants, identity by decent was performed and the PI_HAT parameter was generated to insure there was no underlying cryptic relatedness, using PLINK v1.90b3g. A stringent PI_HAT parameter cutoff was used, where PI_HAT >0.1 considered to be cryptic relatedness.

	4	**1**	**2**	**5**	**3**	**6**	**7**
4	X	0	0	0	0	0.031	0
1	X	X	0.034	0	0	0	0
2	X	X	X	0	0	0.0276	0.0123
5	X	X	X	X	0	0	0
3	X	X	X	X	X	0	0
6	X	X	X	X	X	X	0
7	X	X	X	X	X	X	X

Supplementary Table 1: There were concerns about 4& 3 being related because they both have multiple CHADL mutations. An identity by descent estimation was performed and the PI_HAT parameter was assessed**.**

	Start_POS	End_POS	Start_SNPnumber	End_SNPnumber	score	chrom
BLOCK	41631250	41631282	253	254	8	22

SNPid	SNP_POS	SNP_number	Hap_0_allele	Hap_1_allele		
rs9619954	41631250	253	0	1		
rs9619955	41631282	254	0	1		

	Start_POS	End_POS	Start_SNPnumber	End_SNPnumber	score	chrom
BLOCK	41631250	41631282	370	371	32	22

SNPid	SNP_POS	SNP_number	Hap_0_allele	Hap_1_allele		
rs9619954	41631250	370	0	1		
rs9619955	41631282	371	0	1		

Supplementary Table 2: Haplo blocks using HAPCOMPASS: A fast cycle basis algorithm for accurate haplotype assembly of sequence data. Derek Aguia. Brown University

The mini-gene assay was performed using the pSPL3 vector, in which the mutant and wild type alleles were cloned in. Briefly, primers were designed with restriction enzyme sites and were designed to include flanking introns of ~250bp surrounding the exons (5 or 3) of *CHADL*. These PCR products were cloned into the pSPL3 vector and transformed into DH5 alpha cells. Clones were sequenced to confirm the mutant and wildtype allele presence. The purified vector was transfected into HEK293T cells, RNA was purified, and cDNA was generated, and splicing products were amplified using the SD6-Forward primer (TCTGAGTCACCTGGACAACC) and SA2-Reverse primer (ATCTCAGTGGTATTTGTGAGC). The splicing products were run on a 2% gel for 1 ½ hours at 70 volts.

Discussion:

CHADL is a member of the small leucine rich proteoglycan (SLRP) family of proteins. SLRPs are ECM molecules that regulate collagen fibrillogenesis and inhibit TGFβ activity[2]. These activities have intriguing possibilities for involvement in progression of aortic dilatation. Certainly the turnover of collagens in maintaining aortic wall integrity is critical and deficiencies in collagen structure has long been associated with TAA, such as the *COL3A1* mutations in Ehlers-Danlos syndrome associated TAA. Additionally, in Marfan syndrome dysregulation of TGFβ and in Loeys-Dietz syndrome mutations in TGFβ receptors have been shown to cause TAA (Figure 3)[3]. The association of potentially deleterious *CHADL* mutations with TAA in TS introduces the tantalizing possibility that these mutations could be involved in TAA progression via TGFβ signaling. In addition, TGFβ1 regulates TIMP1 and MMP gene expression so there could be a complex cascade of interactions in TAA progression that is somehow mediated by CHADL[4].

Given our findings we propose that CHADL plays a role in tissue homeostasis in the aorta and that the TAA-associated variants are deleterious and disrupt the tissue integrity of the extracellular matrix of the aorta. Three of the variants, rs9619954 (p.Q710R), rs9619955 (p.D721N), and rs6002293 (in

the 5' UTR) were found together in 2 subjects from our cohort, in which we demonstrated that there was no cryptic relatedness between them. Hence, the three SNPs are probably historically the result of a single genetic event and remain in linkage disequilibrium.

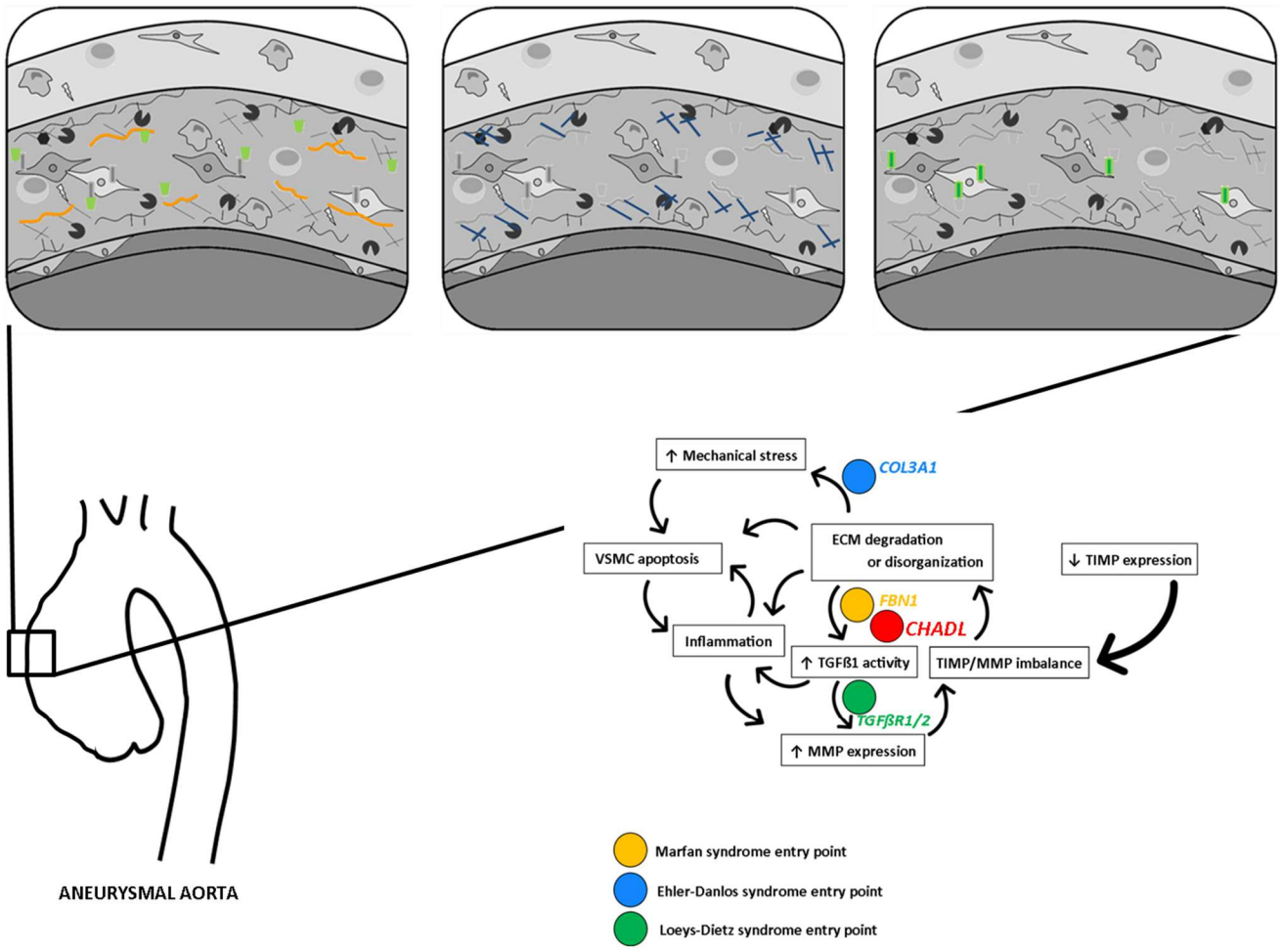


Figure 3: Current model of the genetic and molecular etiology of some syndromes associated with aortopathy that affect the TGFß pathway. The addition of *CHADL* to this model shows how it could also feed into the disease cycle by increasing TGFß signaling.

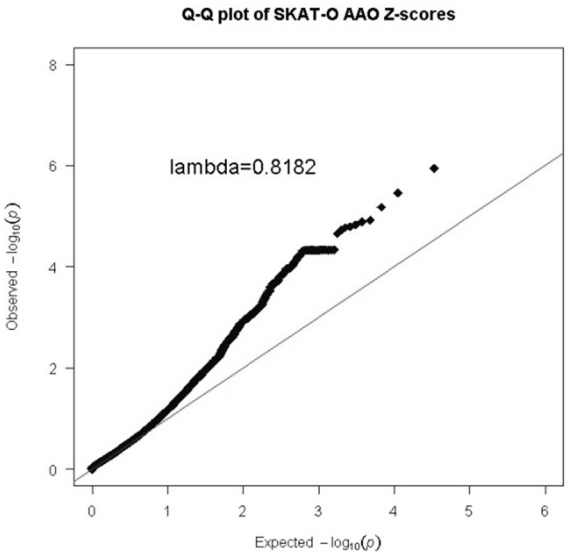
Four individuals were compound heterozygous for *TIMP3* and *CHADL* variants (Table 2). All four were heterozygous for the TIMP3 rs11547635 SNP, two also carry the linked *CHADL* variants described above (p.Q710R, p.D721N and rs6002293), one has the p.T733I missense mutation, and one is heterozygous for synonymous variant p.R696. Table 2 shows the genetic and clinical data for the BAV status, AR Z-scores, and AAO Z-scores for the compound heterozygotes mentioned above.

There is a correlation between being a *CHADL* +3 and *TIMP3* compound heterozygote, and AAO size. In fact, these are the two largest AAO Z-scores in our study cohort, suggesting that these two variants have a combinatorial effect. The CHADL 5' UTR variant is of unknown significance and may or may not contribute to pathogenesis. Although the sample size is small, it does not appear that being a *TIMP3/CHADL* compound heterozygote is significantly correlated with increased AAO Z-score as the Z-scores for those with a single *TIMP3* and *CHADL* variant are similar to those of individuals who have only a *CHADL* variant. All have a single copy of *TIMP1*.

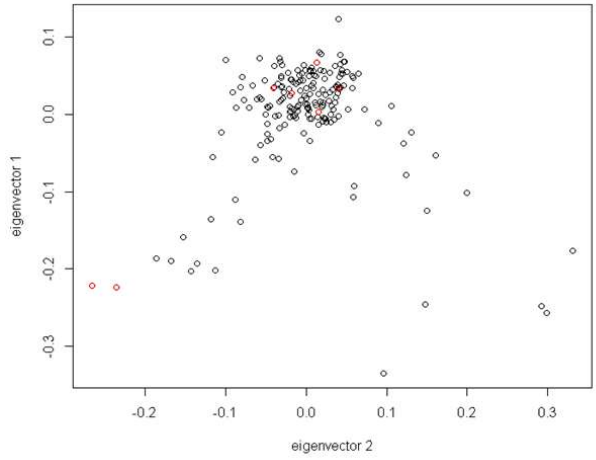
Subject /Age	BAV	AAO Z-score	AR Z-score	*CHADL* Variants	*CHADL* C-Score	*TIMP3* SNP	*TIMP1* copy number
1 / 46	yes	2.00	1.35	p.T733I	14.2	p.His83/p.Ser87	1
2 / 36	yes	2.25	1.57	p.R696=	10.27	p.His83/p.Ser87	1
3 / 10	yes	5.19	4.38	p.Q710R/p.D721N/5'UTR	15.05/21.4/na	p.His83/p.Ser87	1
4 / 64	yes	7.31	2.36	p.Q710R/p.D721N/5'UTR	15.05/21.4/na	p.His83/p.Ser87	1

Table 2: TIMP3/CHADL compound heterozygous study subjects with TIMP1 copy number. All had a BAV and significantly increased ascending aorta (AAO) Z-scores (>2). Subject age listed is the age (in years) at which the aortic dimensions were taken.

We hypothesize that the unique combination of TS and altered TIMPs and CHADL function destabilizes aortic extracellular matrix resulting in aortic disease. This study is critical to achieve our goal to identify biomarkers that could better predict who is at risk for serious aortic disease, instead of the currently insufficient imaging based assessments. Studies of TIMP1, TIMP3, and CHADL protein functions, in the context of TS or euploid males, will hopefully lead to targeted therapies that would slow or halt the progression of aortic disease in these high risk populations.



Supplementary Figure 1: Q-Q plot for the SKAT-O analysis of AAO Z-scores. Although it shows some deviation from the norm, the lambda value is <1, so no adjustment is needed



Supplementary Figure 2: PCA plot with samples that have a CHADL variant are highlighted in red. The subjects with a CHADL variant are not clustered or skewed toward a certain vector.

1. Aguiar, D. & Istrail, S. HapCompass: a fast cycle basis algorithm for accurate haplotype assembly of sequence data. *J Comput Biol* **19**, 577-90 (2012).
2. Tillgren, V. *et al.* The Novel Small Leucine-rich Protein Chondroadherin-like (CHADL) Is Expressed in Cartilage and Modulates Chondrocyte Differentiation. *J Biological Chemistry*. **290**, 918-925 (2015)
3. Neptune, E.R. *et al.* Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* **33**, 407-11 (2003).
4. Hall, M.C. *et al.* The comparative role of activator protein 1 and Smad factors in the regulation of Timp-1 and MMP-1 gene expression by transforming growth factor-beta 1. *J Biol Chem* **278**, 10304-13 (2003).

Chapter III:

A critical primary cilium component, CEP290, identified in a whole exome sequencing study of Down syndrome-associated atrioventricular septal defects

Holly Corbitt[1]; Stephanie Sherman, PhD[2]; Eleanor Feingold, PhD[4]; Roger Reeves, PhD[5]; Michael Zwick, PhD[3]; Cheryl L. Maslen, PhD[1,2,*]

[1]Knight Cardiovascular Institute, Oregon Health & Science University, Portland, Oregon; [2]Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, Oregon; [3]Department of Human Genetics, Emory University, Atlanta, GA 30033, USA; [4]Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA 15261; [5]Department of Physiology and McKusick Nathans Institute for Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

*Corresponding author

Email: maslenc@ohsu.edu (CLM)

In preparation for submission

Abstract

The goal of the current study was to identify the contribution of genetic variants to Down syndrome (DS) associated congenital heart defects. The defects associated with DS, such as congenital heart defects, are poorly understood. For example, the risk of an AVSD is 2000-fold higher in DS compared to the general population, yet the underlying etiology is still unknown. The partial prevalence of AVSD, 1 in 5 infants with DS, supports the premise that trisomy 21 acts as a significant risk factor, but other factors must be contributing to disease onset. To address this hypothesis, we performed a whole exome study (WES) composed of cases (DS with an AVSD) and controls (DS with no CHD) to investigate the presence of genetic modifiers that may be contributing to AVSD in the DS population (syndromic-AVSD). We implemented a kernel-based method, SKAT, to take a cumulative approach to detect the effects of rare variants. We preformed SKAT on approximately 300 genes representing the ciliome, based on previous work showing that Shh through the primary cilium is essential for the septation process. This analysis yielded *CEP290*, a critical component of the primary cilium, as a candidate gene contributing to syndromic AVSD. There were six case-specific variants found within *CEP290* that were predicted to be deleterious, all of which are in a functional or binding domain. We hypothesize that the rare variants within *CEP290* will cause defects in ciliogenesis and primary cilium function leading to aberrant Shh signaling.

Introduction

Congenital heart defects (CHD) are the most common birth defects and the largest contributor to infant mortality and morbidity, aside from infections[1-3]. With an incidence rate of nearly one percent globally, CHDs pose a serious global health concern, in spite of the advances made to improve their diagnoses and treatment[4-6]. Numerous studies have established that CHD are heritable and have revealed specific genetic variants contributing to their occurrence[9-14].

Trisomy 21, the cause of Down syndrome (DS), is the most common chromosomal aneuploidy in humans with a prevalence of 1 in every 700 live births. Even with such a high prevalence, it is unknown why nearly 50% of newborns with DS have some form of CHD[15,16]. The most common and also severe CHD in infants with DS is an atrioventricular septal defect (AVSD), which requires surgery at a very young age. With a prevalence of 1in 10,000 live births, AVSD is quite rare in the general population[17]. Conversely, in the DS population, the risk is increased by 2,000 fold, where ~20% of those with DS have an AVSD (syndromic-AVSD)[15]. Based on the known genetic component and the partial prevalence seen in DS, we hypothesize a threshold model in which the theoretical level of risk conferred by trisomy 21 with the addition of autosomal genetic variants are required to reach the disease threshold. Additionally, we propose that these genetic variants could have a large effect size in the DS population, but little to no effect in the general population.

An AVSD is a defect caused by in which the fusion of the mesenchymal cap of the primary atrial septum, the dorsal mesenchymal protrusion, and the atrioventricular (AV) endocardial cushions does not occur properly. The result is a large whole in the heart, instead of the typical four chambers, which causes mixing of the oxygen rich and oxygen poor blood.

Using genetic inducible fate mapping, it has been shown that Sonic hedgehog (Shh) signaling is required for AV septation during early cardiac morphogenesis[18]. In vertebrates, Shh signaling is transduced exclusively through the primary cilium. Therefore, the mechanisms that regulate ciliogenesis and the functions of the ciliary components governing this complex nano-machine are vital features of proper Shh signaling. We hypothesized that we would identify variants in gene(s) within the primary cilium network (the ciliome) to be associated with syndromic-AVSD. To this end, a WES study was conducted on the largest DS cohort to date. Our carefully phenotyped cohort consisted of 107 cases (DS with an AVSD) and 102 controls (DS with no CHD). We took a cumulative gene-based approach (SKAT) to test the burden of rare variants in genes within the

ciliome. As a result, we identified a major component of the primary cilium, *CEP290*, as a candidate gene contributing to the risk of syndromic-AVSD. This finding noticeably echoes a recent ENU recessive forward genetic screen of CHDs which also found mutations in *CEP290*, further supporting the link between Cep290 and AVSDs.

Methods

Study Subjects and Ethical approval

All of the DS individuals in this cohort have a complete trisomy 21, by karyotype.

Whole Exome Sequencing, Quality Control and Data Cleaning

In total, 281 DNA samples were submitted for exome sequencing and the exome capture kit Nimblegen V2 exome target kit was used to prepare the sequencing libraries. WES was performed by the NHLBI Resequencing & Genotyping Service at the University of Washington (D. Nickerson, US Federal Government contract number HHSN268201100037C). In summary, 7 samples failed post-sequencing QC and 274 samples passed post-sequencing QC. The average read depth for the targeted exome was 63X, with 82% of the target regions covered at greater than 20X. Reads were mapped to the hg19 UCSC genome build using the Burrows-Wheeler aligner, version 0.7.10. Variants were called using the GATK Unified Genotyper pipeline and annotated using SeattleSeq Annotation 138. In total, 295,611 final variants for the 274 samples were transferred to the Maslen lab for evaluation.

Data cleaning and filtering was performed using PLINK v1.90b3g[45], which 1) removed any variants with less than 99% genotyping rate, where 21,780 variants were removed; 2) removed individuals with more than 5% missing genotypes, where no individuals were removed; 3) excluded markers that fail the Hardy-Weinberg equilibrium test using a threshold of 1.0×10^{-6}, where 3,536 variants were removed.

A principal components analysis (PCA) was performed using the R package SNPRelate to calculate the eigenvectors (EVs) for each subject {Zheng et al. 2012}. Data were prepared for PCA analysis by taking common SNPs (MAF >5%) and pruning out SNPs in linkage disequilibrium with an $r^2 > 0.2$, stepping along five SNPs at a time within 50kb windows. PCA analysis of these variants were projected onto the 1000 genomes data, using the Peddy package, was used to detect population outliers that could confound the analysis (Supplementary Figure 1A) {Pedersen et al. 2017}. Population outliers were removed and the analysis was repeated in which there were no more outliers (Supplementary Figure 1B). In total, 65 subjects were detected to be of non-Europeans ancestry and were removed from the dataset. Further, we plotted cases and controls according eigenvector 1 and eigenvector 2 and saw no separation which confirmed no addition bias in population structure. We use the first three eigenvectors as covariates in most downstream analysis.

The final dataset contained 172,476 variants across 209 subjects, providing a total genotyping rate of 0.999866.

Gene-Based Statistical Analyses

We implemented a kernel-based method, SKATBinary linear kernel, to take a cumulative approach to detect the effects of variants and their association with syndromic-AVSD[19]. This is a statistically powerful method because it aggregates variants by gene, instead of analyzing individual variants. Based on our hypothesis that genes within the ciliome are implicated in syndromic-AVSD, we took a hypothesis driven approach and only tested genes within the ciliome. The gene list we used was the "gold standard" ciliome list in which there are 300 well curated and annotated genes[20]. These variants were grouped into their respective genes, hg19_refGene, and were analyzed using the sequenced based kernel approach. In summary, 4,586 total SNPs were grouped into 299 total

genes; therefor the significance threshold is $1.67X10^{-4}$. A quantile-quantile plot was generated to investigate any departure from the observed and expected p-values.

Variant Validation and Annotation

All variants of interest were Sanger sequenced for validation purposed and confirmed. We used multiple computational tools and standardized guidelines to assess the pathogenicity of each variant of interest, including Combined Annotation Dependent Depletion (CADD) scores, their allele frequency in the ExAC database, Genomic Evolutionary Rate Profiling (GERP) scores, and PolyPhen-2 scores[21-24]. Variants of interest were mapped to their location within their respective gene and the genes' domains and functional sites using PROSITE myDomains application[25].

Results

Association of *CEP290* with syndromic AVSD

The SKAT analysis yielded one gene, *CEP290*, which reached the significance threshold corrected for multiple testing (Figure 1). Cep290 is a critical component of the primary cilium, where it is involved in regulating ciliogenesis, trafficking, and functions of the primary cilium[30-32,36]. A quantile-quantile plot was generated to investigate any departure from the observed and expected p-values (Supplementary Figure 2). All variants were validated using Sanger sequencing.

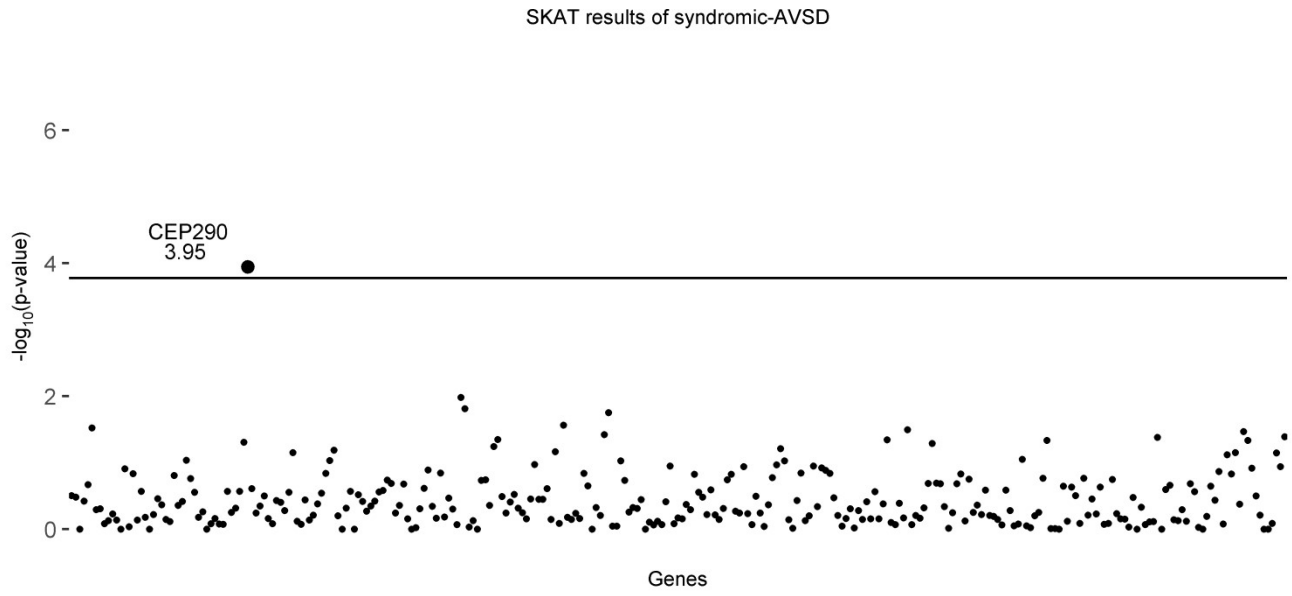


Figure 1: Scatter plot of SKAT results on the WES data for syndromic-AVSD. On the x-axis are genes and on the y axis is the negative log transformed p-value associated with each gene. The blue line at 3.775 is the significance threshold after correcting for multiple testing, Bonferroni. *CEP290* is the only gene that is significantly associated with syndromic-AVSD.

CEP290 variants and annotation

We used multiple computational tools and standardized guidelines to assess the pathogenicity of each *CEP290* variant. In total, six variants were classified as pathogenic, summarized in Table 1, and they each fall within an important functional or binding domain of Cep290.

Variant	ExAC Allele freq. European (Non-Finn)	CADD	GERP	POLY PHEN
p.D299N	0.0002%	25.1	4.77	0.999, probably damaging
p.D433G	0.2%	24.7	5.84	0.943, probably damaging
p.R1622C	na	35	5.67	0.959, probably damaging
p.R1729W	0.04%	26.4	4.81	1, probably damaging
p.E1985K	na	34	5.23	0.976, probably damaging
p.D2426N	0.012%	34	5.67	0.999, probably damaging

Table 1: The *CEP290* mutations identified to be associated with syndromic-AVSD.

All of these variants were not found in our control cohort and are absent from, or at very low frequency, in the ExAC database. Computational algorithms used to predict the variants effect

(CADD, GERP, and PolyPhen-2) all indicate a high probability that these are deleterious variants, or mutations. A CADD score >20 indicates the top 1% most deleterious variants possible throughout the human genome, and note that all six *CEP290* variants have a score much greater than 20. Of the six variants, four of the *CEP290* mutations are in the ciliogenesis regulatory domains, and four mutations are in protein-protein binding domains of *CEP290*, with two in both. Hence, all six of the mutations have the potential to interfere with ciliogenesis or ciliary function. Mapping the variants to their location within *CEP290* shows that each fall within a functional or binding domain (Figure 2).

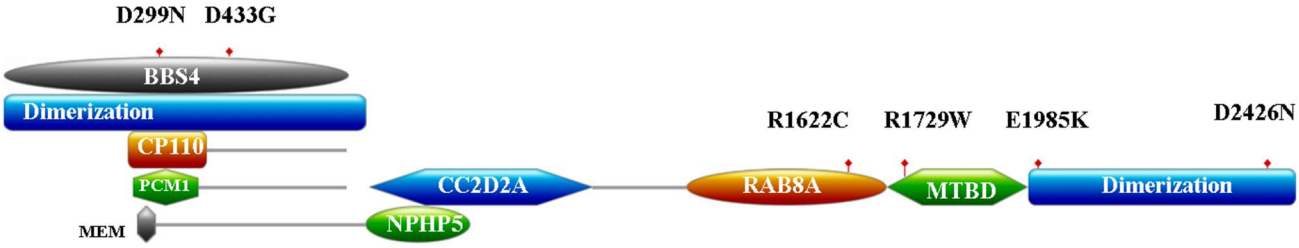


Figure 2: The *CEP290* mutations identified to be associated with syndromic-AVSD were mapped (red dots) to *CEP290*. Mapping the identified rare variants to Cep290 domains reveals that all six of the variants are within a functional region.

Discussion

We have completed a WES study, on the largest DS cohort to date, to identify genetic variants that are associated with AVSD in infants with DS. Our carefully phenotyped cohort consisted of 107 cases (DS with an AVSD) and 102 controls (DS with no CHD). By taking a cumulative gene-based approach to test the effects of rare variants in the ciliome, we have identified a major component of the primary cilium, *CEP290*, as a candidate gene contributing to the risk of syndromic-AVSD. Evidence for the fundamental role of the primary cilium in heart development has been shown through several mouse studies. Although these studies have focused on different components or mechanisms of disease, the common theme which arose from all of them is that CHDs occur when the primary cilium is disrupted. More specifically, it has been shown that the disruption of proper ciliogenesis and ciliary function can alter Shh signaling and result in CHDs, including AVSD[26-28]. Our

candidate gene, *CEP290*, is a critical component of the cilium and plays a role in regulating ciliogenesis and protein recruitment to the cilium.

The primary cilium is the exclusive compartment for Shh signal transduction. Therefore, the mechanisms that regulate ciliogenesis, including the timing, length and number of cilia, are important features of proper Shh signaling[27,29]. Previous protein-protein interaction studies have established that Cep290 interacts with its protein partners at two locations at the primary cilium, the transition zone and centriolar satellites (Figure 3)[30]. At the cellular level, Cep290 is involved the formation of the cilium, ciliogenesis, and the function of the cilium[31-32]. The N and C terminus of Cep290 undergo a conformational change to dimerize into a close inactive state. This auto-inhibitory mechanism of Cep290's N and C terminus, along with the binding of Cp110, prevents the primary cilium from forming while cells are mitotically active. This inhibition is released when cells are in the G0/G1 phase of the cell cycle and ciliogenesis begins, which is regulated through the release of Cp110 and of the N and C-terminal dimer[33-35]. Further, to mediate ciliogenesis and the function of the cilium, Cep290 interacts with Rab8a, the BBSome, microtubules, etc.[36,32,37]. Proper ciliogenesis to yield a functional primary cilium is required for proper Shh signaling and disturbance to the normal formation and function of the primary cilium can alter signaling sensitivity[27].

.

Although the specific binding and functional regions within Cep290 have been identified, the residues that participate have not yet been resolved. Hence, all six of the mutations have the potential to interfere with ciliogenesis, ciliary function, and ultimately Shh signaling.

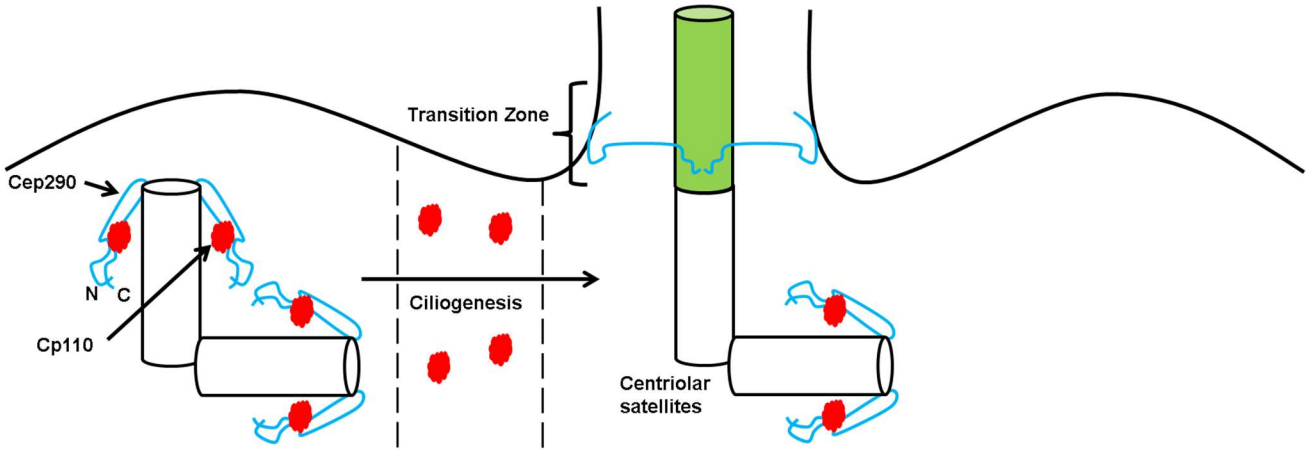


Figure 4: Model of Cep290 localization during ciliogenesis and regulation of the dimerization of the N and C-terminal domains. Cep290 is in a closed and inhibited state by its N and C termini and CP110 during the cell cycle. Upon release of CP110 and a conformational change releasing the N and C termini, the protein frees from its closed state. This event triggers ciliogenesis and in its open form it is able to bind its partners at the transition zone. Additionally, Cep290 has been shown to localize and interact with some of its partners at the centriolar satellites. At both of these sites, Cep290 interacts with its partners to regulate vesicle trafficking and entry into the primary cilium.

Both the N and C terminus of Cep290 have a predicted coiled-coil structure[33]. In this assembly, hydrophobic residues are buried in the core and hydrophilic side chains are exposed on the outside of the protein. To further stabilize the coiled-coil, positive and negative charged amino acids are opposite one another on the edge of the core, which forms a salt bridge. This quaternary protein folding could be affected by the proposed variants. Variants p.D299N, p.D433G, p.E1985K, and p.D2426N all change the chemical properties of the side chains while also introducing some steric hindrance concerns. Therefore, we hypothesize that the variants located in the N and C-terminal dimerization domains will affect the stability of the inactive close state, resulting in aberrant ciliogenesis which could lead to altered cilium length and number of cilia.

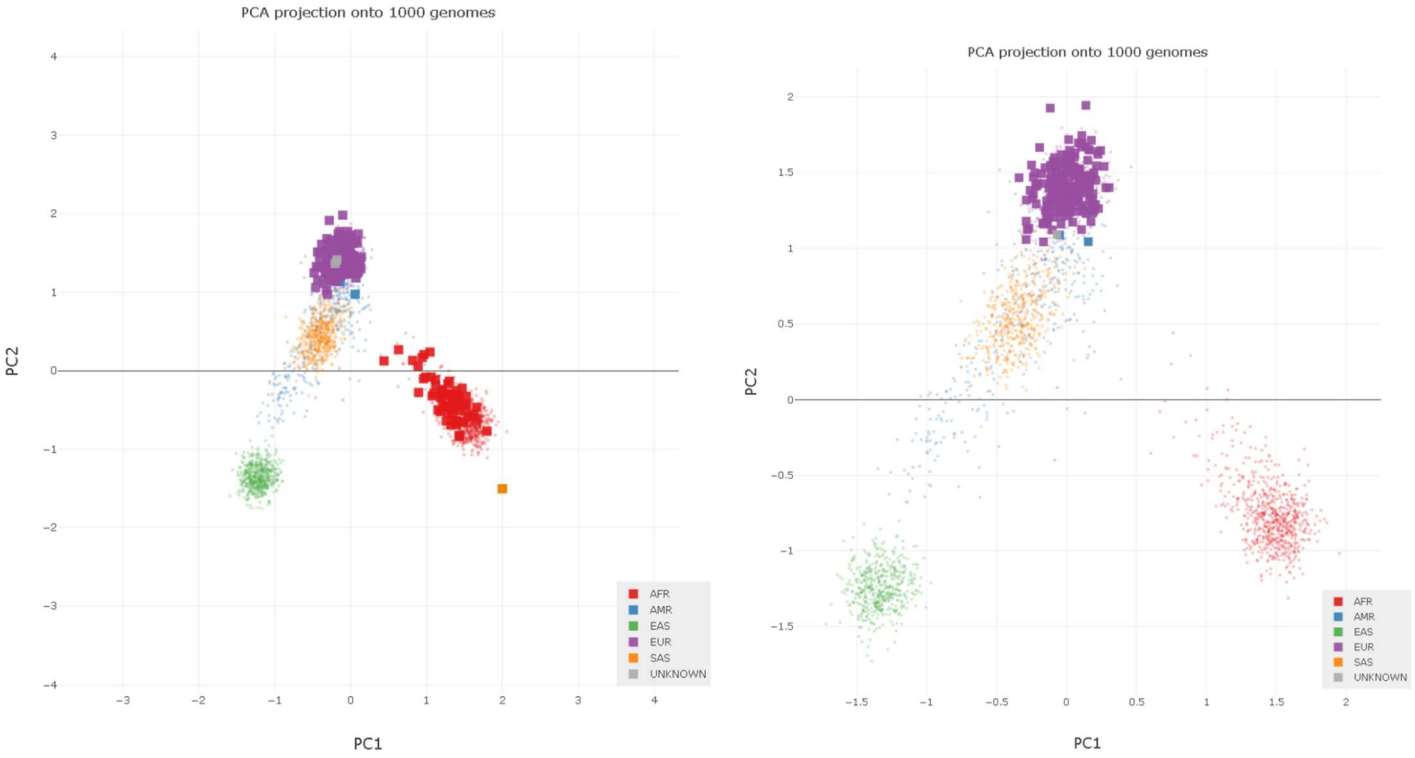
Variants p.D299N, p.D433G, and p.R1622C are located in critical binding regions of Cep290 and could affect canonical protein-protein interactions. Through these different interactions Cep290 functions in specific pathways that facilitate protein trafficking to and from the cilia. Both p.D292N and p.D427G are located in the BBS4 binding domain and Cep290 interacts with the BBSome at the

centriolar satellites and transition zone. The BBSome is a holo-complex of eight core BBS proteins[38]. Cep290 interacts with the BBSome to regulate trafficking and entry of ciliary components, including Ptch1 and Smo[31,37,39]. The stability of the BBSome is critical to its function, as it must be in its holo-complex form to gain entry into the cilium; therefore disruption leads to altered Shh signaling[38,40,41]. Variant p.R1616C is located in the Rab8a binding domain of Cep290. Cep290 mediated targeting of Rab8a has been shown to collaborate with the BBSome to promote ciliogenesis[32]. Rab8a is fundamental in modulating ciliary cargo packaged in vesicles to the primary cilium. Cep290 directly binds Rab8a and may play a role in its recruitment and docking. Additionally, the BBSome regulates Rab8a entry into the cilium, where only the GTP-active form of Rab8a is allowed to enter[42,43]. Therefore, these variants in CEP290 could affect the formation/localization of the BBSome could influence the localization of Rab8a.

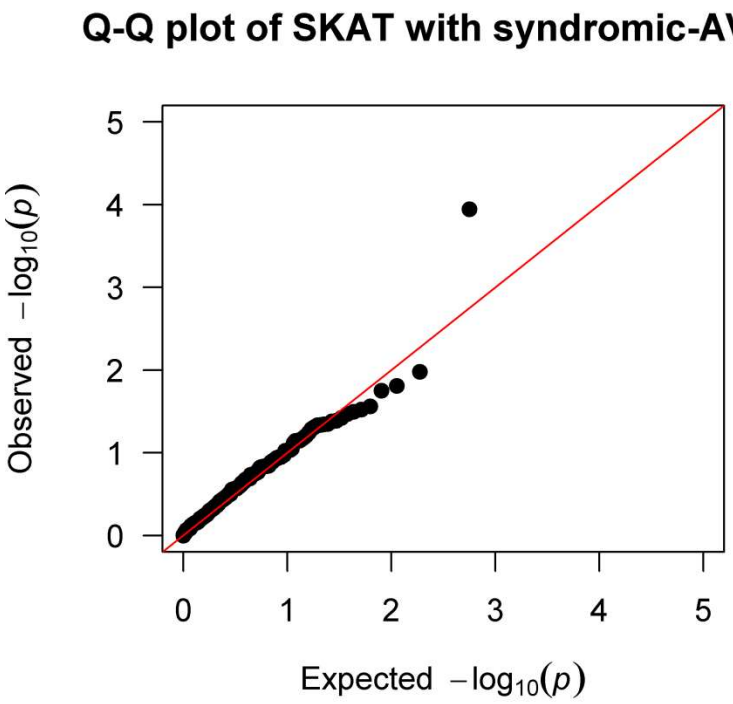
The varying effects and manifestations of variants in the context of different genetic backgrounds have been well established. For example, our lab along with our collaborator Dr. Roger Reeves has confirmed this phenomenon on the trisomy 21 genetic background using a widely used DS mouse model, Ts65Dn26. *CRELD1* was one of the first genes identified as a risk factor for AVSD. Subsequent mouse studies showed that null mutations of *CRELD1* were benign on a euploid background, but had a high incidence rate of CHDs on the Ts65Dn background38. These observations are evidence that the trisomy 21 genetic background is likely influencing the effects of genetic variants on the development of a CHD phenotype. We hypothesize that these mutations in *CEP290* are an underlying cause of AVSDs in infants with DS and will cause disruption of primary cilium function and Shh signaling. Interestingly, cells from a trisomy 21 mouse model were found to be deficient in their Shh signaling capability and response[44]. Thus, we hypothesize a threshold model in which trisomy 21 Shh deficiency predisposes infants with DS to AVSD, but that a second hit leads to an AVSD. This study adds significant knowledge to the etiology of syndromic-AVSDs and to

the potential complex effect of the trisomic 21 genetic background leading to the high prevalence of syndromic-AVSD.

Supplemental Information



Supplemental Figure 1: A principal component analysis results detected population stratification in our WES cohort. Using the Peddy package, common variants were used to detect samples that were not of European ancestry. **1A** shows 65 samples (red) of AFR ancestry and after their removal **1B** shows no more samples as outliers, leaving a final sample size of 209 cases and controls.



Supplemental Figure 2: Q_Q plot shows no deviation from observed and expected p-values.

1. Moller, J.H., Allen, H.D., Clark, E.B., Dajani, A.S., Golden, A., Hayman, L.L., Lauer, R.M., Marmer, E.L., McAnulty, J.H., Oparil, S., et al. (1993). Report of the task force on children and youth. American Heart Association. Circulation 88, 2479-2486.
2. Boneva, R.S., Botto, L.D., Moore, C.A., Yang, Q., Correa, A., and Erickson, J.D. (2001). Mortality associated with congenital heart defects in the United States: trends and racial disparities, 1979-1997. Circulation 103, 2376-2381.
3. Shuler, C.O., Tripathi, A., Black, G.B., Park, Y.M., and Jerrell, J.M. (2013). Individual risk factors and complexity associated with congenital heart disease in a pediatric medicaid cohort. South Med J 106, 385-390.
4. van Rijen, E.H., Utens, E.M., Roos-Hesselink, J.W., Meijboom, F.J., van Domburg, R.T., Roelandt, J.R., Bogers, A.J., and Verhulst, F.C. (2005). Current subjective state of health, and longitudinal psychological well-being over a period of 10 years, in a cohort of adults with congenital cardiac disease. Cardiol Young 15, 168-175.
5. Elixhauser, A., and Owens, P. (2006). Adverse Drug Events in U.S. Hospitals, 2004: Statistical Brief #29. In Healthcare Cost and Utilization Project (HCUP) Statistical Briefs. (Rockville (MD).
6. Fahed, A.C., Gelb, B.D., Seidman, J.G., and Seidman, C.E. (2013). Genetics of congenital heart disease: the glass half empty. Circ Res 112, 707-720.
7. Dennis, N.R., and Warren, J. (1981). Risks to the offspring of patients with some common congenital heart defects. Journal of medical genetics 18, 8-16.
8. Emanuel, R., Somerville, J., Inns, A., and Withers, R. (1983). Evidence of congenital heart disease in the offspring of parents with atrioventricular defects. Br Heart J 49, 144-147.
9. Cripe, L., Andelfinger, G., Martin, L.J., Shooner, K., and Benson, D.W. (2004). Bicuspid aortic valve is heritable. Journal of the American College of Cardiology 44, 138-143.
10. Flaquer, A., Baumbach, C., Pinero, E., Garcia Algas, F., de la Fuente Sanchez, M.A., Rosell, J., Toquero, J., Alonso-Pulpon, L., Garcia-Pavia, P., Strauch, K., et al. (2013). Genome-wide linkage analysis of congenital heart defects using MOD score analysis identifies two novel loci. BMC genetics 14, 44.
11. Cordell, H.J., Bentham, J., Topf, A., Zelenika, D., Heath, S., Mamasoula, C., Cosgrove, C., Blue, G., Granados-Riveron, J., Setchfield, K., 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, 822-824.
12. Soemedi, R., Wilson, I.J., Bentham, J., Darlay, R., Topf, A., Zelenika, D., Cosgrove, C., Setchfield, K., Thornborough, C., Granados-Riveron, J., et al. (2012). Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. American journal of human genetics 91, 489-501.
13. Robinson, S.W., Morris, C.D., Goldmuntz, E., Reller, M.D., Jones, M.A., Steiner, R.D., and Maslen, C.L. (2003). Missense mutations in CRELD1 are associated with cardiac atrioventricular septal defects. American journal of human genetics 72, 1047-1052.
14. Al Turki, S., Manickaraj, A.K., Mercer, C.L., Gerety, S.S., Hitz, M.P., Lindsay, S., D'Alessandro, L.C., Swaminathan, G.J., Bentham, J., Arndt, A.K., et al. (2014). Rare variants in NR2F2 cause congenital heart defects in humans. American journal of human genetics 94, 574-585.
15. Freeman, S.B., Bean, L.H., Allen, E.G., Tinker, S.W., Locke, A.E., Druschel, C., Hobbs, C.A., Romitti, P.A., Royle, M.H., Torfs, C.P., 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, 173-180.
16. Freeman, S.B., Taft, L.F., Dooley, K.J., Allran, K., Sherman, S.L., Hassold, T.J., Khoury, M.J., and Saker, D.M. (1998). Population-based study of congenital heart defects in Down syndrome. Am J Med Genet 80, 213-217.
17. Hartman, R.J., Riehle-Colarusso, T., Lin, A., Frias, J.L., Patel, S.S., Duwe, K., Correa, A., Rasmussen, S.A., and National Birth Defects Prevention, S. (2011). Descriptive study of nonsyndromic atrioventricular septal defects in the National Birth Defects Prevention Study, 1997-2005. American journal of medical genetics Part A 155A, 555-564.
18. Hoffmann, A.D., Peterson, M.A., Friedland-Little, J.M., Anderson, S.A., and Moskowitz, I.P. (2009). sonic hedgehog is required in pulmonary endoderm for atrial septation. Development 136, 1761-1770.

19. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. American journal of human genetics 89, 82-93.
20. van Dam, T.J., Wheway, G., Slaats, G.G., Group, S.S., Huynen, M.A., and Giles, R.H. (2013). The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. Cilia 2, 7.
21. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics 46, 310-315.
22. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-291.
23. Cooper, G.M., Stone, E.A., Asimenos, G., Program, N.C.S., Green, E.D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. Genome Res 15, 901-913.
24. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Current protocols in human genetics Chapter 7, Unit7 20.
25. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., Cuche, B.A., de Castro, E., Lachaize, C., Langendijk-Genevaux, P.S., and Sigrist, C.J. (2008). The 20 years of PROSITE. Nucleic Acids Res 36, D245-249.
26. Li, Y., Klena, N.T., Gabriel, G.C., Liu, X., Kim, A.J., Lemke, K., Chen, Y., Chatterjee, B., Devine, W., Damerla, R.R., et al. (2015). Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature 521, 520-524.
27. Caspary, T., Larkins, C.E., and Anderson, K.V. (2007). The graded response to Sonic Hedgehog depends on cilia architecture. Developmental cell 12, 767-778.
28. Berbari, N.F., Lewis, J.S., Bishop, G.A., Askwith, C.C., and Mykytyn, K. (2008). Bardet-Biedl syndrome proteins are required for the localization of G protein-coupled receptors to primary cilia. Proc Natl Acad Sci U S A 105, 4242-4246.
29. Rohatgi, R., Milenkovic, L., and Scott, M.P. (2007). Patched1 regulates hedgehog signaling at the primary cilium. Science 317, 372-376.
30. Craige, B., Tsao, C.C., Diener, D.R., Hou, Y., Lechtreck, K.F., Rosenbaum, J.L., and Witman, G.B. (2010). CEP290 tethers flagellar transition zone microtubules to the membrane and regulates flagellar protein content. J Cell Biol 190, 927-940.
31. Barbelanne, M., Hossain, D., Chan, D.P., Peranen, J., and Tsang, W.Y. (2015). Nephrocystin proteins NPHP5 and Cep290 regulate BBSome integrity, ciliary trafficking and cargo delivery. Hum Mol Genet 24, 2185-2200.
32. Kim, J., Krishnaswami, S.R., and Gleeson, J.G. (2008). CEP290 interacts with the centriolar satellite component PCM-1 and is required for Rab8 localization to the primary cilium. Hum Mol Genet 17, 3796-3805.
33. Schafer, T., Putz, M., Lienkamp, S., Ganner, A., Bergbreiter, A., Ramachandran, H., Gieloff, V., Gerner, M., Mattonet, C., Czarnecki, P.G., et al. (2008). Genetic and physical interaction between the NPHP5 and NPHP6 gene products. Hum Mol Genet 17, 3655-3662.
34. Spektor, A., Tsang, W.Y., Khoo, D., and Dynlacht, B.D. (2007). Cep97 and CP110 suppress a cilia assembly program. Cell 130, 678-690.
35. Tsang, W.Y., Bossard, C., Khanna, H., Peranen, J., Swaroop, A., Malhotra, V., and Dynlacht, B.D. (2008). CP110 suppresses primary cilia formation through its interaction with CEP290, a protein deficient in human ciliary disease. Developmental cell 15, 187-197.
36. Drivas, T.G., Holzbaur, E.L., and Bennett, J. (2013). Disruption of CEP290 microtubule/membrane-binding domains causes retinal degeneration. J Clin Invest 123, 4525-4539.
37. Zhang, Y., Seo, S., Bhattarai, S., Bugge, K., Searby, C.C., Zhang, Q., Drack, A.V., Stone, E.M., and Sheffield, V.C. (2014). BBS mutations modify phenotypic expression of CEP290-related ciliopathies. Hum Mol Genet 23, 40-51.
38. Zhang, Q., Seo, S., Bugge, K., Stone, E.M., and Sheffield, V.C. (2012). BBS proteins interact genetically with the IFT pathway to influence SHH-related phenotypes. Hum Mol Genet 21, 1945-1953.
39. Jin, H., White, S.R., Shida, T., Schulz, S., Aguiar, M., Gygi, S.P., Bazan, J.F., and Nachury, M.V. (2010). The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. Cell 141, 1208-1219.

40. Lechtreck, K.F., Johnson, E.C., Sakai, T., Cochran, D., Ballif, B.A., Rush, J., Pazour, G.J., Ikebe, M., and Witman, G.B. (2009). The Chlamydomonas reinhardtii BBSome is an IFT cargo required for export of specific signaling proteins from flagella. J Cell Biol 187, 1117-1132.
41. Blacque, O.E., Li, C., Inglis, P.N., Esmail, M.A., Ou, G., Mah, A.K., Baillie, D.L., Scholey, J.M., and Leroux, M.R. (2006). The WD repeat-containing protein IFTA-1 is required for retrograde intraflagellar transport. Mol Biol Cell 17, 5053-5062.
42. Nachury, M.V., Loktev, A.V., Zhang, Q., Westlake, C.J., Peranen, J., Merdes, A., Slusarski, D.C., Scheller, R.H., Bazan, J.F., Sheffield, V.C., et al. (2007). A core complex of BBS proteins cooperates with the GTPase Rab8 to promote ciliary membrane biogenesis. Cell 129, 1201-1213.
43. Loktev, A.V., Zhang, Q., Beck, J.S., Searby, C.C., Scheetz, T.E., Bazan, J.F., Slusarski, D.C., Sheffield, V.C., Jackson, P.K., and Nachury, M.V. (2008). A BBSome subunit links ciliogenesis, microtubule stability, and acetylation. Developmental cell 15, 854-865.
44. Currier, D.G., Polk, R.C., and Reeves, R.H. (2012). A Sonic hedgehog (Shh) response deficit in trisomic cells may be a common denominator for multiple features of Down syndrome. Prog Brain Res 197, 223-236.
45. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7.

Chapter IV:

AVSD has heterogeneous genetic contributions that include genes important in cilia function and cardiovascular system development

Holly Corbitt; Cheryl L. Maslen, PhD*

Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR
Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR

*Corresponding author

Email: maslenc@ohsu.edu (CLM)

In preparation for submission

Abstract

Congenital heart defects (CHD) are the most common birth defect affecting ~40,000 births per year. One of the most severe forms of CHD is an atrioventricular septal defect (AVSD), which is the result of aberrant development and fusion in the atrioventricular junction of the developing heart. While most occurrences of AVSDs are associated with genetic syndromes, such as Down syndrome, they can also occur sporadically and without any other defect as non-syndromic AVSD. It has been established that AVSDs have an extremely complex and oligogenic architecture. While some genes and risk factors have been identified, the majority of sporadic non-syndromic AVSD cases remain unknown. To further our understanding of the genetic underpinnings of AVSD we used whole exome sequencing (WES) investigate the genetic variation in a cohort of non-syndromic AVSD cases. By applying both a case-control gene burden analysis and a deleterious rare variant filtering strategy we identified genes previously known to play a role in CHDs and novel genes in pathways not yet associated with CHDs. Specifically, we identified known mechanisms such as cardiac transcription regulators (*GATA6*, *ZFPM2*, and *FGF10*), ciliome genes (*GLI3*, *GLI1*, *PTCH1*, and *SMO*), and a recently discovered modifier *JAM2*; and novel mechanisms such as transcription factors (*ATF1* and *MYC*), signaling regulators (*COPS3*, *ERN1*, *GNG5*, and *P2RX5*), and an ECM component (*LAMB1*).

Introduction

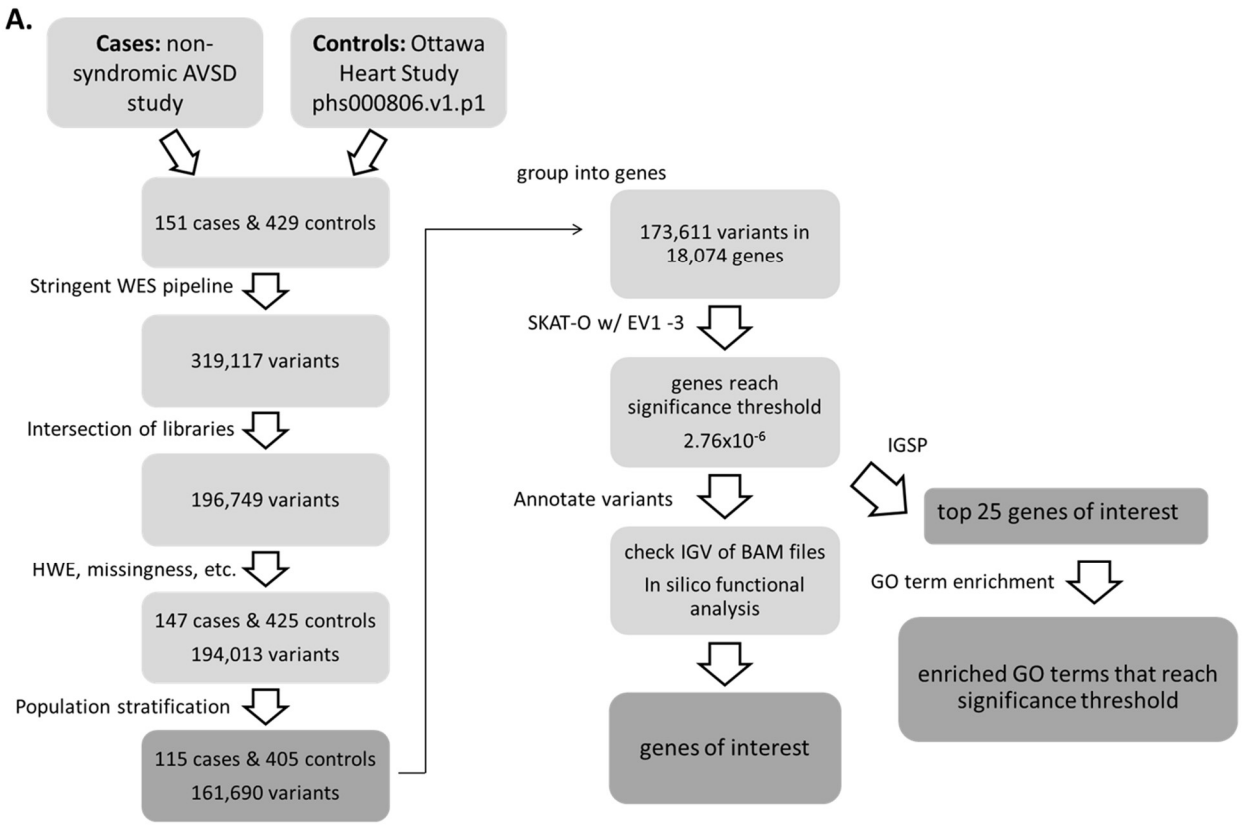
Atrioventricular septal defects (AVSD) are a clinically severe congenital heart defect (CHD) that accounts for about 7% of all CHDs and is most often seen in children with Down syndrome[1]. The majority of the remaining cases are sporadically occurring non-syndromic AVSD. While the risk of AVSD is relatively low in the general population, the overall life impact and poor prognosis is significant even after surgical correction[2,3] There have been numerous investigations into the genetic mechanisms underlying both familial and sporadic non-syndromic AVSD that point to a heterogeneous contribution of genes involved in transcription, epigenetic modifiers, early development, and signal transduction[4-7]. In human studies, *de novo* or inherited mutations have

been identified in key transcription regulators, including the *GATA* family[6,8,9], *NR1D2*[5], *NR2F2*[10], *NKX2-5*[11], *ZFPM2*[8,12], the *T box* family[6,13], and the *NOTCH* family[14]. One of the first genes to be associated with AVSD, *CRELD1*, has been found to have both recurrent and private missense mutations in humans[15]. Traditional human genetic strategies have been limited due to small number of cases, locus heterogeneity, and incomplete and variable penetrance in AVSD. Therefore, much of what we know about the genetic and molecular pathogenesis of AVSD is from animal models. Notably, a large forward genetic screen of mice with CHDs, including AVSDs, showed that the cilia and cilia transduced signaling was the primary mechanism for the pathogenesis of CHDs[16]. Further supporting these genetic studies, biochemical and molecular experiments in mice have elucidated proteins involved in critical signaling pathways s that when disrupted lead to complete or partial AVSDs such as the Gata family, Creld1, Notch, Tbx2, Shh, and Pdgfr-a among others[15,17-21]. These studies and others have elucidated the molecular and cellular events critical for proper heart development.

One of the key steps in the development of the heart is the formation of the valves and septa, which is critical for the separation of oxygen-rich blood from oxygen-poor blood. The molecular events of early heart development begin when two distinct populations of cells emerge, the first and second heart fields[22]. The second heart field progenitor cells receive Sonic hedgehog (Shh) paracrine signaling that trigger their migration and differentiation to populate the dorsal mesenchymal protrusion and the mesenchymal cap, which forms the base of the primary atrial septum[23]. When these two structures fail undergo coordinated development and properly fuse with the endocardial cushions, an atrioventricular septal defect can result. The migration and differentiation of the second heart field has been shown to be essential for this cardiac septation process which is mediated by Sonic hedgehog (Shh) signaling through the primary cilia[23,24]. The primary cilia is a complex structure that is composed of hundreds of tightly regulated components that play a variety of roles in the cell, such as chemical sensation, signal transduction, and control of cell growth. This network of

proteins, lipids, and chemicals is referred to as the ciliome. While it is clear that Shh signaling and proper functioning of the ciliome is critical for septation, we acknowledge that it could also be regulated through unknown mechanisms within or outside of the ciliome.

The goal of this study is to further our knowledge of the genetic underpinnings of AVSD. Our study has yielded new evidence about the genes that may play a role in the genetic mechanisms of AVSD. Specifically, we have identified contributory variants in genes within the ciliome, genes known to play a role in cardiovascular system development, and novel genes and pathways not yet known to contribute to congenital heart development.



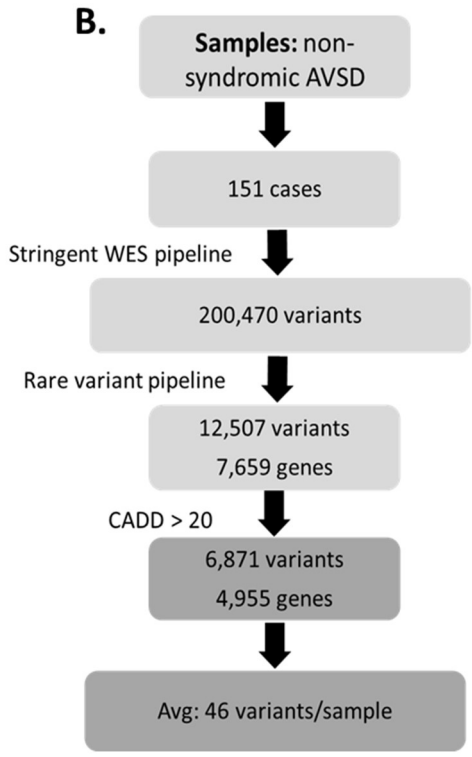


Figure 1: Approach for the analysis of the non-syndromic AVSD whole exome sequencing study. **A.** A cohort of 115 cases and 405 controls was analyzed using SKAT-O and IGSP. Briefly: 1) case and control raw files were combined and variants were called using a stringent pipeline 2) only variants in regions intersecting both libraries was kept 3) quality control metrics for HWE, missingness, etc. were used further filter variants 4) PCA analysis was used to remove population outliers 5) remaining cases and controls and their variants were grouped into genes and analyzed using a gene-based association analysis 6) genes and their respective p-values were analyzed through an integrated gene network and phenotype information analysis where the top 25 genes were investigated for GO term enrichment. **B.** The case cohort of 151 samples was analyzed separately to identify rare variants. Deleterious rare variant criteria was: variants with MAF > 1% in any population, variants not predicted to impact transcription or protein amino acid content, or to be annotated as a protein interaction amino acid, variants with quality score < 500, depth < 20 if not a known dbSNP site or depth < 10 if known dbSNP site, paralogous sequence (^MUC,/-/, ^OR), genotype quality score < 90 for novel variant, and a scaled CADD score > 20.

Methods:

Phenotyping and subject recruitment

The phenotype of interest in this study is an atrioventricular septal defect or AVSD (including complete AVSD, an intermediate transitional AVSD, partial AVSD ASD primum, and a single ventricle with an unbalanced AV canal) that was confirmed by a clinical diagnosis.

Whole exome sequencing of the non-syndromic and syndromic AVSD cohort

Library capture using the Niblegen V2 kit and whole exome sequencing (WES) using an Illumina HiSeq 2000 was performed by the University of Washington where VCF files and raw BAM files were transferred to the Maslen lab. In total, 151 non-syndromic AVSD samples passed sequencing quality control metrics and were released for analysis.

dbGAP control data set: Ottawa Heart Study phs000806.v1.p1

The control WES cohort was obtained through the MIGen_ExS: Ottawa Heart Study in which only the control group (429 samples) was utilized in the case-control approach of the study.

Case-control WES pipeline and analysis

A pipeline was created based on the GATK best practices to combine the case and control datasets. The two WES datasets were sequenced using different library capture kits and under different experimental conditions. To solve any potential biases or discrepancies, all BAM files were un-map, realigned to hs37d5, and merged for joint high stringency variant calling using picard, samtools, and GATK. Only regions in the intersecting the two libraries, 39.4Mb, were used for quality control statistics and variants were only called for this intersecting region, where 196,749 variants were called.

In summary, there was an average mean read depth of 40X, with 83% of the target regions covered at greater than 15X. Data cleaning and filtering was performed using PLINK v1.90b3g, which 1) removed individuals with more than 5% missing genotypes, where 4 individuals were removed 2) removed any variants with less than 99% genotyping rate, where 1,858 variants were removed 3) excluded markers that fail the Hardy-Weinberg equilibrium test using a threshold of $1.0x10^{-6}$, where 787 variants were removed. After removing individuals that failed the PCA analysis (see below), the

final dataset contained 194,013 variants across 572 samples (147 non-syndromic cases and 425 controls), providing a total genotyping rate of 0.99987.

Principal component analyses of the case-control WES analysis

To investigate any potential batch effects due to merging the case and control datasets, several principal component analyses (PCA) were performed on alignment, variant, and population level data.

For the alignment level data the PCA analysis was based on metrics from the final BAM files including: mean coverage, percent of bases above 15X, and library size. These metrics were plotted in a scatter plot and separated based on case or control groups. Using the prcomp package in R, a PCA analysis was performed on these three metrics and separation between the case and control groups was investigated.

For the variant level data the PCA analysis was based on metrics from the final VCF file including: average depth, number of singletons, ratio of heterozygotes to homozygotes, and the transition/transversion ratio. These metrics were plotted in a scatter plot and separated based on case or control groups. Using the prcomp package in R, a PCA analysis was performed on these three metrics and separation between the case and control groups was investigated.

For the population level PCA, data were prepared for PCA analysis by taking common SNPs (MAF >5%) and pruning out SNPs in linkage disequilibrium with an r2 > 0.2, stepping along five SNPs at a time within 50kb windows. PCA analysis of these variants were projected onto the 1000 genomes data, using the Peddy package, was used to detect population outliers that could confound the analysis (Supplementary Figure 3a & 3b).

SKAT-O of the case-control WES

A WES case-control analysis was used on a non-syndromic AVSD case cohort compared to a control group, in which we took a gene-burden approach to identify genes that might be implicated in AVSD (Figure 1A). To this end, the final combined variant call file set was composed of high stringency and high confidence variants. These variants were grouped into their respective genes, hg19_refGene, and were analyzed using the sequenced based kernel approach. In summary, 173,611 total SNPs were grouped into 18,074 total genes. The number of total SNPs is greater than the number of total called SNPs because variants were allowed to be in more than one gene since the test compares gene burden in the same gene not between different genes. The first three principal components were added as covariates to adjust for any underlying population structure. SKAT-O was run to test for the association with AVSD as a dichotomous variable[25]. A quantile-quantile plot was generated to investigate any departure from the observed and expected p-values (Supplementary Figure 4). Genes reaching the exome-wide significance threshold of 2.77×10^{-6} were considered genes of interest and investigated further.

Integrated gene signaling processing of the case-control WES

To prioritize risk genes from the SKAT-O analysis, p-values corresponding to each gene from SKAT-O were used as the input variable for the integrated gene signaling processing (IGSP) analysis[26]. Method parameter of network plus phenotype, full integration, and 2 percent risk gene percentage was used. Each gene was assigned an average network score, an average phenotype score, and a finial IGSP score. The top 25 genes from the IGSP analysis, based on the final score, were annotated and investigated for enrichment with biological process gene ontology (GO) terms using STRING, in which p-values were false discovery rate adjusted[27].

http://zdzlab.einstein.yu.edu/1/igsp.html

Validation and annotation of variants and genes of interest

Due to the lack of addition DNA from these samples, Sanger sequencing to validate variants was not possible. Instead, a careful examination of variants in the BAM files using the integrative genome viewer (IGV) was employed (Supplemental Figure 5). Combined Annotation Dependent Depletion (CADD) scores, PHRED-scaled, were used as a tool for scoring the deleteriousness of variants of interest[28]. Further, we annotated on the gene level with the number of variants observed, if they were driven by cases, controls, or both, and if they had known cardiovascular phenotype (Supplemental Table 1). As a validation of our control cohort variant calls, we annotated each variant with its allele frequency in the ExAC database, European non-Finish population[29]. A flow chart of this approach is summarized in Figure 1A.

Deleterious rare variant analysis of non-syndromic AVSD

Additionally, a deleterious rare variant analysis was used for the non-syndromic AVSD case cohort to identify potential rare variants of large effect size, that maybe missed or underpowered by the traditional case-control approach (Figure 1B).

Variants were re-called using the BAM files from the nonsyndromic AVSD case cohort where 151 samples passed QC metrics. The variants in this VCF file are overlapping and also unique in comparison to the combined case-control VCF calls by nature of the algorithm used to call variants. For example, if a particular region had good coverage/quality in the case dataset, but not in the control dataset; then that variant would not appear in the combined case-control cohort VCF, but would be in just the case cohort VCF file. Additionally, variants from both datasets were thrown out if they were not in the regions intersecting the two libraries. For this analysis, no samples were removed due to population ancestry prediction. By re-calling variants on just the case cohort we are able to get a higher resolution and complete set of variants. These variants were then filtered using highly stringent parameters leading to a set of high confidence rare variants to be analyzed. Briefly,

variants were filtered out of the dataset according to these parameters: variants with MAF > 1% in any population, variants not predicted to impact transcription or protein amino acid content, or to be annotated as a protein interaction amino acid, variants with quality score < 500, depth < 20 if not a known dbSNP site or depth < 10 if known dbSNP site , paralogous sequence (^MUC,/-/, ^OR), genotype quality score < 90 for novel variant, and a scaled CADD score > 20. The variants and their respective genes were further annotated and prioritized. A flow chart of this approach is summarized in Figure 1B.

Results:

Principal component analyses and quality control

Population level PCA analysis revealed that 33 cases and 20 controls were not predicted to be of European ancestry and were removed from the data set for the case-control study. Further, we plotted cases and controls according eigenvector 1 and eigenvector 2 and saw no separation which confirmed no addition bias in population structure (Supplementary Figure 3c).

For our final dataset contained 161,690 variants across 520 samples (115 non-syndromic cases and 405 controls) providing a total genotyping rate of 0.99997 and had a genotype concordance with dbSNP of 99.02%, which passed our threshold of 95%. Variant level quality control metrics for whole exome sequencing were used to confirm the alignment and joint variant calling pipeline.

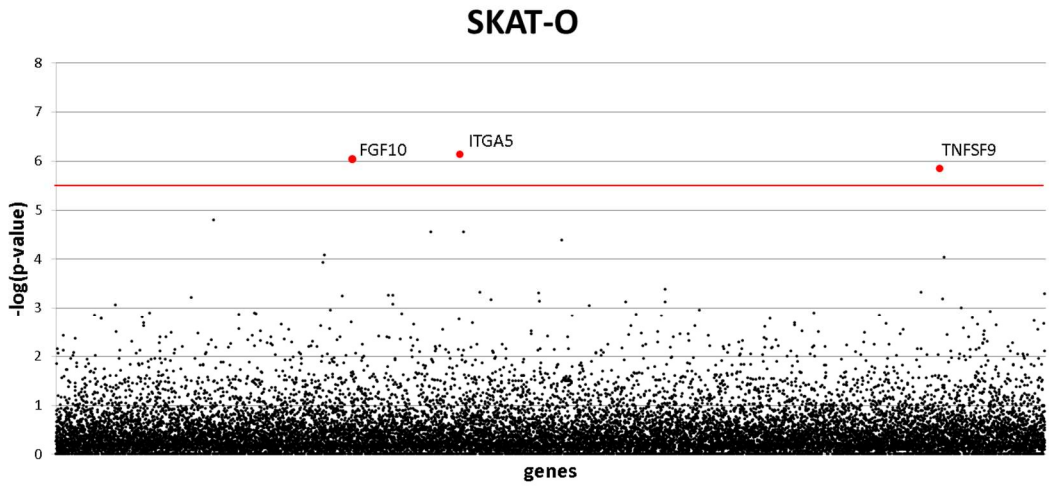
Combining sequencing data from multiple studies is becoming a common theme in large genotype-phenotype studies. Because of the availability of high quality data and well characterized cohorts through databases like dbGAP, this is now more feasible, accurate, and precise. Despite this, there can be other biases and batch effects to investigate and mitigate. Therefore, PCA analyses were performed at the alignment and variant level to look for any separation in our case and control datasets. For the alignment level metrics: our case dataset the average mean coverage of 53X and

our control dataset was at 70X, the average percent bases above 15X for the case dataset was at 85.0% and our control dataset was at 81.7%, and the average library size for our case dataset was $1.86x10^8$ and our control dataset was at $1.50x10^8$. A PCA analysis of these three alignment level metrics showed a clear separation between our case and control dataset (Supplementary Figure 1).

Because of this result, we developed a highly stringent and high confidence pipeline for calling variants to mitigate these observed batch effects. After applying this, we performed the same analysis, but on metrics from the variant call format file. For these metrics: our case dataset had an average depth of 36X and our control dataset was at 43X, the average number of singletons in our case dataset was 140 and our control dataset was at 150, the average ratio of heterozygotes to homozygotes for our case dataset was 1.75 and our control dataset was 1.77, and the average ratio of transitions to transversions was 3.09 for our cases dataset and 3.08 in our control dataset. A PCA analysis of these four variant calling level metrics showed little to no separation between our case and control dataset (Supplementary Figure 2).

SKAT-O reveals ITGA5, FGF10, and TNFSF9 as genes associated with non-syndromic AVSD

Utilizing SKAT-O, a case-control association test was performed to investigate on the gene-level by aggregating variants into their respective genes. This is a powerful analysis that can overcome the lack of power in traditional single variant association test. Our cohort composed of 115 non-syndromic AVSD cases and 405 controls with no heart defects revealed three genes above the exome-wide significance threshold which was Bonferroni adjusted for testing 18,074 genes: *ITGA5*, *FGF10*, and *TNFSF9* with p-values of $7.18x10^{-7}$, $8.99x10^{-7}$, and $1.41x10^{-6}$ respectively (Figure 2).



gene	p-value	-log(p-value)	variants in gene
ITGA5	7.18E-07	6.14	19
FGF10	8.99E-07	6.05	6
TNFSF9	1.41E-06	5.85	2

Figure 2: Optimal Sequence Kernel Association Test (SKAT-O) results shows that three genes are associated with AVSD. Scatter plot showing the exome-wide significant finding that variants in *FGF10*, *ITGA5,* and *TNFSF9* are associated with AVSD. The horizontal line is the threshold for exome-wide significance (based on testing 18,074 genes, the exome-wide significance p-value=2.77×10^{-6}). The table shows the SKAT-O results and the corresponding number of variants identified in each significant gene.

GENES OF INTEREST: SKAT-O				ANNOTATED VARIANTS					
GENE	**# of case driven variants**	**overall driver**	**MGI heart phenotype**	**SNP**	**ref/ alt**	**ExAC**	**maf_a**	**maf_u**	**CADD**
ITGA5	7	both	yes	rs143445624	G/A	0.03%	0.43%	0.00%	5.927 (G993=)
				rs138831238	G/A	0.08%	0.87%	0.00%	9.613 (H621=)
				rs113261097	G/A	1.04%	4.35%	0.25%	7.054 (splice)
				rs112878226	T/G	1.03%	4.35%	0.25%	5.459 (potential splice)
				rs2230393	G/A	1.04%	4.35%	0.25%	12.25 (G217=, splice)
				rs12314853	A/G	1.04%	4.35%	0.25%	7.863 (intron)
				rs12301470	C/C	1.04%	4.35%	0.25%	5.972 (splice)
FGF10	3	both	yes	5:44310580	A/G	na	3.91%	0.00%	10.42 (I126)
				5:44388566	A/G	na	1.74%	0.00%	1.425 (L73)
				5:44388734	G/C	na	2.17%	0.00%	8.641 (P17)
TNFSF9	2	both	no	19:6531148	GG CT/ G	0.51%	3.91%	0.00%	4.202 (inframe L41del)

Table 1: Annotated variants in the genes of interest from the SKAT-O analysis. The name of the gene along with: how many case driven genes, if it was found in just cases/controls/both which drove the association p-value, and if that gene is known to be associated with congenital heart defects in mice. Each variant is

further annotated with its rsID from dbSNP (if available), the reference and alternate allele, the ExAC allele frequency for the non-Finish European population, the allele frequency in our AVSD affected case cohort (maf_a) and the unaffected control cohort (maf_u), and its corresponding CADD score and consequence.

All of the genes from the SKAT-O analysis and their respective variants are annotated in Supplemental Table 4. Traditional gene burden test are powerful when all the variants in a gene are causal, but often have limitations when the gene has many non-causal variants or when causal variants have different directions of association (both protective and deleterious). This is actually one of the strengths of the SKAT-O algorithm in that it can detect both scenarios[30]. Because of this complex combination of variants that goes into the actual SKAT-O p-value calculation, we focus on the case driven variants within the genes identified to reach exome wide significance (Table 1).

Integrin subunit alpha 5*, ITGA5*, is a part of the integrin family that associates with its beta 1 subunit that is a receptor for fibrillin-1 and mediates cell adhesion[31]. Fibrillin-1 is an extracellular matrix protein that has been shown to be essential for cardiovascular development. In our study, we identified 7 variants in *ITGA5* which were present in our non-syndromic AVSD cases. Specifically, two of the variants rs143445624 and rs138831238, although synonymous, are extremely rare and only observed in cases. The other five variants are in a haplotype block we will term *ITGA5* +5. The variants in *ITGA5* +5 have a 1% minor allele frequency in the ExAC European non-Finnish population, but have a minor allele frequency of 4.35% in our non-syndromic AVSD case cohort. Interestingly, this block has also been associated with biliary tract disease[32]. Four of the five variants are predicted to impact splice sites, but rs2230393 is likely the causal variant in the block because of its high CADD score and actual proximity to the splice acceptor site in exon 6.

Fibroblast growth factor 10, *FGF10*, is a paracrine growth factor involved in several biological process including embryonic development, cell proliferation, and cell differentiation[33,34]. It is also a key signaling component of the second heart field as a downstream target of the Shh pathway[35]. There were three variants identified in *FGF10* that were only observed in our non-syndromic AVSD

cases cohort and not observed at all in the ExAC database, speaking to their pathogenic potential. Via the transcription factor ChIP-seq track in the UCSC genome browser[36,37], we identified that variants 5:44388566 and 5:44388734 are in a region occupied by the EZH2 transcription factor, in multiple cell types (Figure 3). EZH2, enhancer of zeste 2 polycomb repressive complex 2 subunit, plays a critical role in healthy embryonic development through the regulation of epigenetic maintenance of genes. It is specifically involved in early heart development and ablation of EZH2 in mice results in multiple CHDs, including atrial septal defects, atrioventricular septal defects, and ventricular septal defects[38]. Also in this region, an additional variant in *FGF10*, P47Q with a CADD score of 23, was identified in 10 of our cases.

Figure 3: UCSC genome browser of *FGF10* gene region 44,388,566 to 44,388,734 of genome hg19, showing the EZH2 ChIP-seq peak. http://genome.ucsc.edu

Tumor necrosis factor ligand superfamily member 9, *TNFSF9*, is a transmembrane cytokine that acts as a ligand to activate T-cells, may play a role in activation-induced cell death, and be involved in interactions between T-cells and macrophages[39]. Only 1 variant were found in *TNFSF9* of which was non-syndromic AVSD case specific and there is no evidence for this gene in heart development.

The top 25 IGSP prioritized risk genes are enriched for cardiovascular development

To further increase the studies power in detecting potential genes involved in the pathogenesis of AVSD, we also integrated known network and phenotype information to uncover risk genes that may have only had a marginal association signal. To this end, we employed the newly developed IGSP algorithm[26]. Using IGSP, we prioritized an additional 25 risk genes, three of which were previously identified in our SKAT-O analysis (Table S1). These top 25 risk genes were significantly enriched for anatomical structure development, vascular development, and cardiovascular system development, etc. (Figure 4). These enrichment results are noteworthy because IGSP is an unbiased approach

algorithm that had no prior knowledge of what phenotype was being studied. Therefore, these enrichments show that IGSP prioritized genes are of high interest to the potential pathogenesis of AVSD. IGSP identified several risk genes - *ZFPM2* (Zinc Finger Protein, Multitype 2), *MAP2K1* (Mitogen-Activated Protein Kinase Kinase 1), *ITGAV* (Integrin Subunit Alpha V), *CXCL12* (C-X-C Motif Chemokine Ligand 12), *JAM2* (Junctional Adhesion Molecule 2), and *GATA6* (GATA-binding factor 6) - that are known to play a critical role for CHDs in humans or in heart development in animal models[8,40-44]. Further investigation into the genotypes of cases and controls in these genes found that they are composed of a complex mixture of case specific variants, control specific variants, and those that occur in both. This primarily speaks to the methodology used to compute p-values by the burden test SKAT-O which accounts for underlying genetic mechanisms that genes can possess casual variants in different directions and of different magnitudes[30].

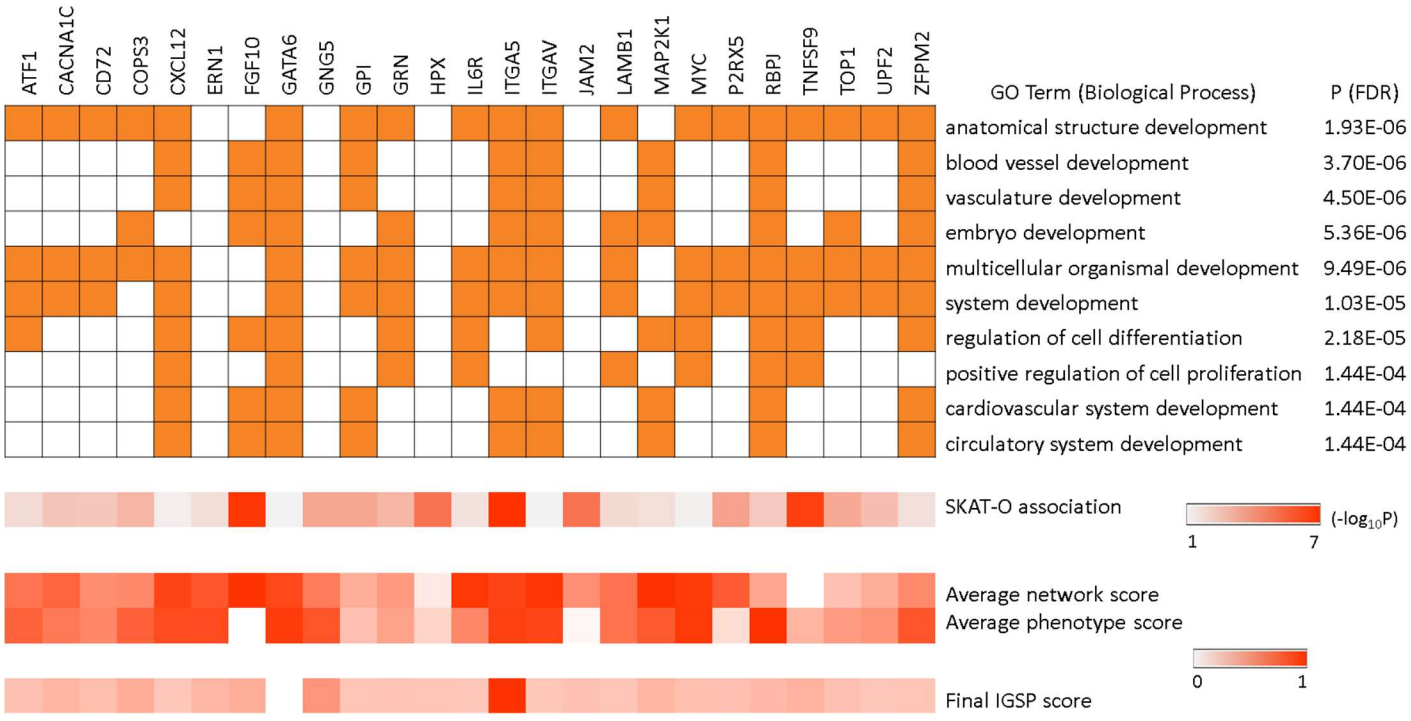


Figure 4: Integrated gene signaling processing analysis reveals risk factors involved in AVSD. The top 25 genes from the IGSP analysis are enriched with GO terms related to early development and specifically cardiovascular development. The genes listed on top are in highlighted in orange if they are included in the GO term listed to the right with their respective enrichment P values, false discover rate adjusted. Below are heat maps for the original SKAT-O association P values, average network score and phenotype score, and the final IGSP score.

Results of the deleterious rare variant analysis of the non-syndromic AVSD cohort

The 151 non-syndromic AVSD samples were subjected to a stringent WES pipeline in which a total of 200,470 variants were called. After the rare variant filters and those with a CADD > 20 were applied, a total of 6,871 variants in 4,955 genes remained with an average of 46 variants per sample. To further prioritize the genes with rare variants, they were clustered into groups including: those with a Mouse Genome Informatics (MGI) cardiovascular phenotype, those with a CADD scaled score > 35 representing the top 0.05% of all deleterious variants in the genome, those involved in the primary cilia, and those involved in the Shh signaling pathway (Figure 5A, Supplemental Table 2). Additionally, the genes from the case-control SKAT-O analysis were re-analyzed to look for additional variants.

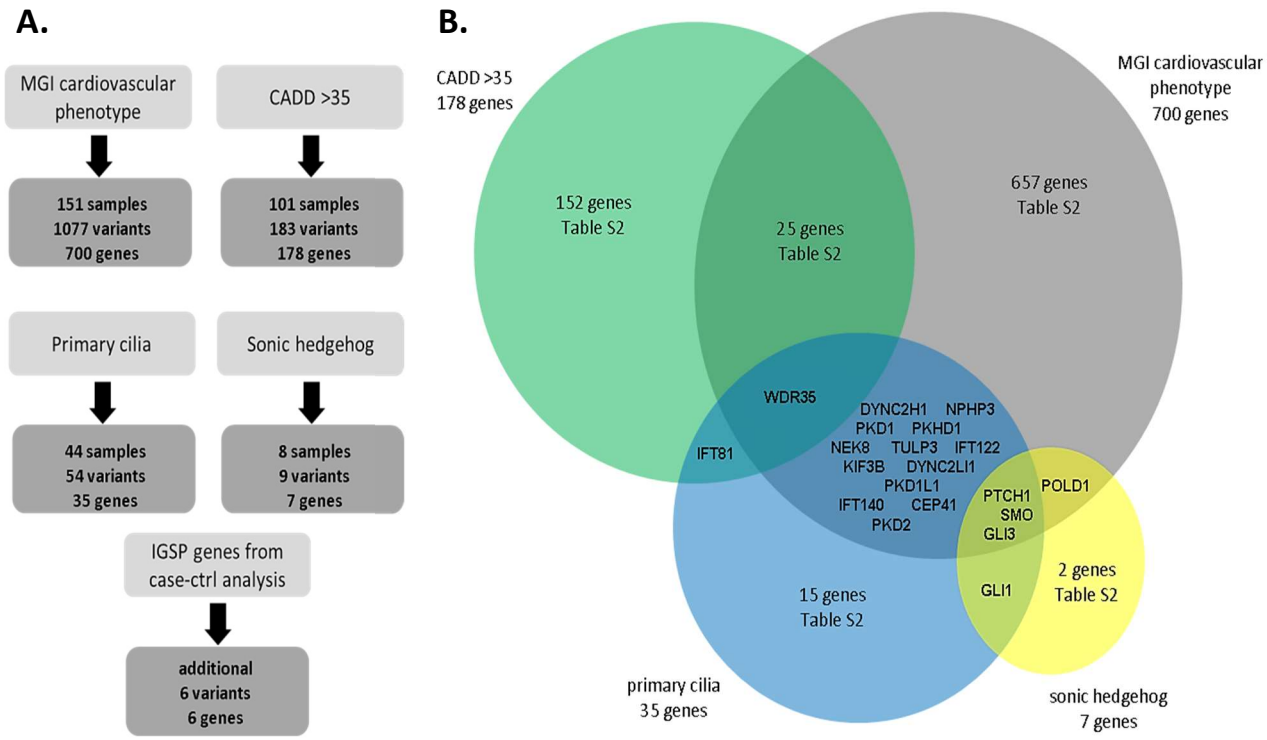


Figure 5: Deleterious rare variants in genes involved in sonic hedgehog signaling and the primary cilia were found in non-syndromic AVSD cases. A. Groups in which the genes and their variants were clustered into. Listed below the groups are the number of samples, the number of variants, and the number of genes in each. **B.** A Venn diagram of each group and the genes that overlap.

Overlapping clusters containing deleterious rare variants (Figure 5B)

The MGI database is a robust and carefully curated database that has experimental data on knockout or mutated genes and their associated phenotypes in mice[45]. This was the broadest group in which all 151 samples had one or multiple variants in a total of 700 genes. Of these 700 genes with a cardiovascular phenotype, 31 genes specifically were associated with an atrioventricular septal defect (Supplemental Table 3).

CADD scores are generated using over 60 metrics and annotations. These scores are widely used to assess and rank the deleteriousness of variants of interest[28]. We applied this to all of our rare variants and chose a very high cutoff of 35 to prioritize the most deleterious variants in our dataset. In total, 178 genes had 183 variants with a scaled CADD score > 35 (Supplemental Table 2).

Extensive studies in mouse models, molecular biology, and biochemistry have been utilized to better understand the etiology of AVSDs. From this work we know that Shh signaling through the primary cilia is critical for proper septation. Therefore, we annotated all the genes from the rare variant analysis using the Gene Ontology cellular component classifier[27] for the primary cilia (35 genes with 54 variants) and the ConsensusPathDB[46,47] database for Shh signaling (7 genes with 9 variants).

Genes and their variants that are in multiple overlapping groups are of high interest to this study and will be prioritized for follow up studies (Table 2). Specifically of interest are variants in *GLI3* (GLI Family Zinc Finger 3), *GLI1* (GLI Family Zinc Finger 1) *PTCH1* (Patched 1), and *SMO* (Smoothened), which had predicted deleterious rare variants occurring in 6 non-syndromic AVSD cases. These genes are central to proper Shh signaling through the primary cilia and mouse models have cardiovascular defects. Additionally, the *WDR35* (WD Repeat Domain 35) stop gained variant, L641*, has a high CADD score of 42 and is classified as pathogenic in the ClinVar database.

	I D	Gene	chr:pos	Functional Prediction
Sonic hedgehog & Primary cilia & MGI cardiovascular phenotype	S005	GLI3	chr7:42005552	E1040V
	S035	GLI3	chr7:42063163	K467N
	S053	GLI3	chr7:42007425	D734N
	S030	PTCH1	chr9:98238438	R536G
	S121	SMO	chr7:128851979	P684L
Primary cilia & MGI cardiovascular phenotype & CADD >35	S141	WDR35	chr2:20141557	L641*
Sonic hedgehog & Primary cilia	S107	GLI1	chr12:57861891	R398C
Sonic hedgehog & MGI cardiovascular phenotype	S063	POLD1	chr19:50905989	G321S
Primary cilia & CADD >35	S090	IFT81	chr12:110630488	R512*
Primary cilia & MGI cardiovascular phenotype	S132	CEP41	chr7:130039945	T303I
	S016	DYNC2H1	chr11:103124170	R3407Q
	S029	DYNC2H1	chr11:103124170	R3407Q
	S075	DYNC2H1	chr11:103091424	R3007C
	S101	DYNC2H1	chr11:103086453	G2900S
	S089	DYNC2LI1	chr2:44014370	R71K
	S091	DYNC2LI1	chr2:44016779	D80G
	S046	IFT122	chr3:129202416	S632N
	S130	IFT140	chr16:1642474	P162L
	S074	IFT46	chr11:118416555	T280M
	S105	IFT46	chr11:118422535	E264A
	S084	KIF3B	chr20:30915459	Y655H
	S035	NEK8	chr17:27068575	L679P
	S022	NPHP3	chr3:132433942	D315V
	S143	NPHP3	chr3:132402290	S1217R
	S032	PKD1	chr16:2140002	S4213F
	S119	PKD1L1	chr7:47924213	S1083L
	S130	PKD1L1	chr7:47913531	R1288C
	S150	PKD2	chr4:88959654	Splicing Acceptor/Donor Loss

	S033	PKHD1	chr6:51637567	L2859V
	S139	PKHD1	chr6:51923171	R488W
	S042	TULP3	chr12:3042666	R260P
	S081	TULP3	chr12:3048591	Splicing Acceptor/Donor Loss

Table 2: Annotated variants in the genes of interest from the rare variant analysis. Samples in gray have more than one variant.

Samples with more than one variant of interest

There were five samples that had compound variants of interest:

S033 has one variant in the primary cilia and MGI cardiovascular phenotype overlapping cluster (PKHD1) and also has three variants in FGF10 variants identified in through SKATO and the FGF10 variant identified in the rare variant analysis.

S053 one in the Sonic hedgehog, primary cilia, and MGI cardiovascular phenotype overlapping cluster (GLI3) and also has the FGF10 variant identified in the rare variant analysis.

S105 one variant in the primary cilia and MGI cardiovascular phenotype overlapping cluster (IFT46) and also has the five linked variants in ITGA5 identified in SKATO.

S130 has two variants in the primary cilia and MGI cardiovascular phenotype overlapping cluster (PKD1L1 and IFT140).

S035 has one variant in the primary cilia and MGI cardiovascular phenotype overlapping cluster (NEK8) and one in the Sonic hedgehog, primary cilia, and MGI cardiovascular phenotype overlapping cluster (GLI3).

Discussion:

AVSD is a clinically significant congenital heart defect that has lifelong implications. It is most often seen in Down syndrome, but also occurs as a simplex trait, where it is a sporadic defect without any discernable extracardiac anomalies. Although advances have been made in identifying contributing genes, it is clearly an oliogenic defect with only a small fraction of genes identified to date.

Therefore, we designed a comprehensive unbiased study of the genetics of those with non-syndromic AVSD. Utilizing the exomes of a large case-control cohort, we took two approaches to identify and prioritize genes of interest. In this study, we found several genes critical to the primary cilia and cardiovascular development were significantly associated with non-syndromic AVSD in both case-control and rare variant analyses. Additionally, novel genes not previously known to have a role in AVSD were also discovered, leading to new potential mechanisms.

Evidence for the fundamental role of the primary cilia in heart development has been shown through several mouse studies[13,23,24,34,48,49]. Although these studies have focused on different components or mechanisms of disease, the common theme which arose is that congenital heart defects occur when the primary cilium is disrupted. Specifically, we know that Shh signaling, which is a highly conserved pathway that occurs exclusively in the primary cilia, is the critical signaling event. Briefly, when the Shh ligand binds to its receptor Ptch1, it exits the cilia and releases its repression on Smo. Subsequently, the Shh transcription factors Gli2 and Gli3, now in their active form, are shuttled to the nucleus where they act as transcription effectors of Shh by turning on target genes Gli1 and Ptch1, among others[50-52]. Shh signaling is dependent on the integrity and proper functioning of the primary cilia, such as trafficking, microtubule structure, and membrane components. Hence we call this expansive network, the ciliome.

It is well established that the network of genes that make up the ciliome are critical for proper development of the atrioventricular septum[53]. Therefore, these ciliome components are the principal candidate genes, that when mutated, could result in an AVSD. We combined the knowledge of the ciliome's role in pathogenesis of AVSD with the MGI phenotype database and CADD scores to prioritize genes of interest. Our study has identified several ciliome genes that have predicted deleterious rare variants in our non-syndromic AVSD cases. Notably, there were predicted rare deleterious variants in *GLI3* in 3 cases and 1 case with a *GLI1* variant. Mutations in this family of

transcription factors have been implicated in severe congenital malformations and ciliopathies[54,55] Genes that directly interact with the Gli family were also identified in our AVSD cases in which non-synonymous variants in *PATCH1* and *SMO* were in one case each. Although it remains to be confirmed, we predict that these rare pathogenic variants in these 6 cases act through the same mechanism by directly disrupting Shh signaling.

Bi-directional, anterograde, and retrograde motors that traffic proteins along microtubules to and from the primary cilia are responsible for cilia formation, maintenance, and signaling. We identified deleterious rare variants in *DYNC2H1* and *IFT122,* which are IFT-A retrograde components involved in Gli1 trafficking[56]. Mutations in these genes have been shown to cause Gli1 misslocalization and are associated with abnormal Shh signaling[57-59]. Rare variants in *IFT140* are also significant with regards to Gli1 because expression was decreased in an *IFT140* deletion study[60]. Other transport proteins *IFT81*, *IFT146*, and *KIF3B* are in the IFT-B complex responsible for anterograde trafficking and *WDR35* and *DYNC2LI1* are in the IFT-A complex responsible for retrograde trafficking in the primary cilia[61]. We hypothesize that these rare deleterious variants in these 13 cases indirectly disrupt Shh signaling and/or other vital functions of the cilia through trafficking.

Utilizing an unbiased gene-based scan of the genome, SKAT-O, we found that *FGF10* reached exome wide significance. Progenitor cells in the second heart field that express Fgf10 give rise to the right ventricle and the outflow track of the human heart though Wnt/β-catenin and Notch signaling[14,62,63]. While mouse studies have shown that a knockout of Fgf10's receptor Fgfr2-IIIb leads to ventricular septal defects, Fgf10 knockout does not, but did have an abnormal right ventricle[64].

Another gene identified in the SKAT-O analysis, *ITGA5*, may speak to different mechanism leading to an AVSD, outside of effecting Shh signaling. Itga5 is a membrane protein that forms a complex with its beta counterpart to form a receptor for fibronectin, which is an important part of the

extracellular matrix (ECM)[65,66]. The ECM's role in heart development is complex and extensive in that it provides communication and interaction between cells[67]. We hypothesize that Itga5 affects cell adhesion via fibronectin which is required for the migration and proliferation of the second heart field cells.

To gain a broader understanding of our case-control results from the SKAT-O analysis, we ran a secondary analysis in which known network and phenotype information was integrated, using IGSP. The top 25 risk genes ranked from this analysis were overrepresented in anatomical structure development, vascular development, and cardiovascular system development, among others, speaking to the power of our study in that the algorithm is blinded to the phenotype being studied. Of particular interest is the junctional adhesion molecule, *JAM2*, which is on chromosome 21 and has been implicated as an effector gene in Down syndrome associated-AVSD[40]; *GATA6* and *ZFPM2*, which are a known cardiac transcription factors that when mutated cause an AVSD[8]. This analysis also gives us insight into several genes known to play a role in embryonic development, but not specifically cardiovascular development, such as transcription factors: *ATF1* (Activating Transcription Factor 1) and *MYC* (MYC Proto-Oncogene); and signaling regulators: *COPS3* (COP9 Signalosome Subunit 3), *ERN1* (Endoplasmic Reticulum To Nucleus Signaling 1), *GNG5* (G Protein Subunit Gamma 5), *P2RX5* (Purinergic Receptor P2X 5); and ECM component: *LAMB1* (Laminin Subunit Beta 1). The complex architecture of variants identified in genes in our case-control study need to be investigated further in order to fully understand their effects on the gene product and ultimately on heart development.

As with most sequencing studies, the power to detect variants and risk genes is limited by the samples which were available to us, therefore we utilized approaches that were designed to mitigate and add power. Future studies should encompass a higher number of cases and controls to uncover additionally variants and genes. The rare nature of the AVSD phenotype and the heterogeneity of

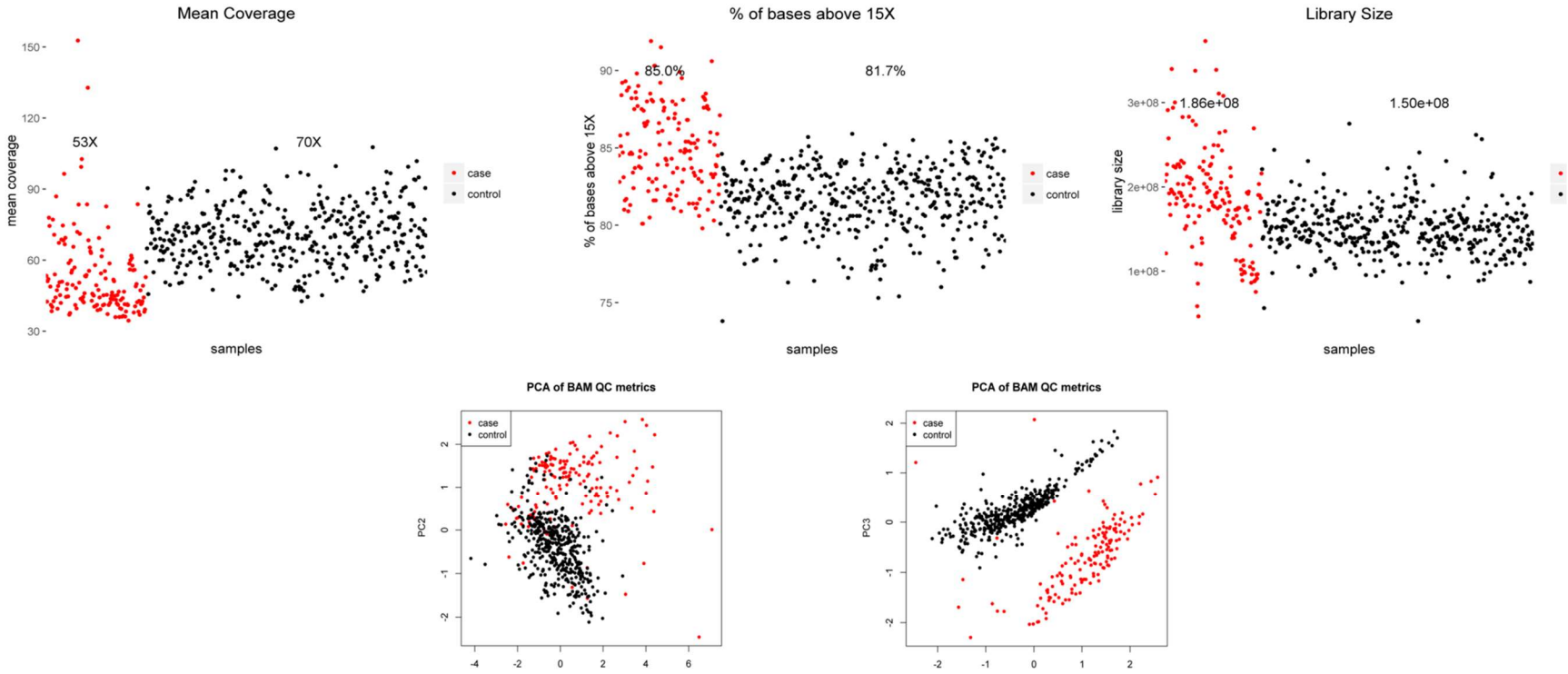
the genetic mechanisms have led to uncertainty and inconsistency in the expected effect size of causal variants. Therefore, we did not take this into account in our present study, although the SKAT-O algorithm does weigh variants based on minor allele frequency. Because of the known overrepresentation of rare variants in complex disease, we were particularly interested in investigating their presence in our case cohort.

Our large sequencing study of non-syndromic AVSD used two approaches to discover risk genes implicated in AVSD. Most of the genes we discovered correlate with previous studies and literature that demonstrate the critical nature of the second heart field and Shh signaling in the proper formation of the atrioventricular septum. We hypothesize that variants in these gene directly or indirectly cause aberrant Shh signaling. Additionally, we have identified genes that are involved in novel signaling pathways, trafficking, cell adhesion, and transcription factor networks that have not yet been elucidated in cardiovascular development.

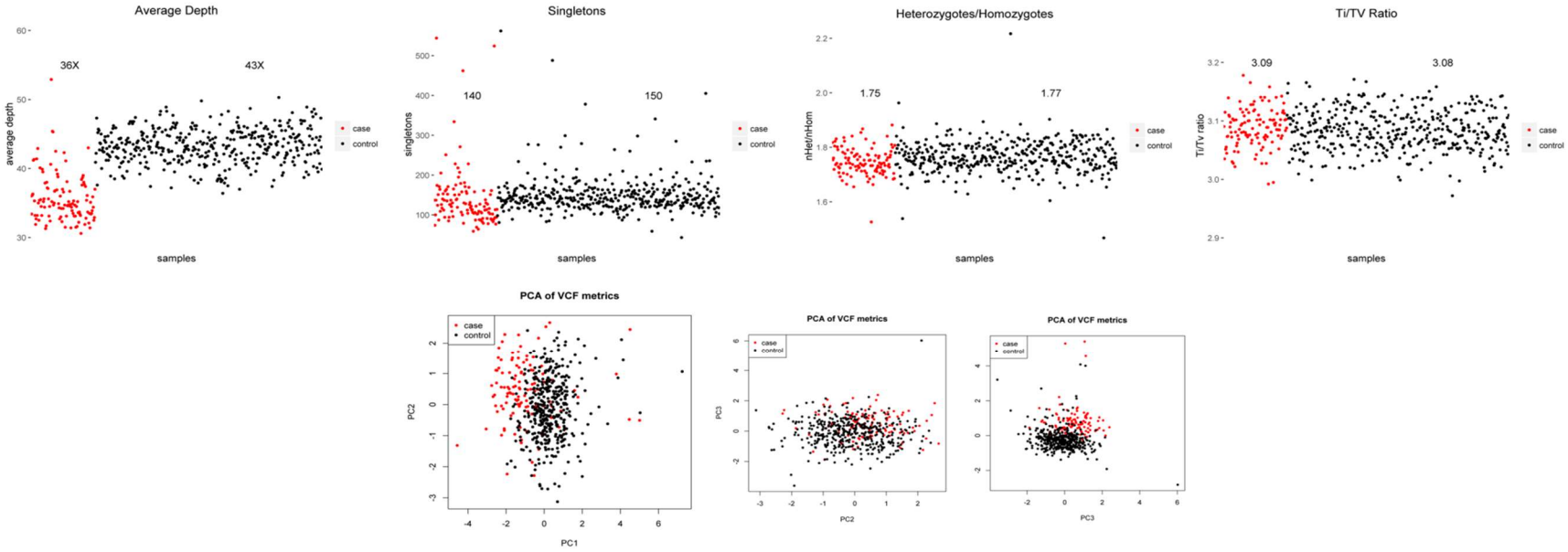
With multiple lines of evidence supporting the heterogenetic nature of AVSD, our goal was to add to the growing list of risk genes throughout the genome. With this study we pave the way for systems biology and follow up studies which will be facilitated by our risk gene discoveries.

Supplemental figures and tables

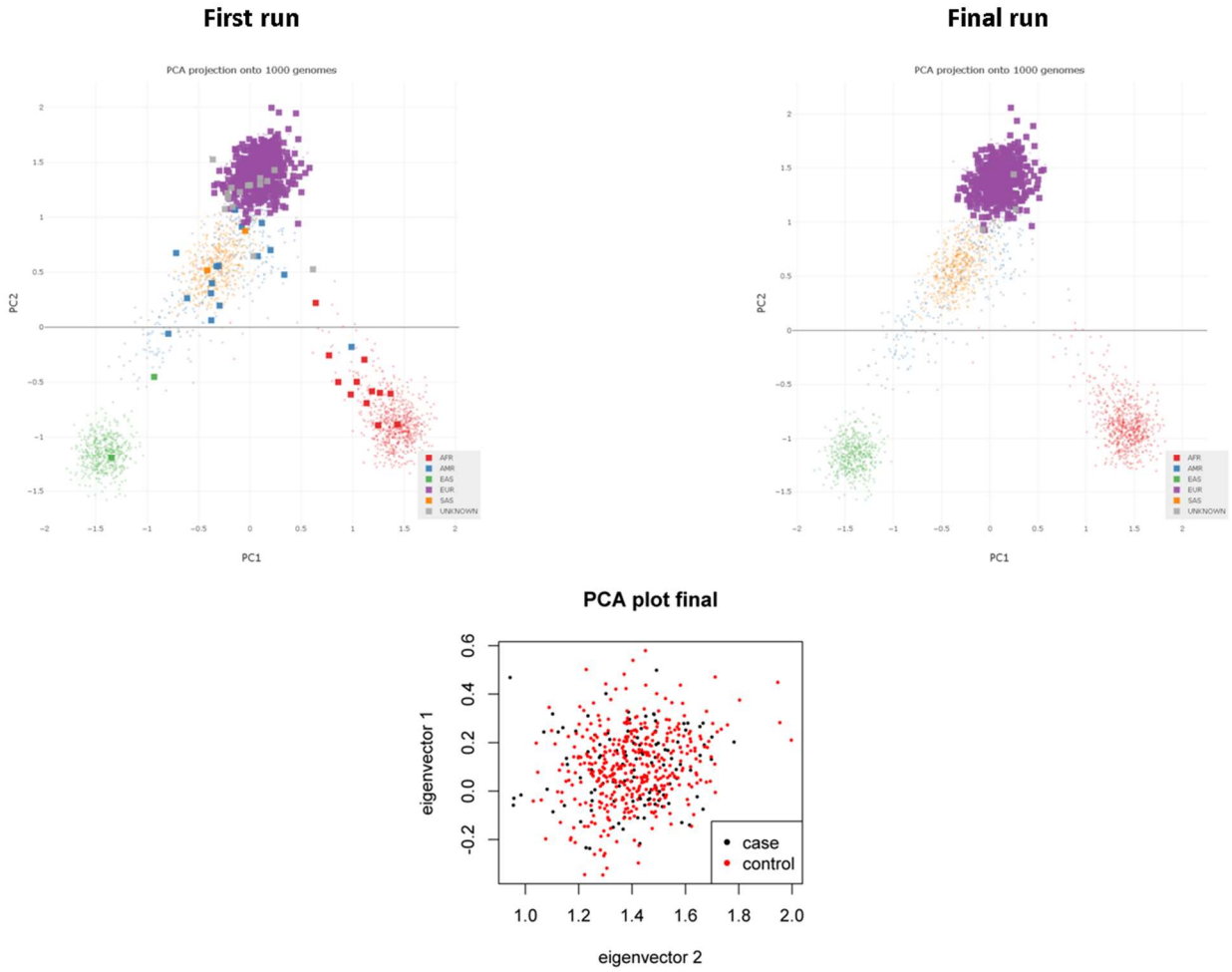
Supplementary Figure 1: alignment level PCA



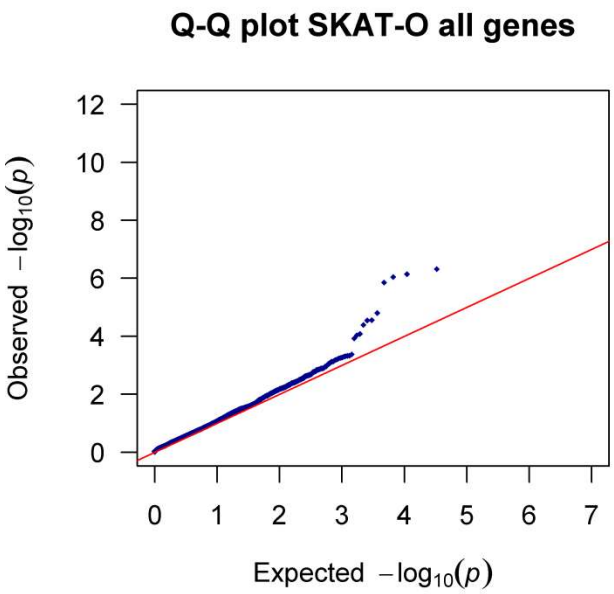
Supplementary Figure 2: variant level PCA



Supplementary Figure 3: population level PCA



Supplementary Figure 4: Q-Q plot of SKAT-O analysis



Supplementary Figure 5: example IGV captures of variants of interest

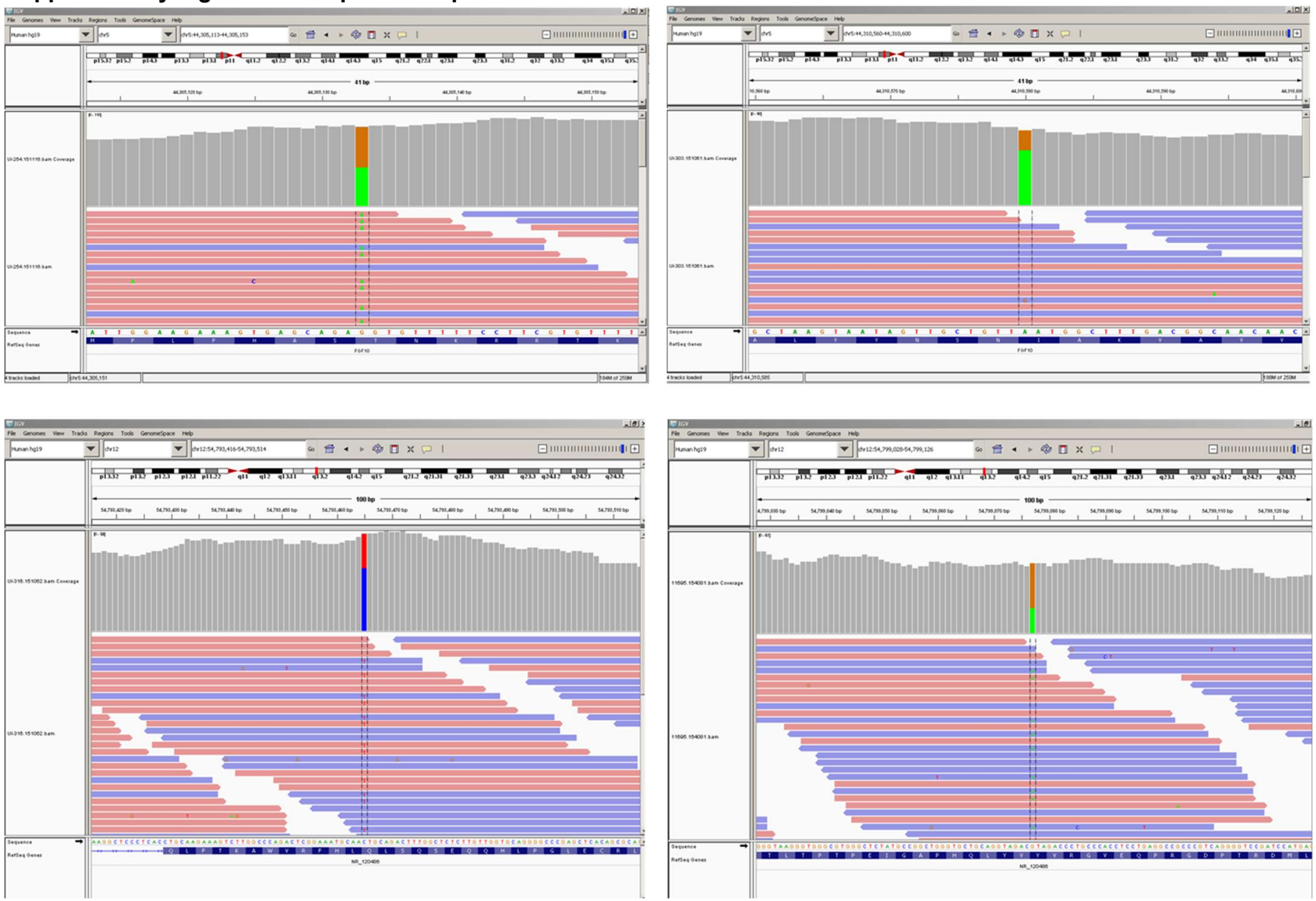


Table S1: Annotated variants and genes of interest from the IGSP analysis.

GENES FROM IGSP ANALYSIS								ANNOTATED VARIANTS				
GENE	#	DRIVER	MGI or HUMAN HEART PHENO	FINAL SCORE	ASSOC SCORE	AVG NETWORK SCORE	AVG PHENOTYPE SCORE	SNP	REF/ALT	ExAC	MAF_A	MAF_U
CXCL12	1	case	yes	0.21	1.51	0.90	0.87	10:44876173	A/G	0.01%	0.43%	0.00%
ITGA5	19	both	yes	0.77	6.14	0.90	0.92	rs143445624	G/A	0.03%	0.43%	0.00%
								rs138831238	G/A	0.08%	0.87%	0.00%
								rs113261097	G/A	1.04%	4.35%	0.25%
								rs112878226	T/G	1.03%	4.35%	0.25%
								rs2230393	G/A	1.04%	4.35%	0.25%
								rs12314853	A/G	1.04%	4.35%	0.25%
								rs12301470	C/C	1.04%	4.35%	0.25%
								rs200141012	A/G	0.03%	0.00%	0.12%
								rs200896534	G/A	0.02%	0.00%	0.12%
								rs151324601	G/C	0.08%	0.00%	0.12%
								rs200284742	A/G	0.02%	0.00%	0.12%
								12:54797063	G/A	0.00%	0.00%	0.12%
								12:54797739	C/T	0.00%	0.00%	0.12%
								rs185312486	C/A	0.04%	0.00%	0.12%
								rs201235779	A/C	0.00%	0.00%	0.12%
								rs138614417	A/C	0.05%	0.43%	0.12%
								rs7306692	C/T	25.65%	27.78%	27.76%
								rs12426629	C/G	5.15%	5.22%	4.82%
								rs1922254	T/G	8.99%	11.74%	11.60%
ITGAV	17	both	yes	0.21	1.35	0.96	0.89	2:187455255	G/A	na	0.34%	0.00%
								rs16828136	G/A	0.36%	0.34%	0.00%
								rs191593642	G/A	0.00%	0.34%	0.00%
								2:187520976	TTCCAGG/T	na	0.34%	0.00%
								rs200299043	G/A	0.04%	0.34%	0.00%
								rs3738918	A/G	0.42%	1.69%	0.00%
								rs201118947	T/C	0.00%	0.00%	0.12%
								rs145167954	C/T	0.04%	0.00%	0.12%
								rs147403786	G/C	0.08%	0.00%	0.24%
								rs147413915	C/T	0.09%	0.00%	0.12%
								2:187533446	ATT/A	0.11%	0.00%	0.59%
								rs200924395	G/A	0.01%	0.00%	0.12%
								2:187541549	A/G	0.00%	0.00%	0.12%
								rs9333288	A/G	26.40%	21.62%	25.76%
								rs9333289	T/C	28.51%	27.03%	29.06%
								rs2595391	T/G	10.09%	9.80%	9.65%
								rs56316527	G/A	0.13%	0.34%	0.12%
FGF10	6	both	yes	0.31	6.05	0.97	0.13	5:44310580	A/G	na	3.91%	0.00%
								5:44388566	A/G	na	1.74%	0.00%
								5:44388734	G/C	na	2.17%	0.00%
								rs147715509	T/G	0.33%	0.00%	0.24%
								5:44388406	G/T	na	0.00%	0.12%
								rs2290070	C/G	19.84%	22.17%	20.25%
ZFPM2	16	both	yes	0.22	1.83	0.65	0.82	8:106814120	A/G	0.00%	0.43%	0.00%
								rs28374544	A/G	0.05%	0.43%	0.00%
								rs139368368	C/T	0.00%	0.43%	0.00%
								rs16873741	C/T	0.05%	0.43%	0.00%
								rs200389635	C/T	0.02%	0.00%	0.12%
								rs202204708	A/G	0.08%	0.00%	0.25%
								8:106813515	G/C	0.00%	0.00%	0.12%
								rs35843564	G/A	0.00%	0.00%	0.12%
								rs187043152	G/A	0.50%	0.00%	0.25%
								rs182216711	G/C	0.26%	0.43%	0.12%
								rs121908601	A/G	0.42%	0.87%	0.37%

								rs16873732	T/C	2.36%	4.78%	2.10%
								rs920628	A/G	2.42%	5.22%	2.10%
								rs11995760	C/T	2.33%	5.22%	2.10%
								rs11993776	C/G	9.13%	10.87%	7.65%
								rs3735953	T/C	40.90%	39.13%	40.25%
RBPJ	8	both	yes	0.26	2.38	0.54	0.97	4:26322598	G/A	na	0.87%	0.00%
								rs115117936	G/A	0.30%	0.87%	0.12%
								rs3113014	T/C	0.34%	2.17%	0.37%
								rs185171306	G/A	0.27%	1.30%	0.25%
								rs377744078	G/C	0.03%	0.00%	0.37%
								rs146691754	T/G	0.26%	0.00%	0.49%
								rs151051045	T/C	0.01%	0.00%	0.12%
								rs13116873	G/C	44.07%	46.96%	43.21%
TNFSF9	2	both	no	0.29	5.85	0.21	0.44	19:6531148	GGCT/G	0.51%	3.91%	0.00%
								rs199772971	G/A	0.07%	0.00%	0.12%
COPS3	6	case	yes*	0.31	2.86	0.65	0.78	17:17150407	T/C		0.43%	0.00%
								rs180952638	G/A		0.43%	0.00%
								17:17174227	C/T		0.43%	0.00%
								rs4985761	A/G		48.70%	48.77%
								rs3182911	A/G		47.39%	48.52%
								rs7219012	A/T		11.30%	12.96%
TOP1	1	case	no	0.23	3.19	0.44	0.54	rs61756255	C/T		2.61%	0.37%
GRN	15	both	no	0.22	2.88	0.58	0.52	17:42426459	G/C		0.43%	0.00%
								rs138473783	G/A		0.43%	0.00%
								17:42428736	G/T		0.43%	0.00%
								rs140298583	G/A		0.43%	0.00%
								rs57745105	C/T		0.43%	0.00%
								rs25646	T/C		6.09%	1.98%
								17:42426554	G/A		0.00%	0.12%
								17:42428486	G/C		0.00%	0.12%
								17:42429455	C/G		0.00%	0.12%
								rs142926942	G/A		0.00%	0.12%
								rs63751088	G/A		0.43%	0.12%
								rs63750412	C/T		0.43%	0.25%
								rs72824736	G/A		7.83%	9.75%
								rs9897526	G/A		16.96%	11.48%
								rs850713	G/A		25.22%	23.46%
GPI	11	control	yes	0.22	3.26	0.51	0.39	19:34868485	C/T		0.00%	0.12%
								19:34868792	A/G		0.00%	0.12%
								rs368742453	G/A		0.00%	0.12%
								rs202009325	C/T		0.00%	0.25%
								19:34884272	G/A		0.00%	0.12%
								rs140676743	C/T		0.00%	0.12%
								19:34890849	C/G		0.00%	0.12%
								rs144111841	C/T		0.87%	0.74%
								rs8191371	T/C		4.78%	1.36%
								rs1801015	A/G		5.22%	7.41%
								rs1864139	G/A		8.26%	2.96%
GATA6	6	control	yes	0.21	1.37	0.89	0.93	rs143026087	C/T	0.23%	0.00%	0.74%
								rs199592618	T/C	0.04%	0.00%	0.12%
								rs200556553	A/G	0.13%	0.00%	0.12%
								18:19762939	A/G	0.00%	0.00%	0.12%
								rs146949786	A/G	0.58%	0.00%	1.24%
								rs3764962	A/G	1.42%	0.87%	1.36%
ATF1	6	both	no	0.24	1.95	0.72	0.77	rs117004521	C/T		3.04%	0.49%
								rs138054326	A/C		0.00%	0.12%
								rs2230674	C/G		2.17%	2.22%
								rs17291650	A/G		11.30%	10.00%
								rs4986838	T/C		42.61%	41.98%

								rs1129406	C/T		43.04%	41.98%
								rs371837699	G/A		0.43%	0.00%
								12:2760811	A/G		0.43%	0.00%
								rs199538058	C/A		0.43%	0.00%
								rs112414325	G/A		0.43%	0.00%
								rs1051360	G/T		1.30%	0.00%
								rs55792866	G/A		4.35%	0.99%
								rs200282707	A/G		0.00%	0.49%
								rs200289321	A/G		0.00%	0.12%
								rs201756421	T/C		0.00%	0.25%
								rs200022722	C/G		0.00%	0.12%
								rs377345545	C/T		0.00%	0.12%
								rs112002520	C/T		0.00%	0.12%
CACNA1C	26	both	yes	0.27	2.49	0.79	0.68	rs374295479	C/T		0.00%	0.12%
								rs141633456	G/A		0.00%	0.25%
								12:2778209	C/T		0.00%	0.12%
								12:2786965	G/A		0.00%	0.12%
								rs112940259	C/T		0.00%	0.12%
								12:2797767	G/A		0.00%	0.12%
								rs201258230	C/A		0.43%	0.12%
								rs41276710	T/C		0.87%	0.62%
								rs2302731	C/T		4.35%	2.72%
								rs56180838	C/T		4.78%	7.16%
								rs72552065	C/T		6.09%	5.19%
								rs10774053	G/A		20.00%	15.56%
								rs216008	C/T		20.43%	17.65%
								rs10848683	T/C		20.87%	16.67%
								rs142573111	C/G		0.43%	0.00%
								rs35185125	C/A		3.48%	0.62%
								9:35615921	C/T		0.00%	0.12%
CD72	8	both	no	0.24	2.47	0.63	0.63	rs369565307	T/C		0.00%	0.12%
								rs140933365	C/T		0.00%	0.12%
								rs117344011	G/A		1.30%	2.47%
								rs34791102	G/A		6.96%	5.80%
								rs2095858	T/A		46.09%	44.32%
								rs56371725	C/T		3.91%	0.86%
								rs117431067	T/C		0.00%	0.25%
								17:62130702	G/A		0.00%	0.12%
								rs16947420	C/T		0.43%	0.12%
ERN1	10	both	yes*	0.27	1.89	0.83	0.87	17:62122901	C/T		0.87%	0.25%
								rs56201815	G/A		0.87%	0.12%
								rs77619112	T/G		2.17%	2.72%
								rs113336518	C/T		3.04%	2.72%
								rs55939178	G/A		5.22%	3.21%
								rs196912	C/T		23.48%	23.95%
GNG5	1	case	yes	0.40	3.27	0.70	0.83	1:84967676	GA/G		3.48%	0.62%
								11:6452426	A/G		0.43%	0.00%
								rs77660779	C/A		0.43%	0.00%
								11:6452982	C/T		0.43%	0.00%
								11:6459624	C/T		0.43%	0.00%
								rs188970040	C/T		0.43%	0.00%
HPX	11	both	no	0.22	4.55	0.29	0.31	rs75099526	T/C		0.43%	0.00%
								11:6461979	A/G		0.43%	0.00%
								rs113295804	G/A		0.87%	0.00%
								rs117272845	A/G		2.17%	0.86%
								rs201631304	G/C		0.00%	0.12%
								rs35862450	G/A		20.00%	20.86%
								1:154408541	G/T		0.43%	0.00%
IL6R	7	both	no	0.21	1.78	0.95	0.62	1:154422408	G/A		0.43%	0.00%
								rs2229237	C/T		2.17%	0.49%
								rs28730735	C/T		0.00%	0.25%

								1:154437621	C/T		0.00%	0.12%
								rs143810642	C/T		0.00%	0.12%
								rs2228145	A/C		38.26%	41.23%
JAM2	8	case	yes	0.24	4.55	0.63	0.16	rs201699419	T/C		0.43%	0.00%
								rs56256482	T/G		0.43%	0.00%
								rs149216934	CT/C		0.43%	0.00%
								rs8134513	C/A		15.22%	8.64%
								rs2276220	G/A		14.78%	8.77%
								rs8133602	A/G		13.91%	8.89%
								rs2829870	T/A		11.74%	8.40%
								rs2829877	T/C		11.30%	8.03%
LAMB1	28	both	no	0.22	1.98	0.73	0.71	rs113389946	G/A		0.43%	0.00%
								rs200654865	G/T		0.00%	0.12%
								rs368805835	C/G		0.00%	0.12%
								7:107569936	G/C		0.00%	0.12%
								rs35915664	A/G		0.00%	2.96%
								rs57244800	C/A		0.00%	0.12%
								7:107572754	G/A		0.00%	0.12%
								rs138364317	G/A		0.00%	0.12%
								rs25660	G/A		0.00%	0.12%
								rs141390544	C/T		0.00%	0.12%
								7:107594112	T/A		0.00%	0.12%
								7:107595987	G/A		0.00%	0.12%
								7:107599826	C/T		0.00%	0.12%
								rs143597483	G/T		0.00%	0.25%
								7:107600992	G/A		0.00%	0.12%
								rs373152146	G/A		0.00%	0.12%
								7:107605053	A/G		0.00%	0.12%
								7:107605081	A/G		0.00%	0.12%
								rs140146478	G/C		0.00%	0.12%
								7:107643277	C/G		0.00%	0.12%
								rs139880642	T/TCC		0.87%	1.11%
								rs35710474	C/T		3.91%	6.54%
								rs11770141	C/T		11.74%	9.38%
								rs25659	G/A		16.96%	14.07%
								rs2230156	A/C		27.39%	28.15%
								rs20556	C/T		37.39%	37.41%
								rs7561	G/T		43.48%	41.85%
								rs2701034	A/G		44.78%	45.80%
MAP2K1	4	both	yes	0.27	1.86	0.97	0.82	15:66782108	CTATT/C		0.43%	0.25%
								rs16949939	C/T		3.91%	1.36%
								rs41306345	C/T		10.00%	8.64%
								rs16949924	G/C		31.74%	37.53%
MYC	4	both	yes*	0.24	1.42	0.94	0.94	8:128750586	CGAG/C		0.43%	0.00%
								rs61752959	G/A		1.30%	0.00%
								8:128752871	T/C		0.00%	0.12%
								rs4645959	A/G		4.78%	4.57%
P2RX5	14	both	no	0.24	3.38	0.82	0.27	rs1131057	C/T		0.43%	0.00%
								rs373182156	G/A		0.43%	0.00%
								17:3593388	C/T		0.43%	0.00%
								rs142264131	G/A		0.43%	0.00%
								rs186137135	G/A		0.87%	0.00%
								rs151100959	C/T		3.48%	0.49%
								17:3591871	G/A		0.00%	0.12%
								rs147009070	C/T		0.00%	0.12%
								17:3594965	A/AT		0.00%	0.12%
								rs200015059	T/C		0.00%	0.12%
								rs61748727	C/T		0.43%	0.49%
								rs144498633	G/T		0.43%	0.49%
								rs142863822	A/T		0.87%	0.37%
								rs3215407	T/TG		29.57%	30.12%

UPF2	14	both	no	0.21	2.65	0.51	0.57	rs374821084	C/T		0.43%	0.00%
								10:11978547	C/T		0.43%	0.00%
								rs144752256	T/C		0.43%	0.00%
								rs118004016	C/G		2.17%	0.25%
								10:11971932	C/A		0.00%	0.12%
								rs144437160	G/A		0.00%	0.12%
								rs145418095	A/G		0.00%	0.12%
								rs202183189	T/C		0.00%	0.12%
								10:11998385	G/A		0.00%	0.12%
								rs375701393	C/A		0.00%	0.12%
								rs35258162	A/T		0.00%	0.37%
								rs78315861	A/G		1.30%	0.25%
								rs11257490	G/A		14.78%	16.79%
								rs7921794	T/G		46.09%	44.32%

Table S2: Additional annotated variants and genes of interest from the rare variant analysis.

Table S3: Genes of interest from the rare variant analysis with a MGI AVSD phenotype

1. Freeman, S.B. et al. Ethnicity, sex, and the incidence of congenital heart defects: a report from the National Down Syndrome Project. Genet Med 10, 173-80 (2008).
2. Pradat, P., Francannet, C., Harris, J.A. & Robert, E. The epidemiology of cardiovascular defects, part I: a study based on data from three large registries of congenital malformations. Pediatric Cardiology 24, 195-221 (2003).
3. Harris, J.A., Francannet, C., Pradat, P. & Robert, E. The epidemiology of cardiovascular defects, part 2: a study based on data from three large registries of congenital malformations. Pediatric Cardiology 24, 222-35 (2003).
4. Zaidi, S. et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature 498, 220-3 (2013).
5. Priest, J.R. et al. De Novo and Rare Variants at Multiple Loci Support the Oligogenic Origins of Atrioventricular Septal Heart Defects. PLoS Genet 12, e1005963 (2016).
6. Garg, V. et al. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. Nature 424, 443-7 (2003).
7. Oyen, N. et al. Recurrence of congenital heart defects in families. Circulation 120, 295-301 (2009).
8. Maitra, M., Koenig, S.N., Srivastava, D. & Garg, V. Identification of GATA6 sequence variants in patients with congenital heart defects. Pediatric Research 68, 281-5 (2010).
9. (!!! INVALID CITATION !!!).
10. Al Turki, S. et al. Rare Variants in NR2F2 Cause Congenital Heart Defects in Humans. Am J Hum Genet 98, 592 (2016).
11. Winston, J.B. et al. Complex trait analysis of ventricular septal defects caused by Nkx2-5 mutation. Circ Cardiovasc Genet 5, 293-300 (2012).
12. Marino, B., Marcelletti, C., Giannotti, A. & Dallapiccola, B. Di George anomaly with atrioventricular canal. Chest 99, 242-3 (1991).
13. Xie, L. et al. Tbx5-hedgehog molecular networks are essential in the second heart field for atrial septation. Dev Cell 23, 280-91 (2012).
14. High, F.A. et al. Murine Jagged1/Notch signaling in the second heart field orchestrates Fgf8 expression and tissue-tissue interactions during outflow tract development. Journal of Clinical Investigation 119, 1986-96 (2009).
15. Maslen, C.L. et al. CRELD1 mutations contribute to the occurrence of cardiac atrioventricular septal defects in Down syndrome. Am J Med Genet A 140, 2501-5 (2006).
16. Li, Y. et al. Global genetic analysis in mice unveils central role for cilia in congenital heart disease. Nature 521, 520-4 (2015).
17. Franco, A., Zhang, L., Matkovich, S.J., Kovacs, A. & Dorn, G.W., 2nd. G-protein receptor kinases 2, 5 and 6 redundantly modulate Smoothened-GATA transcriptional crosstalk in fetal mouse hearts. Journal of Molecular and Cellular Cardiology 121, 60-68 (2018).
18. Qian, Y. et al. Multiple gene variations contributed to congenital heart disease via GATA family transcriptional regulation. J Transl Med 15, 69 (2017).
19. Watanabe, Y. et al. Activation of Notch1 signaling in cardiogenic mesoderm induces abnormal heart morphogenesis in mouse. Development 133, 1625-34 (2006).
20. Bax, N.A. et al. Cardiac malformations in Pdgfralpha mutant embryos are associated with increased expression of WT1 and Nkx2.5 in the second heart field. Developmental Dynamics 239, 2307-17 (2010).
21. Zhou, L. et al. Gata4 potentiates second heart field proliferation and Hedgehog signaling for cardiac septation. Proceedings of the National Academy of Sciences of the United States of America 114, E1422-E1431 (2017).
22. Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two sources of myocardial cells. Nat Rev Genet 6, 826-35 (2005).
23. Goddeeris, M.M. et al. Intracardiac septation requires hedgehog-dependent cellular contributions from outside the heart. Development 135, 1887-95 (2008).
24. Hoffmann, A.D., Peterson, M.A., Friedland-Little, J.M., Anderson, S.A. & Moskowitz, I.P. sonic hedgehog is required in pulmonary endoderm for atrial septation. Development 136, 1761-70 (2009).
25. Lee, S. et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 91, 224-37 (2012).
26. Lin, J.R., Zhang, Q., Cai, Y., Morrow, B.E. & Zhang, Z.D. Integrated rare variant-based risk gene prioritization in disease case-control sequencing studies. PLoS Genet 13, e1007142 (2017).
27. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25, 25-9 (2000).
28. Kircher, M. et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46, 310-5 (2014).
29. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285-91 (2016).
30. Wu, M.C. et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89, 82-93 (2011).

31. Mittal, A., Pulina, M., Hou, S.Y. & Astrof, S. Fibronectin and integrin alpha 5 play requisite roles in cardiac morphogenesis. Developmental Biology 381, 73-82 (2013).
32. Fritsche, L.G. et al. Association of Polygenic Risk Scores for Multiple Cancers in a Phenome-wide Study: Results from The Michigan Genomics Initiative. Am J Hum Genet 102, 1048-1061 (2018).
33. Hart, A., Papadopoulou, S. & Edlund, H. Fgf10 maintains notch activation, stimulates proliferation, and blocks differentiation of pancreatic epithelial cells. Developmental Dynamics 228, 185-93 (2003).
34. Kelly, R.G., Brown, N.A. & Buckingham, M.E. The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. Dev Cell 1, 435-40 (2001).
35. Watanabe, Y. et al. Fibroblast growth factor 10 gene regulation in the second heart field by Tbx1, Nkx2-5, and Islet1 reveals a genetic switch for down-regulation in the myocardium. Proceedings of the National Academy of Sciences of the United States of America 109, 18273-80 (2012).
36. Kent, W.J. et al. The human genome browser at UCSC. Genome Research 12, 996-1006 (2002).
37. Raney, B.J. et al. Track data hubs enable visualization of user-defined genome-wide annotations on the UCSC Genome Browser. Bioinformatics 30, 1003-5 (2014).
38. Chen, L. et al. Conditional ablation of Ezh2 in murine hearts reveals its essential roles in endocardial cushion formation, cardiomyocyte proliferation and survival. PLoS One 7, e31005 (2012).
39. Gaur, U. & Aggarwal, B.B. Regulation of proliferation, survival and apoptosis by members of the TNF superfamily. Biochemical Pharmacology 66, 1403-8 (2003).
40. Li, H. et al. Penetrance of Congenital Heart Disease in a Mouse Model of Down Syndrome Depends on a Trisomic Potentiator of a Disomic Modifier. Genetics 203, 763-70 (2016).
41. Pizzuti, A. et al. Mutations of ZFPM2/FOG2 gene in sporadic cases of tetralogy of Fallot. Human Mutation 22, 372-7 (2003).
42. Rodriguez-Viciana, P. et al. Germline mutations in genes within the MAPK pathway cause cardio-facio-cutaneous syndrome. Science 311, 1287-90 (2006).
43. Turner, C.J., Badu-Nkansah, K., Crowley, D., van der Flier, A. & Hynes, R.O. alpha5 and alphav integrins cooperate to regulate vascular smooth muscle and neural crest functions in vivo. Development 142, 797-808 (2015).
44. Nagasawa, T. et al. Molecular cloning and characterization of a murine pre-B-cell growth-stimulating factor/stromal cell-derived factor 1 receptor, a murine homolog of the human immunodeficiency virus 1 entry coreceptor fusin. Proceedings of the National Academy of Sciences of the United States of America 93, 14726-9 (1996).
45. Bult, C.J. et al. Mouse Genome Database (MGD) 2019. Nucleic Acids Res (2018).
46. Kamburov, A., Stelzl, U., Lehrach, H. & Herwig, R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res 41, D793-800 (2013).
47. Kamburov, A., Wierling, C., Lehrach, H. & Herwig, R. ConsensusPathDB--a database for integrating human functional interaction networks. Nucleic Acids Res 37, D623-8 (2009).
48. Snarr, B.S., Wirrig, E.E., Phelps, A.L., Trusk, T.C. & Wessels, A. A spatiotemporal evaluation of the contribution of the dorsal mesenchymal protrusion to cardiac development. Developmental Dynamics 236, 1287-94 (2007).
49. Mommersteeg, M.T. et al. Two distinct pools of mesenchyme contribute to the development of the atrial septum. Circulation Research 99, 351-3 (2006).
50. Stone, D.M. et al. The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. Nature 384, 129-34 (1996).
51. Marigo, V., Davey, R.A., Zuo, Y., Cunningham, J.M. & Tabin, C.J. Biochemical evidence that patched is the Hedgehog receptor. Nature 384, 176-9 (1996).
52. Hui, C.C. & Angers, S. Gli proteins in development and disease. Annual Review of Cell and Developmental Biology 27, 513-37 (2011).
53. Blacque, O.E. et al. Functional genomics of the cilium, a sensory organelle. Current Biology 15, 935-41 (2005).
54. Ahn, S. & Joyner, A.L. Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. Cell 118, 505-16 (2004).
55. Dai, P. et al. Sonic Hedgehog-induced activation of the Gli1 promoter is mediated by GLI3. J Biol Chem 274, 8143-52 (1999).
56. Mill, P. et al. Human and mouse mutations in WDR35 cause short-rib polydactyly syndromes due to abnormal ciliogenesis. Am J Hum Genet 88, 508-15 (2011).
57. Ocbina, P.J., Eggenschwiler, J.T., Moskowitz, I. & Anderson, K.V. Complex interactions between genes controlling trafficking in primary cilia. Nat Genet 43, 547-53 (2011).
58. Qin, J., Lin, Y., Norman, R.X., Ko, H.W. & Eggenschwiler, J.T. Intraflagellar transport protein 122 antagonizes Sonic Hedgehog signaling and controls ciliary localization of pathway components. Proceedings of the National Academy of Sciences of the United States of America 108, 1456-61 (2011).
59. Keady, B.T. et al. IFT25 links the signal-dependent movement of Hedgehog components to intraflagellar transport. Dev Cell 22, 940-51 (2012).

60. Jonassen, J.A., SanAgustin, J., Baker, S.P. & Pazour, G.J. Disruption of IFT complex A causes cystic kidneys without mitotic spindle misorientation. Journal of the American Society of Nephrology 23, 641-51 (2012).
61. Prevo, B., Scholey, J.M. & Peterman, E.J.G. Intraflagellar transport: mechanisms of motor action, cooperation, and cargo delivery. FEBS J 284, 2905-2931 (2017).
62. Cohen, E.D. et al. Wnt/beta-catenin signaling promotes expansion of Isl-1-positive cardiac progenitor cells through regulation of FGF signaling. Journal of Clinical Investigation 117, 1794-804 (2007).
63. Klaus, A. et al. Wnt/beta-catenin and Bmp signals control distinct sets of transcription factors in cardiac progenitor cells. Proceedings of the National Academy of Sciences of the United States of America 109, 10921-6 (2012).
64. Marguerie, A. et al. Congenital heart defects in Fgfr2-IIIb and Fgf10 mutant mice. Cardiovascular Research 71, 50-60 (2006).
65. Yang, J.T., Rayburn, H. & Hynes, R.O. Embryonic mesodermal defects in alpha 5 integrin-deficient mice. Development 119, 1093-105 (1993).
66. Bax, D.V. et al. Cell adhesion to fibrillin-1 molecules and microfibrils is mediated by alpha 5 beta 1 and alpha v beta 3 integrins. J Biol Chem 278, 34605-16 (2003).
67. Rentschler, S., Jain, R. & Epstein, J.A. Tissue-tissue interactions during morphogenesis of the outflow tract. Pediatric Cardiology 31, 408-13 (2010).

Conclusions

Next generation sequencing technology has given us unprecedented ability to investigate the genetic basis of diseases, which is leading to a rapid increase in the understanding of the human genome. Despite this, there are many complex genetic diseases in which the genetic origins are still unknown. In this thesis, we use whole exome sequencing to focus on elucidating the genetic mechanisms of congenital heart defects in both its syndromic and non-syndromic presentations.

While congenital heart defects are associated with numerous aneuploidies, this thesis limits its focus to two of the most common human aneuploidies, Down syndrome (DS) and Turner syndrome (TS). In DS, 20% are born with an atrioventricular septal defect (AVSD) and in TS, 30% are born with bicuspid aortic valve (BAV) and a thoracic aortic dilation (TAD). Although not as common, AVSD, BAV, and TAD are also seen in the non-syndromic population.

Because congenital heart defects are known to be complex and genetically heterogeneous, we utilized a sophisticated statistical genetics algorithm, SKAT, to test the cumulative effects of variants on the gene level. Through this approach, we have identified several candidate genes that are major components of the cardiovascular developmental pathway, primary cilia signaling pathway, and extracellular remodeling pathway, among others.

For TS, we found that hemizygosity of *TIMP1* and genetic variants in its paralogue, *TIMP3*, synergistically increase the risk for BAV and TAD. Because of the thorough understanding of TIMPs in extracellular remodeling and this pathway's role in the pathogenesis of TAD, the results of this study have clear biological relevance. These results also point to common genetic mechanism in those with BAV with TAD in the non-syndromic population. In the non-syndromic population, *TIMP1* and *TIMP3* expression has been reported to be significantly reduced and *MMP2* and *MMP9* expression a significantly increased. The results of our syndromic BAV/TAD study in combination with the previously reported non-syndromic studies leads us to the conclusion that the MMP/TIMP ratio imbalance is the genetic entry point for TS associated BAV/TAD.

For DS, we found that rare deleterious variants in a primary cilium component, *CEP290*, is associated with AVSD. Mapping the variants to their location within *CEP290* shows that each fall within a functional or binding domain and have the potential to interfere with ciliogenesis or ciliary function. There is ample evidence for the fundamental role of the primary cilium in heart development in which it has been shown that the disruption of proper ciliogenesis and ciliary function can alter Shh signaling and result in CHDs, including AVSD.

AVSDs are a clinically severe congenital heart defect that accounts for about 7% of all CHDs and is most often seen in children with Down syndrome, but the majority of the remaining cases are sporadically occurring non-syndromic AVSD. We aimed to add to the growing list of candidate genes and risk factors of sporadic non-syndromic AVSD cases remain unknown. Through this study, we found several genes significantly associated with non-syndromic AVSD. These genes are involved in cardiac transcription regulation (*GATA6*, *ZFPM2*, and *FGF10*), ciliome genes (*GLI3*, *GLI1*, *PTCH1*, and *SMO*), and a recently discovered modifier *JAM2*; also novel mechanisms such as transcription factors (*ATF1* and *MYC*), signaling regulators (*COPS3*, *ERN1*, *GNG5*, and *P2RX5*), and an extra cellular matrix component (*LAMB1*) were identified.

Our investigations could have a direct impact on the clinical management of the striking number infants per year that are born with CHDs and the adults who experience cardiovascular disease later in life. Targeted gene panels of candidate genes are quickly becoming the method of choice for identifying if patients have a disease-causing variant, particularly in genetic counseling of prospective parents. Therefore our work could be directly

translated to the clinic. Ultimately, our study could help change the medical management of these patients by identifying new molecular targets and novel methods of intervention.