INSTITUTO DE CIENCIAS E INNOVACIÓN EN MEDICINA
Facultad de Medicina
Clínica Alemana - Universidad del Desarrollo

GENETIC MODIFIERS ASSOCIATED WITH THE VARIABILITY OF THE AORTIC PHENOTYPE IN PATIENTS WITH INHERITABLE CONNECTIVE TISSUE DISORDERS (ICTDs)

BY: YANIRETH JOSÉ JIMENEZ BEJARANO

Thesis presented to the School of Medicine Clínica Alemana-Universidad del Desarrollo in order to apply for doctoral degree (PhD) in Sciences and Innovation in Medicine

SUPERVISOR:
DR. JUAN FRANCISCO CALDERÓN GIADROSIC

MARCH 2022
SANTIAGO

With love and eternal gratitude,

I would like to dedicate my thesis to my family and

especially in memory of my mother

ACKNOWLEDGMENTS

I would like to thank the following people for helping with this research project: first of all, I would like to express my gratitude to my tutor, Dr. **Juan Francisco Calderón**, who has supported me throughout this research project. I am extremely grateful for your patience, for your guidance throughout the doctorate, for our friendly meetings, for your personal support in my academic endeavors, and for your insightful comments that prompted me to sharpen my thinking and took my work to a higher level.

To my thesis committee: **Gabriela Repetto, Báltica Cabieses and Matthieu Miossec**, for dedicating their time to guide and support me during the execution of my thesis. Your thoughtful and detailed feedback has been very important to me.

Thank our **patients** for participating in this study. I also thank the physicians of the Clínica Alemana Universidad del Desarrollo and the Instituto Nacional del Thorax for collaborating in the recruitment and clinical evaluation of patients.

A very special thanks to **David Gorkin**, for giving me his valuable suggestions for the execution of this project.

I would also like to thank the coordinator of the doctoral program, **Andrés Klein**, for his effort to help doctoral students in everything.

I also thank all my friends and co-workers, who were available to help me in my laboratory work and supported me in the development of my thesis, including:

Valentina Jarur, Marcelo Rojas, Jimena Castillo, Glenda Cofre, Valeria Olguin and Juan Carlos Rubilar.

And my biggest thanks to my entire family for all the support they have given me in everything I do. To my son **Mathias**, for filling my days with happiness, to my husband **Leonardo**, for all your love and for being an incredible person who manages to calm me in moments of anguish.

I am grateful to the funding sources that supported this Thesis, the Fondecyt project N ° 11170353, Internal Project UDD2020, CAS-UDD support grant.

CONTENTS

FIGURES

TABLES

LIST OF ABBREVIATIONS

AA	Aortic aneurysm
AAA	Abdominal Aortic Aneurysm
AoR	Aortic root
BAV	Bicuspid Aortic Valve
BMP	Bone Morphogenic Protein
BQSR	Base Quality Score Recalibration
BWA	Burrows-Wheeler Aligner
CAD	Coronary Artery Disease
CLRT	Composite Likelihood Ratio Test
COL3A1	Collagen Type III Alpha 1 Chain
COPD	Chronic Obstructive Pulmonary Disease
ECM	Extracellular Matrix
ELN	Elastin
FBLN4	Fibulin 4
FBN1	Fibrillin 1
FTAAD	Familial Thoracic Aortic Aneurysm And Dissection
GenTAC	Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions
GO	Gene Ontology
ICTD	Inheritable Connective Tissue Disorders
LDS	Loeys-Dietz Syndrome
LTBP	Latent TGF-β Binding Protein
MAP3K1	Mitogen-Activated Protein Kinase Kinase Kinase 1
MFS	Marfan Syndrome

NGS	Next-Generation Sequencing
NICD	Notch Intracellular Domain
NMD	Nonsense Mediated Decay
RBPJ	Recombinant Binding Protein Suppressor
PPARD	Peroxisome Proliferator-Activated Receptor δ
PTC	Premature Termination Codon
PTPTJ	Protein Receptor Tyrosine Phosphatase Type J
SGS	Shprintzen-Goldberg Syndrome
TAA	Thoracic Aortic Aneurysm
TF	Transcription Factor
TGF-β	Transforming Growth Factor Beta
VAAST	Variant Annotation, Analysis, and Search Tool
VAT	Variant Annotation Tool
vEDS	vascular Ehlers-Danlos Syndrome
VEGFR2	VEGF receptor 2
VSMC	Vascular-Type Smooth Muscle Cell
VST	Variant Selection Tool
ES	Exome Sequencing
ZP	Zona Pellucida

ABSTRACT

Inheritable Connective Tissue Disorders (ICTDs) are rare genetic diseases that involve variants in genes that encode for proteins of the extracellular matrix (ECM). ICTDs affect patients from birth and their symptoms are present in the cardiovascular system, musculoskeletal system, the skin, the eye and the respiratory system. This family of diseases includes Marfan Syndrome (MFS), Loeys-Dietz Syndrome types I, II, III and IV (LDS), Ehlers-Danlos Syndrome type IV (vEDS), among others, and are caused by variants in genes that code for different effectors or regulators of the Transforming Growth Factor beta (TGFβ) signaling pathway, including genes that encode for ECM components that play active roles in the signaling activity of this pathway. The cardinal feature of these diseases is the enlargement and rupture of big vessels and, most prominently, the aorta. Aneurysm and dissection of the aorta is the main cause of mortality in these patients but the clinical course of MFS and other ICTDs differs considerably in terms of age of onset and severity, even among individuals who share the same causative variant.

This led us to hypothesize the existence of genetic variants elsewhere in the genome that influence the severity of the cardiovascular phenotype in MFS. We recruited familial and sporadic MFS patients that were subsequently classified as having severe (n=8) or mild aortic phenotype (n=14) according to the age of presentation of the first cardiovascular manifestation and/or

catastrophe related to the aorta. We used Exome Sequencing (ES) to identify genetic variants that may be associated with the severity of this clinical manifestation, and we performed linkage analysis using the software VAAST. We identified 5 genes associated with severe aortic phenotype and 3 genes that could be protective of this phenotype in MFS. These genes regulate components of the ECM, TGFβ pathway and other signaling pathways that are involved in the maintenance of the ECM and/or angiogenesis. Further studies will be required to understand the functional effect of these variants and explore novel, personalized risk management strategies and, potentially, new therapies for these patients.

RESUMEN

Las Enfermedades Heredables del Tejido Conectivo (EHTCs) son enfermedades genéticas raras que involucran mutaciones en los genes que codifican para proteínas de la matriz extracelular (MEC). Las EHTCs afectan a los pacientes desde el nacimiento y sus síntomas están presentes en el sistema cardiovascular, el sistema musculoesquelético, la piel, los ojos y el sistema respiratorio. Esta familia de enfermedades incluye el Síndrome de Marfan (SMF), Síndrome de Loeys-Dietz tipos I, II, III y IV (SLD), Síndrome de Ehlers-Danlos tipo IV (SED-IV) entre otros, todos ellos causados por mutaciones en genes que codifican diferentes efectores o reguladores de la vía de señalización del factor de crecimiento transformante beta (TGFβ por sus siglas en inglés), incluidos los genes que codifican los componentes de la MEC que desempeñan un papel activo en la señalización de esta vía. La característica cardinal de estas enfermedades es la dilatación y la ruptura de los grandes vasos, especialmente la aorta. El aneurisma aórtico y la disección es la principal causa de mortalidad en estos pacientes, pero el curso clínico de SMF y otras EHTCs difiere considerablemente en relación con la edad de inicio y la gravedad, incluso entre individuos que comparten la misma mutación causal.

Esto nos llevó a plantear la hipótesis de la existencia de variantes genéticas en otras partes del genoma que influyen en la gravedad del fenotipo

cardiovascular en el SMF. Reclutamos pacientes con SFM familiar y esporádico que posteriormente se clasificaron como fenotipo aórtico severo (n = 8) o leve (n = 14) según la edad de presentación de la primera manifestación cardiovascular y/o catástrofe relacionada con la aorta. Usamos secuenciación del exoma completo (ES por sus siglas en inglés) para identificar variantes genéticas que pueden estar asociadas con la gravedad de esta manifestación clínica, y realizamos análisis de ligamiento usando el software VAAST. Identificamos 5 genes asociados con el fenotipo aórtico severo y 3 genes que podrían ser protectores de este fenotipo en el SMF. Estos genes regulan los componentes de la MEC, la vía de señalización del TGFβ y otras vías de señalización que están involucradas en el mantenimiento de la MEC y/o la angiogénesis. Se requieren más estudios para comprender el efecto funcional de estas variantes y explorar un manejo de riesgos personalizado y, potencialmente, nuevas terapias para estos pacientes.

1. INTRODUCTION

1.1 Inheritable Connective Tissue Disorders (ICTDs)

Inheritable Connective Tissue Disorders are rare diseases caused by variants in the genes that code for extracellular matrix (ECM) proteins and genes that code for proteins of the TGFβ signaling pathway that have implications in the homeostasis of the extracellular matrix. ICTDs often present with an autosomal dominant mode of inheritance, although other modes of inheritance are observed[1]. The ECM has many functions, including providing mechanical-structural support for neighboring cells and regulating cellular behavior through the modulation of some signaling pathways. In relation to the cardiovascular system, the ECM provides flexibility and elasticity to the heart and vascular tree[2]. The cardiovascular manifestations of the different ICTDs vary enormously, ranging from those requiring relatively standardized clinical management such as mild heart valve defects to the most lethal, in which the aneurysm and dissection of the aorta artery are the most prominent; in diseases such as Marfan Syndrome (MFS), Loeys-Dietz Syndrome (LDS) and Vascular Ehlers-Danlos Syndrome (vEDS), because they are the main cause of death in these patients[3–7].

Since the ECM is present in all tissues and organs, variants that affect any of its components affect multiple systems in the human body. The consequence of this is a phenomenon called **variable expressivity**, in which these diseases show a wide range of symptoms and phenotypic characteristics that may or may not be present

between different patients, even though they may have the same causative variants. This clinical heterogeneity and the overlap of phenotypes between different ICTDs such as those just mentioned makes the diagnostic process an extremely challenging task from a clinical point of view[1,8]. The identification of genetic variants associated with or causing this phenotypic variability would facilitate the decision-making process regarding the optimal moment to carry out surgical repair of the affected blood vessels and the design of drugs aimed at slowing the progression of aortic disease (ideally to the point of avoiding the need for surgery). At present, our understanding of the effects of genetic changes on the frequency and progression of the clinical phenotype of ICTDs is very limited. Deepening our understanding is therefore highly important for this group of patients.

1.2 TGF-β Signaling

Transforming growth factor beta (TGF-β) is a potent cytokine belonging to a family of dimeric polypeptidic growth factors that also include bone morphogenic proteins (BMPs) and activins[9]. TGF-β emerges as a mechanism to program and control cell development and behavior throughout cellular functions and physiological processes that include cell proliferation, differentiation, apoptosis, regulation of embryonic development, wound healing, angiogenesis, among others. Consequently, the dysfunction of these processes can trigger serious diseases[10].

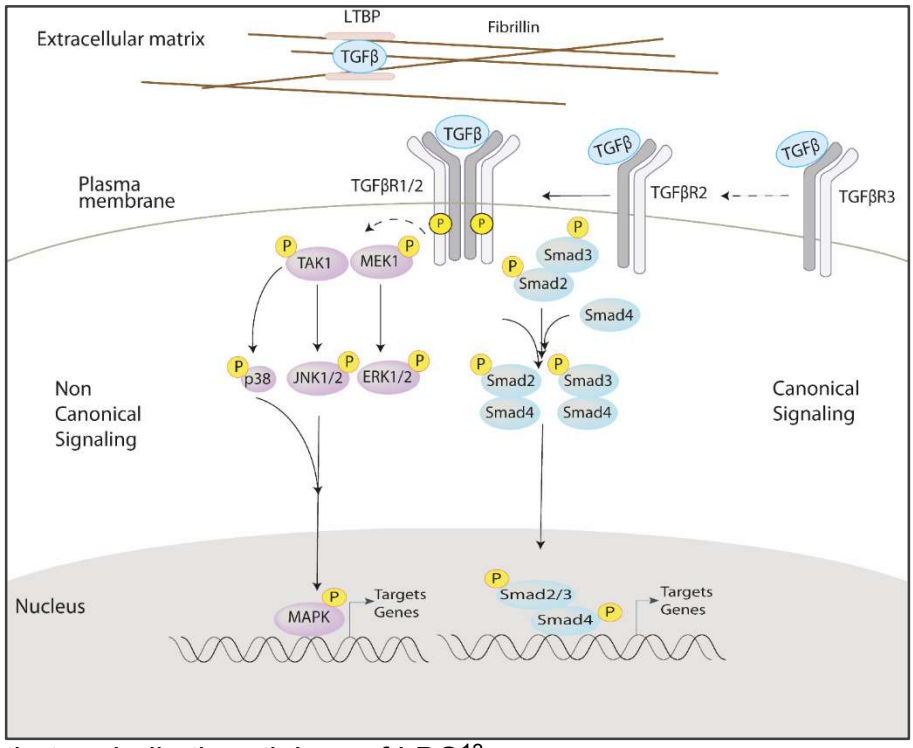
Most of secreted TGF-β is stored in the extracellular matrix as an inactive complex composed of TGF-β and a protein called latent TGF-β binding protein (LTBP). The

binding of TGF-β and LTBP is through a disulfide bond that prevents it from binding to its receptors[10]. The regulatory function of TGF-β cell processes occurs when this molecule binds to three high-affinity cell surface receptors known as TGFβR1, TGFβR2, and TGFβR3 (figure 1). TGF-β binds to the type 3 receptor that presents TGF-β to type 2 receptors, or it can also bind directly to type 2 receptors and subsequently bind to type 1 receptors, thus stimulating its protein kinase activity[9,11]. Once type 1 receptors are activated, both the canonical and non-canonical signaling pathway of TGF-β can be activated in different biological settings[12].

In the canonical TGF-β signaling pathway, after activation of the type 1 receptor, the binding and phosphorylation of Smad2 and Smad3 is induced. The Smad2/3 complex then binds to Smad4 forming a super complex that translocate to the nucleus where it acts in a cell-specific way, where by binding to promoters of genes and in cooperation with different transcription factors (TFs) and cofactors, these complexes control the transcription of hundreds of genes[9,11].

Non-canonical TGF-β signaling refers to the activation of alternative signaling pathways after the binding of TGF-β to the receptor. These include the MAPK (ERK, p38, and JNK), PI3K/Akt, and Rho GTPase pathways[12].

Although the TGF-β signaling pathway is activated by a series of organized and complex regulatory mechanisms that are essential for human development, in some ICTDs such as MFS, TGFβ does not bind to microfibrils because of the presence of a defective fibrillin in the ECM. Therefore, the latent, inactive form is not generated, resulting in elevated levels of circulating TGF-β. It thus binds uncontrollably to its dimeric receptor, activating the phosphorylation cascade[13]. Variants in genes that encode downstream effectors of the TGF-β signaling pathway (*TGFBR1*, *TGFBR2*, and others) cause perturbations of the signaling



pathway that underlie the etiology of LDS[13].

Figure 1 Canonical and non-canonical TGF-β pathways.

1.3 Aortic aneurysm

Aortic disease represents one of the main causes of morbidity and mortality in industrialized society. This is reflected by approximately 2% of all deaths in this society being caused by an aortic aneurysm (AA)[2]. This situation is similar in Chile where, according to the Department of Health, 25,744 people died from "diseases of the circulatory system" (I00 -I99 ICD-10 classification) in 2011. Among them, 6,323 died from complications associated with aortic aneurysms and other blood vessel diseases (Department of Health, 2017). Aneurysms are defined as a local dilation of the artery over 50% of the normal diameter[14]. An aneurysm can progress to fatal aortic dissection and rupture at varying rates if it is not repaired by surgical intervention. The diagnosis of AAs often occurs in late stages of the disease. Based on their location, aneurysms can be classified into two main groups: thoracic aortic aneurysms (TAAs) and abdominal aortic aneurysms (AAAs). Despite constituting the same primary anatomical defect, the pathophysiology of these two types of aneurysms is significantly different due to, among other factors, the different embryonic origins, structural differences and mechanical properties of the aortic wall in the different regions. Moreover, these regions present distinct degrees of susceptibility to suffer atherosclerotic processes, and to immune mediators and cell signaling pathways that have implications in the development of an aortic aneurysm, such as the transforming growth factor TGF-β pathway[15].

When the aorta artery is affected by disease, it can divide (dissection) or dilate (aneurysm) and, in either case, the rupture can have fatal results[16]. This occurs in a variety of diseases and conditions with different origins.

Aortic aneurysms (AAs) of **genetic origin** are caused by Inheritable Connective Tissue Disorders (Marfan syndrome, Loeys – Diez syndrome and Ehlers – Danlos syndrome type IV or vascular type and others), Turner syndrome, Arterial tortuosity syndrome, non-syndromic thoracic aortic aneurysms, and others[17]. Amongst ICTDs, AA is the most life-threatening manifestation and understanding its molecular and cellular basis is the focus of much of current research.

Although AA can be due to a variety of genetic conditions, most of the conditions that trigger the formation of an aneurysm are attributed to chronic diseases of **non-genetic origin**, the most common of these being atherosclerosis (hardening of the arteries) which has a prevalence that is steadily increasing worldwide[18]. The number of adults with hypertension increased from about 600 million in 1975 to 1.13 billion in 2015[19] and it has been estimated that it will increase to 1.5 billion in 2025[20] due to aging, the growth of population and changes in behavioral risk factors[20].

The last group of conditions associated with AA are those of **congenital origin**, they constitute malformations of the heart and large vessels that have important functional consequences for the cardiovascular system. AA of congenital origin have a strong genetic component, but this genetic basis is often poorly defined due

to genetic heterogeneity (different genes causing the same clinical phenotype) and low penetrance of variants in these genes (i.e., an individual may have a pathogenic variant but not present the associated phenotype). The most important of these AAs is the Bicuspid Aortic Valve (BAV), which is the most common congenital heart defect with an estimated prevalence of 0.5 to 1,3%[21], with a genetic component that has been detected in a minority of non-syndromic AVB cases[22]. The most common phenotypic complications in patients with BAV are calcific aortic stenosis and ascending aortic aneurysm, the latter being a 6-fold higher risk in these patients compared to the general population[23,24].

Anatomically, the wall of the aorta artery is made up of three layers of connective tissue, which are composed of different cell types and components of the ECM (figure 2):

• An internal layer (*tunica intima*), composed of endothelial cells that line the lumen of the vessel and are separated from the middle layer by elastic fibers[25].

• An intermediate layer (*tunica media*), populated by VSMCs, also contains elastic fibers, microfibrils containing elastin and fibrillin, collagen fibers and proteoglycans[25].

• An outer layer (*tunica adventitia*), populated by fibroblasts and collagen if separated from the middle layer by elastic fibers[25].

The main mechanisms by which AAs occur are unknown. However, there is significant evidence that the formation of aneurysms is the product of changes in

the ECM of the aortic wall and that they are mediated by disturbances in the TGFβ signaling pathway[14,26].

Uncontrolled TGFβ signaling leads to increased expression and/or activities of enzymes such as metalloproteinases that degrade ECM and elastic fibers, increased proliferation and migration of VSMCs, and excessive secretion and deposition of collagen. Each of these changes leads to maladaptive remodeling of the ECM and VSMC dysfunction that favors the weakening of the aortic wall, making it more prone to dilation, dissection and rupture (figure 2)[27].

In addition, there is evidence of an important role for innate immune cells in the pathophysiology of aneurysms, showing that monocytes/macrophages infiltrate the vessel wall and release proteases (including elastase and other metalloproteinases) that compromise the integrity of the vascular wall through the degradation of the extracellular matrix (ECM)[28,29]. For example, Tang *et al*., found inflammatory infiltrates in the aortic wall that were associated with the production of INF-gamma in patients with ascending TAA; He *et al*., identified an inflammatory component in the medial degeneration of the aorta, composed mainly of T cells and macrophages, in patients with TAA[30,31].

In summary, a better understanding of the molecular mechanisms that lead to the formation of aneurysms is necessary in order to identify predisposing factors, significantly improve early detection protocols and propose individualized treatments for people suffering from this disease.

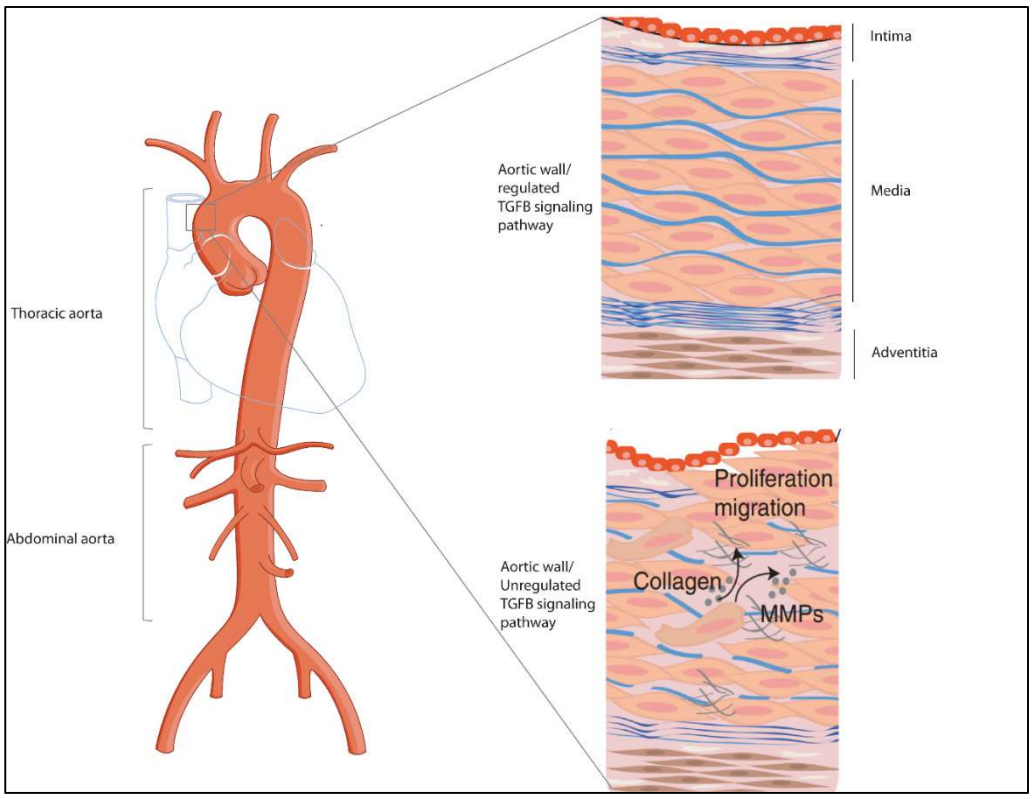


Figure 2. Pathophysiology of aortic artery in unregulated TGFβ signaling pathway. Adapted from https://Smart.servier.com and MacFarlane et al. Cold Spring Harb Perspect Biolog, 2017

Patients with TAAs can be classified into 3 groups: I) inherited TAA syndromes, such as Marfan syndrome (MFS), Ehlers Danlos syndrome (EDS) and Loeys-Dietz syndrome (LDS); II) familial non-syndromic TAA disease; and III) sporadic cases of TAA without a previous family history, detected in advanced age. Genetically caused TAAs represent approximately 30% of all TAAs and are the result of disease-causing variants in genes that code for key proteins in either vascular-type smooth muscle cells (VSMCs), the extracellular matrix (ECM) or transforming growth factor (TGF-β) signaling. These variants are often located in the genes:

fibrillin 1 (*FBN1*), Procollagen type III (*COL3A1*), elastin (*ELN*), fibulin 4 (*FBLN4*), among others (Table 1 and Figure 3)[32–39].

The genetic forms of TAA present a high variability in their phenotype and stand out particularly in cases where an identical causal genetic variant result in different clinical manifestations[40,41]. Understanding this phenotypic variability would allow clinicians to predict clinical outcomes in these patients. This warrants further investigations to achieve a better understanding of the genetic mechanisms related to the pathophysiology of TAA and aneurysm dissection.

Table 1: genes associated with TAA related to the ECM and the TGF-β pathway

Location	Genes	Disease	MIM #
Extracellular matrix	*FBN1*	Marfan Syndrome	154700
	COL3A1	Ehlers-Danlos Syndrome type IV	130050
	FBLN4, ELN	Cutis Laxa with Aneurysm	614437, 123700
TGF-β pathway	*TGFBR1*	Loeys-Dietz Syndrome type I	609192
	TGFBR2	Loeys-Dietz Syndrome tipo II	610168
	SMAD3	Loeys-Dietz Syndrome tipo III	613795
	TGFB2	Loeys-Dietz Syndrome tipo IV	614816
	TGFBR3	Loeys-Dietz Syndrome tipo V	190230
	SKI	Shprintzen-Goldberg Syndrome	182212
Smooth Muscle Cells	*SLC2A10*	Arterial Tortuosity Syndrome	208050
	NOTCH1	Bicuspid Aortic Valve with Aneurysm	109730
	MYH11, ACTA2, MYLK, PRKG1	Familial Thoracic Aortic Aneurysm and Dissection	160745, 611788, 613780, 176894

Below is a brief clinical-molecular description of the most common ICTDs.

1.4 Marfan syndrome (MFS)

Marfan syndrome (MFS, OMIM #154700) is a rare connective tissue disorder with a reported incidence of between 1 in 3000 to 5000 individuals[42,43]. MFS follows an autosomal dominant mode of inheritance and is caused by variants in *FBN1*, the gene encoding for fibrillin-1 an extracellular matrix (ECM) protein that is the major structural component of microfibrils[6,44](Figure 3). The clinical features of MFS include aortic root aneurysm and/or dissection, ectopia lentis, and overgrowth of large bones, among others[3].

Clinical manifestations of MFS are highly variable with significant phenotypic heterogeneity even among individuals harboring the same pathogenic variant in *FBN1*, including intrafamilial variability[45,46]. Prognosis in MFS is related to life-threatening complications such as dissection of the aorta which is the main cause of mortality in these patients. The clinical course of cardiovascular manifestations in MFS differs considerably in age of onset and severity[33]. Studies have shown that some variants in *FBN1* (c.1286 G > C, IVS25+49delTAAGA and IVS40–35C>T) produce variations in its expression and suggest that these variants could change the severity of the disease[47,48]. In contrast, other studies on mice models of MFS have identified other loci that modulate the cardiovascular phenotype of Marfan syndrome[49–51].

There is significant evidence that aneurysms are a consequence of changes in the ECM integrity of the aortic wall and that these would be mediated by perturbations

in the transforming growth factor-β (TGFβ) signaling pathway[52]. Through an alteration in the bioavailability of TGFB produced by a defective fibrillin-1 that can no longer sequester TGFβ increasing its activity levels[53], which triggers inflammation, fibrosis, and activation of metalloproteinases *MMP-2 and MMP-9*, leading to significant loss of vascular smooth muscle Cells (VSMCs)[52]. These factors, together with the decrease in collagen content of ECM, reduce aortic structural integrity and lead to dilation or aneurysm in MFS[54]. In fact, it has been shown in a mouse model of MFS that inhibition of TGF-β signaling prevents the development of cardiovascular phenotypes, and this is independent of the causal variant in the *FBN1* gene[55].

1.5 Loeys-Dietz Syndrome (LDS)

LDS was first described in 2005. It is an autosomal dominant connective tissue disorder. It has craniofacial, neurological, skeletal and cardiovascular manifestations and is characterized by extensive vascular abnormalities that include arterial tortuosity and TAA with dissection, the latter two affecting 95% of patients[5]. LDS is caused by variants in the genes of the TGFβ signaling pathway: TGF-β ligands (*TGFB2* and *TGFB3*), their receptors (*TGFBR1* and *TGFBR2*) or intracellular mediators Smad2 and Smad3 (*SMAD2* and *SMAD3*) (Figure 3)[56]. When LDS was first described, two different subtypes (type 1 and type 2) were described depending on the severity of the craniofacial and cutaneous features. However, in 2014, when the revised nosology was proposed, 6 different LDS subtypes were described according to the gene involved. Thus, subtypes 1, 2, 3, 4,

5 and 6 correspond to variants in the *TGFBR1, TGFBR2, SMAD3, TGFB2, TGFB3* and *SMAD2* genes respectively[57].

Clinical similarities with MFS make it likely for LDS patients to be misdiagnosed. Compared to MFS, it has more severe cardiovascular manifestations including aortic aneurysms that tend to dissect or rupture at a smaller diameter and at a younger age. Furthermore, aortic aneurysms are not confined to the root or ascending aorta, but frequently involve the lateral aortic branches and cerebral vessels. Arterial tortuosity mainly affects the vessels of the neck and head[58].

1.6 Ehlers-Danlos Syndrome type IV (vEDS)

Ehlers-Danlos Syndrome type IV or vascular type (vEDS) is caused by variants in the *COL3A1* gene (Figure 3), which encodes the alpha chain of type III collagen, which is highly expressed in the arterial system and lumen organs. Its estimated prevalence ranges from 1 in 50000 to 1 in100000[59,60] and represents approximately 5% of all EDS cases[61].

vEDS follows an autosomal dominant mode of inheritance, characterized by abnormalities of the skin, joints, hollow organs, and blood vessels and has the worst prognosis amongst all EDS subtypes, due to the rupture of arteries and hollow organs[59]. Its clinical diagnosis is based on the presence of a family history of vEDS, arterial rupture at an early age, spontaneous perforation of the sigmoid colon in the absence of intestinal pathology, and uterine rupture in the absence of a previous cesarean section[62]. Aortic dissection or rupture (but also of

other vessels) is the main cause of death in these patients, with a mean survival of 50 years[63]. Unlike MFS and LDS, vascular accidents and rupture of organs with lumen in patients with vEDS are not preceded by gradual dilation of the vessels, so this progression cannot be monitored and therefore cannot be predicted much less undergo prophylactic repair surgeries, as is done with patients with the other syndromes mentioned[63].

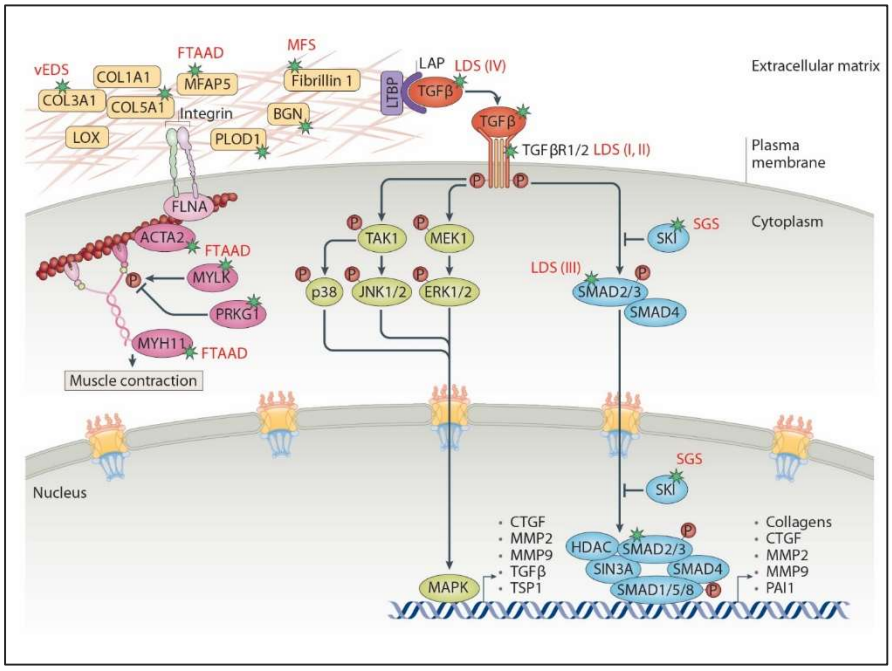


Figure 3. Signaling pathways involved in the pathogenesis of TGFβ vascular diseases.

Genes encoding ECM components (yellow boxes), muscle contraction machinery (purple boxes), and TGFβ signaling pathway (green and blue boxes) can harbor variants that cause different forms of aneurysm syndromes. familiar aortic. In red, diseases caused by variants in the respective genes: vEDS = Vascular Ehlers-Danlos syndrome, FTAAD = familial thoracic aortic aneurysm and dissection, MFS

= Marfan syndrome, LDS = Loeys-Dietz syndrome, SGS = Syndrome of Shprintzen-Goldberg. Adapted from Verstraeten et al, 2017[64].

Because of the great clinical heterogeneity and the clinical overlap between the different ICTDs, the diagnostic process is a difficult task in which any ambiguity could have serious consequences for the patient. The advent of next-generation sequencing (NGS) techniques has revolutionized the detection of new variants and the discovery of genes associated with ICTDs. In some fatal cases, the rupture of the blood vessels is the main cause of mortality in these patients the age of onset and severity of which differs considerably, even between individuals with the same disorder and carrying the same variant. This led us to think about the possibility of the existence of variants in loci other than those of the genes that cause the disease as responsible for these differences. By way of example, it has been reported that, in Mendelian diseases such as cystic fibrosis, caused by point variants in the *CFTR* gene, modifier genes have been identified that considerably modulate the severity of the lung disease[65], which reinforces our hypothesis. The identification of these loci represents an opportunity to improve the quality of life of these patients, particularly in the Chilean population, through the identification and understanding of the molecular mechanisms of the genetic variants that modify the severity of their clinical course.

1.7 Hypothesis/ General and specific aims

Connective tissue disorders are caused by variants in genes involved in ECM and in the TGF-β signaling pathway. It is well described that both Marfan syndrome, Loeys-Dietz syndrome and other ICTDs have a wide clinical variability in individuals with the same disease and even the same causing variant, which makes the genotype-phenotype correlation necessary to be able to address in part difficult. more efficiently to these patients. For all the aforementioned, we consider the following hypothesis:

1.7.1 Hypothesis

The effect of causative variants in patients with ICTDs is modified by variants elsewhere in the genome that modulate the effect of the pathogenic variant to alter the severity of the cardiovascular phenotype.

1.7.2 General Aim

To identify and understand the molecular mechanisms of genetic variants associated with variability in the severity of the aortic phenotype for patients with ICTDs.

1.7.3 Specific Aims:

1.7.3.1 Specific Aim #1:

To document epidemiological, familial, phenotypic (clinical) and molecular diagnostic data of patients with aortic diseases that allow us to identify patients with a genetic origin of this disease (ICTDs) in Chile.

1.7.3.2 Specific Aim #2:

To identify genetic variants associated with extreme presentations of an ICTD through Exome Sequencing (ES).

2. EPIDEMIOLOGICAL, FAMILIAL, PHENOTYPIC (CLINICAL) AND MOLECULAR DIAGNOSTIC DATA OF PATIENTS WITH AORTIC DISEASES IN CHILE.

2.1 INTRODUCTION

Aortic diseases are a group of arterial diseases, among which are: aortic aneurysms (thoracic aortic aneurysm (TAA) and abdominal aortic aneurysm (AAA)), pseudoaneurysms, atherosclerotic and inflammatory conditions, acute aortic syndromes (aortic dissection), intramural hematoma, among others, aortic aneurysm being the most prominent of all. Diagnosis of aortic disease often occurs after a long asymptomatic period, which is why it is usually diagnosed when acute aortic syndrome occurs, which then requires rapid diagnosis and treatment to reduce poor prognostic rates[16]. Over the past two decades, the mortality rate from aortic disease (including TAA, AAA, and acute aortic dissection) has increased from 2.49 per 100,000 in 1990 to 2.78 per 100,000 in 2010 with higher rates in men. This increase appears to be most evident in developing countries, with a median death rate of 0.71 compared to 0.22 in developed countries[16,66,67], this reflects the need to know more about this disease in developing countries like Chile. A better understanding of the molecular basis of aortic disease could help to better clinical management and prevention, which could be reflected in a decrease in the morbidity and mortality rates in these patients.

A key strategy to know more about this disease in Chile, is the creation of the first registry of patients with aortic diseases in Chile. This registry could be an effective tool that allows a better understanding of the natural history of aortic disease,

through the information obtained from patients, as well as various specialists providing clinical, epidemiological, molecular diagnostic, and health perception information. All this effort is justified by the need to find answers to the many questions that arise from these patients: specific clinical manifestations of genes, genotype-phenotype correlations, and specific responses of genes, also factors as well as clinical results and prognosis that allow the design of an efficient strategy to deal with the different clinical presentations of these patients in the best possible way.

In this study we set out to collect the epidemiological, familial, phenotypic (clinical) and molecular data of patients with aortic disease of different etiologies in Chile. This allowed us to identify patients with this disease of genetic origin as ICTDs that served as a platform to identify possible genetic modifiers in the clinical course of Chilean patients with ICTDs that generated genetic and molecular evidence that serves as the basis for implementing individualized treatment options for each patient.

2.2 Methodology

In this study we collected information on a significant number of patients with aortic disease (n = 75) in a database that covers all the relevant clinical aspects necessary to be able to classify these patients according to the different etiologies and that allowed the subsequent stratification into extreme phenotypes informative for the discovery of genetic modifiers of clinical severity in ICTDs.

This database contains epidemiological, imaging, and clinical information from each of the specialists involved in patient care (clinical geneticists, cardiac surgeons, ophthalmologists, and others), as well as family history and other relevant information. We also have information obtained from a survey given to the participants of this study where they provide information, for example, on health perception (Figure 4).

For the creation of our forms, and in observance of Chilean-specific factors, we adapted the forms used by the National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC), which is a longitudinal observational cohort study enrolling patients with conditions relating to genetically induced thoracic aortic aneurysms (TAAs) from 5 regional clinical centers in the United States[68,69].

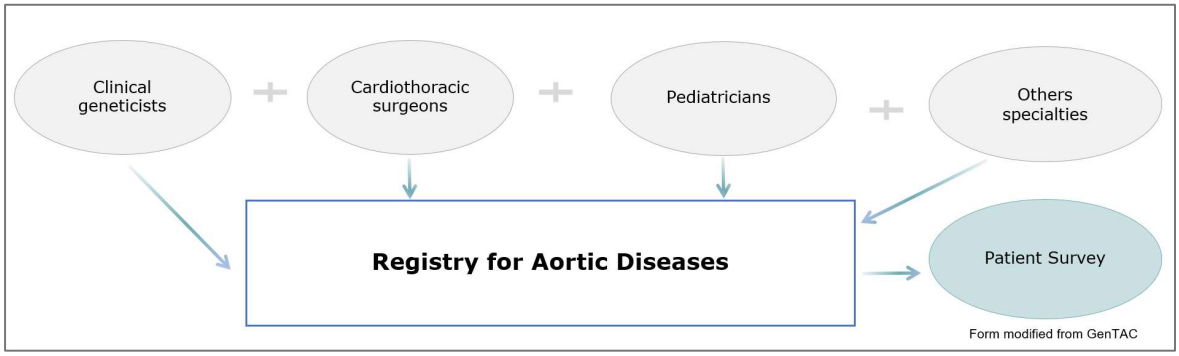


Figure 4. Workflow for Chilean registry for aortic disease

2.2.1 Inclusion criteria

Patients of all ages with a clinical diagnosis of aortic disease were invited to participate in the study. Thanks to the collaboration of clinical geneticists, the surgery service of Hospital del Tórax and Clínica Alemana, we were able to recruit patients with aortic disease of different origins and access the clinical,

epidemiological, and molecular information necessary to carry out the survey of the data that made the construction of this registry possible.

The clinical research protocol was approved by the Scientific Ethics Committee from Servicio de Salud Metropolitano Oriente (Adultos) in the case of Instituto Nacional del Tórax, and the Clinica Alemana Universidad del Desarrollo Scientific Ethics Committee for those patients recruited at Clinica Alemana de Santiago. All patients read and understood the informed consent of the study, and this was signed by each of them to be included in the study. The informed consent included information in relation to the protection of the information provided by them, the handling of the sequencing data, among others.

From this registry, Chilean patients of all ages with a clinical diagnosis of ICTDs were selected. To distinguish different presentations of the cardiovascular phenotype, we classified these patients as mild and severe, according to various criteria, which allowed us to explain the hypothesis of variants in the modifying genes that mitigate or aggravate the severity of the disease. Finally, it was possible to analyze 5 families with affected members, as well as informative healthy relatives and 5 sporadic cases with MFS.

2.3 Results

For this study, we recruited patients of all ages with a clinical diagnosis of aortic disease of different etiologies. Table 2 shows some demographic and clinical characteristics of the recruited patients. Most of the patients were male (60.8%),

with an average age of 44.2 years and a high percentage (78.4%) of these patients already required aortic surgery.

Table 2. Demographic and clinical characteristics of patients wit aortic disease

Characteristics	Patients (n=75)	%
Sex	Male= 46	60.8
	Female= 29	39.2
Age (years)	Mean= 44.22 ± 20.44	
	Range= 4-89	
Aortic surgery	Yes=59	78.4
	No=16	21.6

In our database we registered 35 patients with aortic disease of genetic origin, 28 of non-genetic origin and 12 of congenital origin. A greater number of males versus female patients can be observed for the three groups, so there are no significant differences in the sex distribution in the different etiologies (p = 0.7415) (table3).

Table 3. Classification of aortic disease according to etiology

	Genetic (n=35)	Non-genetic (n=28)	Congenital (n=12)	p-value
Sex	M= 20 F=15	M=17 F=11	M=8 F=4	0.7415*
Age (years)	28.24 ± 15.27	61.14±14.77	51.00 ± 6.47	Gen/non-gen= < $1x10^{-5}$ # Gen/cong= < $1x10^{-5}$ # Non_gen/cong=$9.88x10^{-3}$ #
Age_First_Cx_Ao	33.89± 13.28	61.14±14.77	51.00 ± 6.47	Gen/non-gen=< $1x10^{-5}$ # Gen/cong= 0.0006.4 x 10^{-4} # Non_gen/cong=$9.88x10^{-3}$ #

Mann-Whitney Test

*Fisher exact test

The specific demographic information of the patients with ICTDs who participated in this study is shown in Table 4. We obtained the information shown in this table from a survey answered by the patients in which we consider characteristics such as ethnic perception, education level, socioeconomic situation and the patients through this survey could express their perception of health in a complicated scenario such as suffering from these diseases.

The results showed that the majority (52.1%) responded that they were Latinos of Chilean descent. In addition, 73.9% of those surveyed had FONASA as health coverage and only 26.1% answered that they had an ISAPRE as health insurance, these data reflect what is observed in the general population in Chile, with 78% of the population having FONASA health coverage and only 22% having an Isapre[70].

While the level of education was well varied, most patients had secondary education. This is consistent with the fact these patients are mostly young people with a mean age of (28.24 ± 15.27) and therefore it could be expected that the majority were concentrated at a medium level of education.

Approximately 60% reported that their family group earned less than 1,100,000 Chilean pesos per month. Regarding the habit of smoking and consuming alcohol, many of the participants refrained from answering this question (43% in both cases), however, the answer obtained was mostly positive for both cases in those who answered.

A majority of participants expressed having good or fair health status and most of them maintained their health status equal to or better than the previous year (70%). 65% of surveyed individuals declared themselves able to perform domestic activities, climbing stairs, walking in some cases up to 1 km and doing daily activities such as bathing alone and 78.3% express to perceive pain sporadically without interfering with daily activities in 52.2% and with little interference in 26.1% of cases.

Table 4 Demographic and clinical characteristics for patients with ICTDs

Characteristics		N (%)
Ethnic perception	White	5 (21,7)
	White with European ancestry	1 (4,3)
	Latino not Chilean	1 (4,3)
	Latin with Chilean ancestry	12 (52,1)
	Latin with African ancestry	0 (0)
	Chilean native	8 (34,8)
Health insurance	Fonasa (public, government-issued)	17 (73,9)
	Isapre (private)	6 (26,1)
Level of education	Primary education	4 (17,4)
	Secondary education	10 (43,5)
	University	5 (21,7)
	Postgraduate	3 (13)
	Does not apply	1 (4,3)
Monthly monetary income of the family group (Chilean pesos)	Less than 450000	6 (26,1)
	Between 450000 and 1100000	7(30,4)
	More of 1100000	10 (43,5)
Smoking	Yes	6 (26,1)
	No	7 (30,4)
	Unanswered	10 (26,1)
Alcohol	Yes	9 (39,1)
	No	4 (17,4)
	Unanswered	10 (43,5)
Health perception	Bad	1 (4,3)
	Regular	7 (30,4)
	Good	10 (26,1)
	Very good	4 (17,4)
	Excellent	1 (4,3)
Current health status compared to the previous year	Equal	13 (56,5)
	Better	10 (43,5)
	Worse	0 (0)
Physical activity	Strong	6 (26,1)
	Domestic	20 (86,9)
	Climb ladder	15 (65,2)
	Walk more than 1 Km	15 (65,2)
	Walk more than 1 block	21 (91,3)
	Bathing without help	23 (100)
Pain perception	Constant and moderate	2 (8,7)
	Severe	1 (4,3)
	sporadic	18 (78,3)
	Any	2 (8,7)
Pain interference with daily activities	Less	6 (26,1)
	Moderate	4 (17,4)
	Significative	1 (4,3)
	Any	12 (52,2)

For the identification of patients with extreme phenotype, we first focused on the most life-threatening aspect of MFS: aortic aneurysm and dissection. Patients were stratified according to the criteria published in the revised version of the Ghent criteria for MFS[71]. Since these guidelines define the phenotypes of these diseases, we modify the interpretation of these criteria to generate a stratification, and this can be seen in Table 5.

Patients were classified as having severe (n=8) and mild phenotype (n=14) according to the age of presentation of the first cardiovascular manifestation and/or accident related to the aorta (table 5).

Table 5. Criteria for the stratification and classification of patients with MFS according to the severity of the aortic phenotype.

Family or sporadic patients	Patients	Mild Phenotype			Severe Phenotype			Classification
		AoR* diameter with Z ≤2 in adults, no major cardiovascular event	AoR diameter with Z ≤1 in pediatric patients, no major cardiovascular event	Age at first major cardiovascular event ≥30 years old	AoR diameter with Z ≥3 in adults	AoR diameter with Z ≥2 in pediatric patients	One or more major cardiovascular event before 18 years old	
Family 1	CAS-01-001				X			Severe
	CAS-01-002	X						Mild
	CAS-01-003	X						Mild
	CAS-01-004							Unaffected
	CAS-01-005				X			Severe
	CAS-01-006							Unaffected
	CAS-01-007	X						Mild
Family 2	CAS-01-019			X				Mild
	CAS-01-020		X					Mild
	CAS-01-021		X					Mild
	CAS-01-022		X					Mild
	CAS-01-023			X				Mild
Family 3	CAS-01-024			X				Mild
	CAS-01-025							Unaffected
	CAS-01-026					X		Severe
Family 4	CAS-01-035	X						Mild
	CAS-01-036		X					Mild
Family 5	CAS-01-046			X				Mild
	CAS-01-048		X					Mild
	CAS-01-045						X	Severe
Sporadic 1	CAS-01-016					X		Severe
Sporadic 2	CAS-01-027				X			Severe
Sporadic 3	CAS-01-031						X	Severe
Sporadic 4	CAS-01-043						X	Severe
Sporadic 5	CAS-01-044	X						Mild

Mild and severe phenotypes were defined based on current surgical criteria for MFS. AoR = Aortic root.

2.4 Discussion

Disease registries provide valuable information for research. A registry should be designed so that all relevant information is included to give an idea of the natural evolution of the disease, through the information provided by patients and/or relatives, clinical information, among others, that helps to identify patients who are interested in participating in research studies.

Having a good registry allows us to design a good research study that provides answers to patients with the disease, that allows us to better understand the clinical and molecular basis of the disease and that eventually allows us to design new treatments that improve the quality of life of these patients. The success of this depends on the construction of collaborative relationships between patients as a fundamental axis, researchers, clinical specialists, industry and government agencies. These collaborative relationships could allow taking a step forward in the investigation of diseases such as aortic disease and in particular those ICTD and finally the results of these. The studies have a direct impact on the creation of public policies that protect patients[72–74].

We were able to create a registry of 75 patients with aortic disease. The majority (60.8%) of these were men (Table 2), similar results have been seen in other studies with a marked increase in frequency in male patients with AAA [75]. Similar results were found in patients with AVB and thoracic aortic disease[16,76].

We also observed that the distribution by sex is still skewed towards male patients. However, the age of patients at the time of recruitment (already diagnosed with the

disease) varies enormously according to the origin of the disease. Thus, patients with aortic disease of genetic origin, showed significantly lower ages with respect to non-genetic and congenital origins ($p < 0.00001$). This result is consistent with previously reported data, where it is shown that hypertension (non-genetic origin) is more prevalent in older people, while Marfan syndrome (genetic origin) is observed predominantly in younger patients[77]. These results are also reflected in the age of the first surgery of these patients, which shows a significantly lower age in patients of genetic origin with respect to the other two origins. It is expected that patients who present with the disease earlier will also require aortic surgery earlier.

From the data of patients with aortic disease, we were able to separate the patients of genetic origin and identify the patients with ICTDs, we recruited patients with a clinical diagnosis of MFS in agreement with the Ghent criteria for MFS. Based on these, we stratified all MFS patients in “severe” or “mild” (table 5) depending on the age at which they underwent their first surgical repair of the aorta or the age at which they presented their first event of aortic dissection. With this approach, we sought to differentiate patients according to the point in time at which their aortic wall failed to sustain a normal architecture as a measure of how severe perturbations in the underlying molecular mechanisms present in each of our study participants.

3. IDENTIFICATION OF GENETIC VARIANTS IN PATIENTS WITH INHERITABLE CONNECTIVE TISSUE DISORDERS (ICTDS) THROUGH EXOME SEQUENCING (ES).

3.1 Introduction

Exome sequencing (ES) has become the most widely used strategy to study the genetic basis of disease both in large cohorts of patients with common diseases and in rare disease studies with larger effect sizes and greater potential to influence diseases.[78] The rapid development in DNA sequencing technologies has dramatically reduced the cost of acquiring genetic data. ES makes it possible to interrogate the entire coding portion of the human genome at low cost. To date, ES has been successfully used in many studies and has allowed answering various research questions, it has been possible to identify variants associated with complex diseases; ultra-rare variants have been identified in epileptic patients; Furthermore, ES has shown to be an effective tool for the genetic diagnosis of mitochondrial diseases and has also allowed the identification of candidate genes that modify the phenotype of genetic diseases[79–82]. ES is extremely valuable in the causal variant identification process, because, although it captures only approximately 1% of the entire genome[83], it is estimated that 85% of the variants that cause the genetic disease would be located in this portion of the genome[84,85]. ES is a cost-effective strategy focused on interpreting the information-dense genomic data set in a sample, and the costs involved in both sequencing and storage are much lower when is compared with whole genome sequencing (WGS)[86].

Analyzing the data obtained from ES is a challenge because sequencing platforms generate a large number of short reads that could generate errors in base calling as well as errors in determining the position of each read in the reference genome. Therefore, to mitigate these errors, it is necessary that each position in the reference genome is repeated several times[87]. To counteract all this, bioinformatics algorithms have been developed aimed at finding the best balance between data loss, errors, and analysis time. We implemented a pipeline that, in general terms, first evaluates the quality of reads to exclude low-quality reads, then aligns sequences against the reference genome, then variant calling and filtering process.

In this work we set out to identify genetic variants in patients with ICTDs and their informative relatives through the current best genetic approaches for the discovery of phenotype modifying genes (ES). The variants obtained serve as input for performing a linkage analysis that finally allowed us to identify these candidate genes.

3.2 Method

For this study we recruited patients with aortic disease, from which we selected patients with ICTDs. We were only able to recruit patients with MFS, so all patients had pathogenic variants in the *FNB1* gene. We analyzed the exomes of individuals belonging to 5 families and 5 sporadic patients with this disease. Patients who were classified as having severe (n = 7) and mild (n = 14) phenotypes and their informative relatives (unaffected relatives contributed information to linkage

analyses) were analyzed by exome sequencing. The pedigrees of each of the selected individuals are described below:

Family 1: figure 5 shows the pedigree of family 1. For this study we obtained exomes from patients II-1, II-3, III-6 classified as mild phenotype; patients II-1 and III-8 classified with severe phenotype and informative unaffected relatives II-2 and II-4**.** This is a very interesting family because it has several individuals with a severe phenotype and several with a mild phenotype, which allowed us to do a linkage analysis that could give us a good approach to the search for phenotype modifying genes. through the identification of gene variants in individuals classified as severe but that these variants are not present in individuals classified as mild and some cases in their informative relatives.

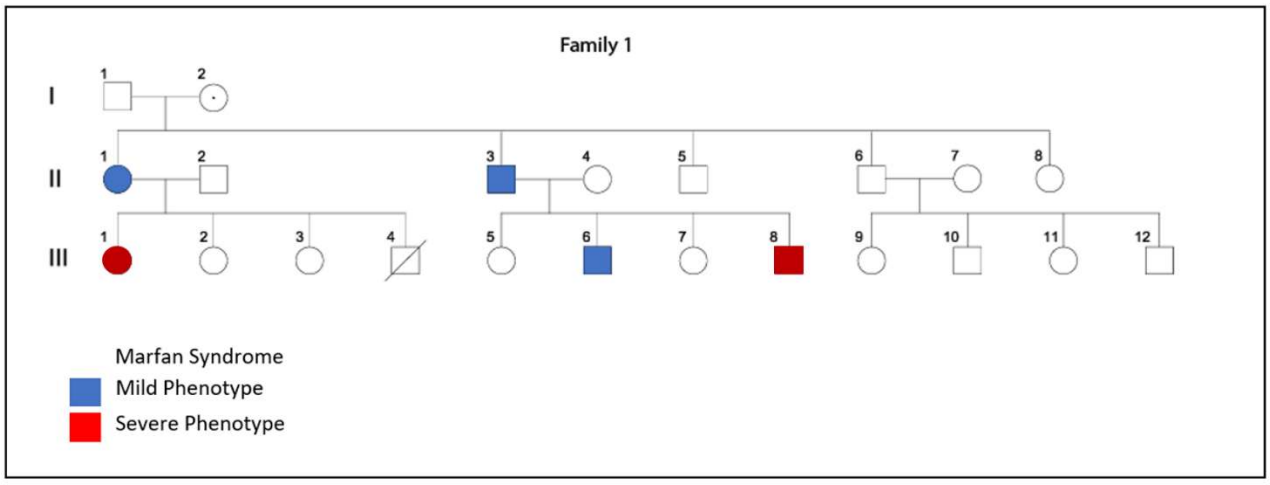


Figure 5. pedigree of family 1 with MFS. Individuals colored in blue are mild patients and those colored in red are severe

- **Family 2:** for family 2, we were able to sequence patients II-2, II-3, III-1, III-2 and III-3, all affected by MFS and classified as having mild cardiovascular phenotype (Figure 6). In this family we did not identify patients with severe phenotype and so were unable to perform an analysis within the family. Nonetheless, we were able to make comparisons of the exomes of these individuals with severe patients belonging to other families.

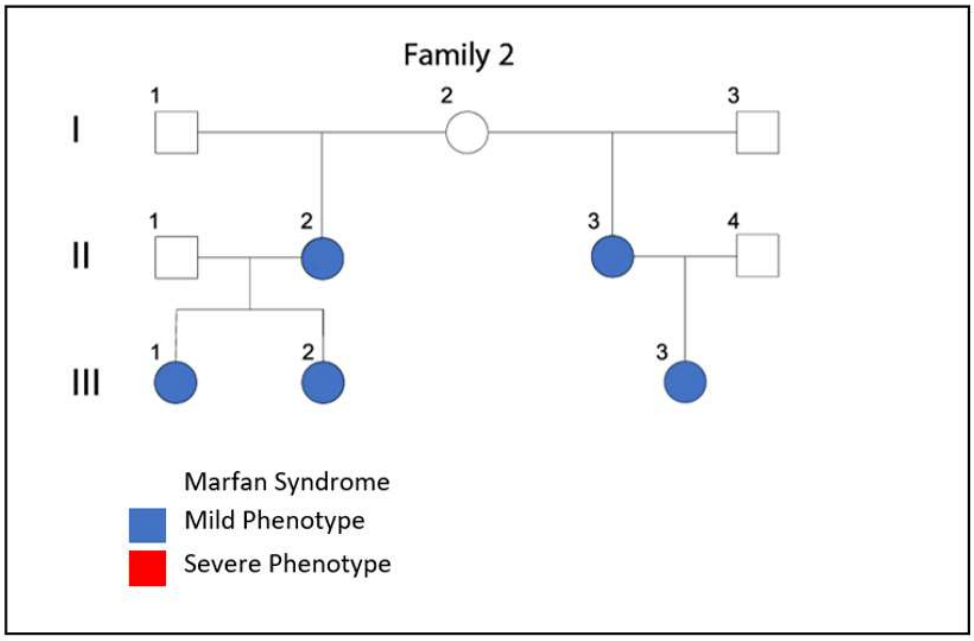


Figure 6. pedigree of family 2 with MFS. Individuals colored in blue are mild patients

- **Family 3:** as shown in figure 7, family 3 consisted of a male individual with a severe phenotype for MFS, his mother also affected with a mild phenotype, and we were also able to analyze the exome of the unaffected individual I-1

(father of II -1). As we had the exomes of mild and severe patients, we were able to do a linkage analysis for this family where we incorporated information from the trio constituted by patient II-1 and their parents.

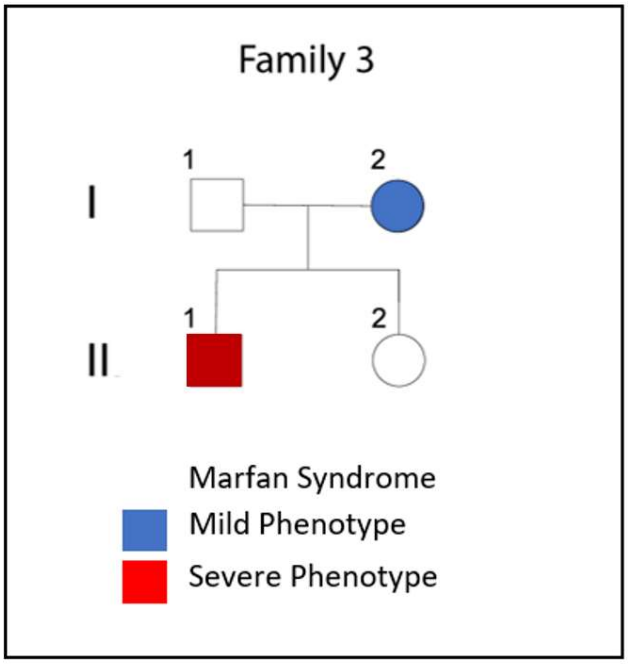


Figure 7. Pedigree of family 3 with MFS. Individuals colored in blue are mild patients and those colored in red are severe

- **Family 4:** In figure 8 the pedigree of family four is shown, we only have affected individuals I-1 and II-1, both with mild phenotype. we were able to compare the exomes of these patients (both mild) with severe phenotype patients belonging to other families

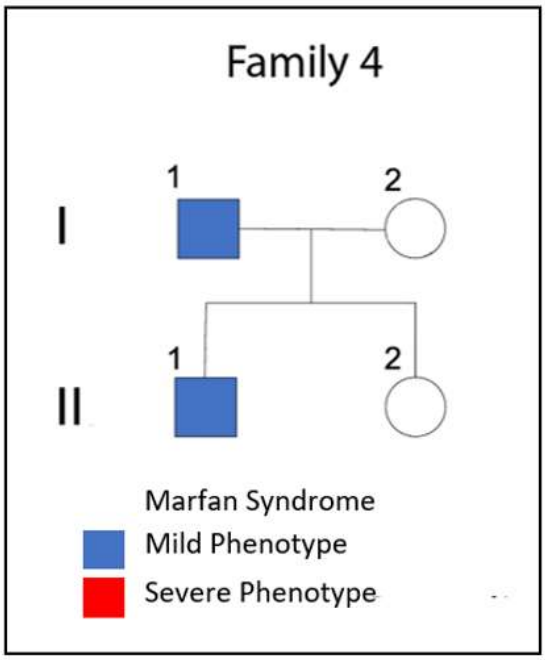


Figure 8. pedigree of family 4 with MFS. Individuals colored in blue are mild patients

Family 5: Although in family 5 there are 5 individuals with a clinical diagnosis of MFS, we were able to obtain exome data from only three of them. We were able to do an analysis in which we incorporated information from these 3 patients (Figure 9).

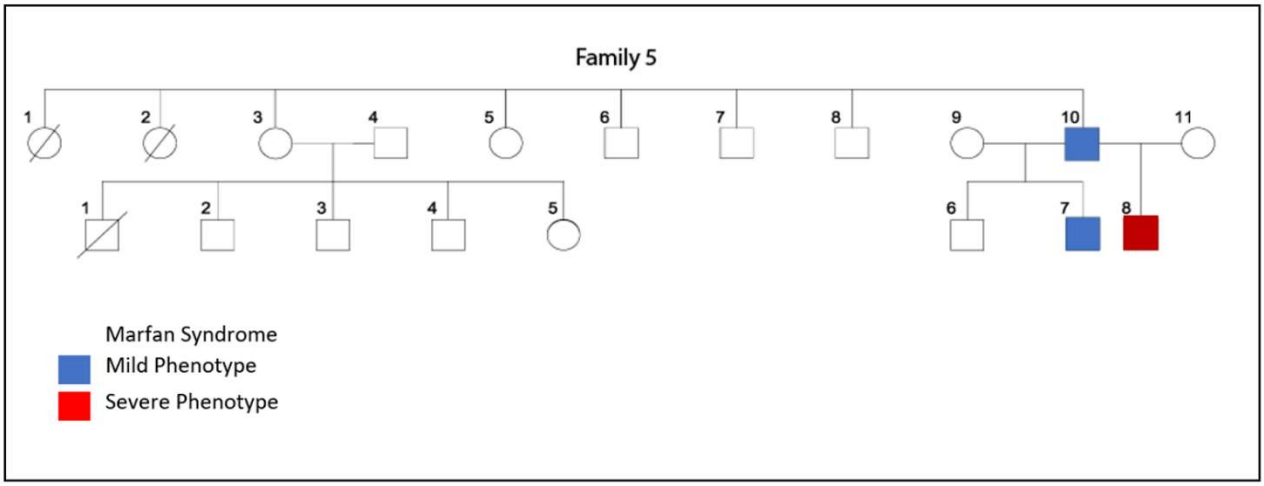


Figure 9. pedigree of family 5 with MFS. Individuals colored in blue are mild patients and those colored in red are severe.

Sporadic Patients: For our study, we had 5 patients with sporadic MFS, 4 of them classified as a severe phenotype and one with a mild cardiovascular phenotype (Figure 10). Exome data from these patients were used in an analysis where we compared all severe patients vs. all patients classified as mild.

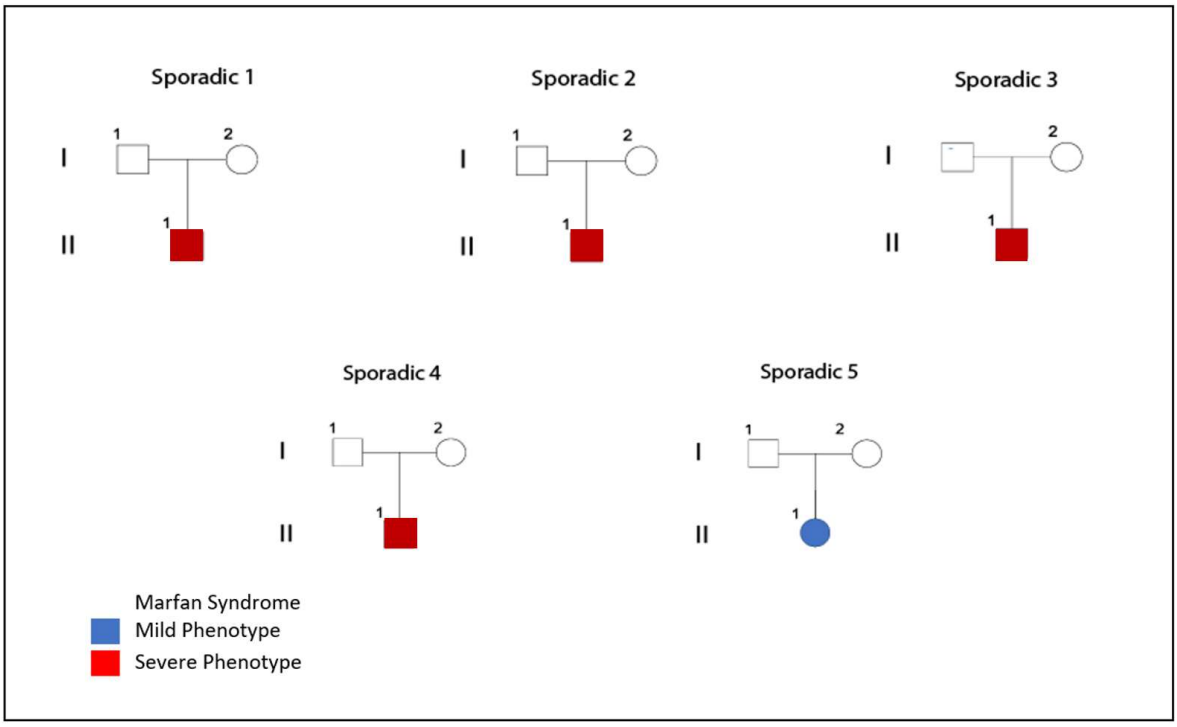


Figure 10. pedigree of sporadic patients with MFS. Individuals colored in blue are mild patients and those colored in red are severe.

3.2.1 Exome sequencing

ES was performed at BGI using their proprietary DNBseq™ NGS Platform, and Agilent V6 + UTR library kit with depth of coverage 100X. All sequencing reads

were aligned to the hg19 build of the human genome. The bioinformatics pipeline for data analysis is described below.

3.2.2 Bioinformatic pipeline for next generation sequencing data

Exome sequencing data from the families and sporadic patients described above were run through a bioinformatics pipeline. The scripts and command line options used to perform various steps of the data analysis are described in appendix, but the general steps were the following (Figure 11).

3.2.2.1 Quality control

The sequenced data generated is stored in a FASTQ file, a text-based format for storing nucleotide sequences and their corresponding quality scores. The certainty of each base call is recorded as a 'Phred' quality score, which measures the probability that a base is called incorrectly[88].

The quality score of a given base is given by the Q value, an estimate of the probability of a base call being wrong. A Q30 value, for example, corresponds to a 1 in 1000 probability of error. In this study, we take Q30 as our threshold for high-quality sequencing. Raw reads were analyzed using FastQC[89]. FastQC runs a series of tests on a FASTQ file to generate a comprehensive quality control report. FastQC assesses data quality by evaluating read length, per base quality score, per sequence quality scores, GC content, nucleotide content, sequence duplication and overrepresented sequences. We run the program MultiQC using all results from FastQC to summarize quality results for multiple sequences and display them in a single report[90].

3.2.2.2 Alignment/Mapping

One of the most important steps in NGS analysis pipelines is the mapping of read sequences to the reference genome. Because of the large volume of reads and the huge size of the whole reference genome, alignment algorithms have been optimized for speed and memory usage. Raw sequences were aligned to the human reference genome (version hg19, UCSC) using Burrows-Wheeler Aligner (BWA)[91],a software package for mapping low-divergent sequences against a large reference genome, such as the human genome and outputs the final alignment in the SAM (Sequence Alignment/Map) format. We used BWA-MEM that is recommended for high-quality queries as it is faster and more accurate.

3.2.2.3 PCR duplicate removal (Samtools and Picard)

Samtools[92] is a suite of tools designed primarily to manipulate SAM files, and the binary transformed BAM files, in preparation for downstream analysis that are tab-delimited text files that contain a header section, which carries information on the project and the genome, and an alignment section, that contain alignment information. For converting SAM files to BAM and for sorting and indexing the alignment files, we used Samtools. One technical artifact of capture-sequencing procedures is the generation of duplicate DNA sequencing reads (defined as reads with the same start point and direction) that are due to PCR-induced duplication. These share the same sequence and have the same alignment position. It is essential to identify, and 'mark' these duplicate reads so that they do not influence variant calling. PCR artifacts were removed with Picardtools[93], which removes all read pairs with identical coordinates, only retaining the pair with the highest

mapping quality and also generates a file that contains information on the percentage of PCR duplicates found in the original aligned file.

3.2.2.4 Base quality score recalibration

Base quality score recalibration (BQSR) was done using BaseRecalibrator and following the best practices of GATK. BQSR stands for Base Quality Score Recalibration. In a nutshell, it is a data pre-processing step that detects systematic errors made by the sequencing machine when it estimates the accuracy of each base call and corrects them. BQSR is a process in which is applied machine learning to model these errors empirically and adjust the quality scores accordingly[94].

3.2.2.5 Realignment around indels

Some artifacts may arise due to the alignment stage, especially around indels where reads covering the start or the end of an indel are often incorrectly mapped. This results in mismatches between the reference and reads near the misalignment region, which can easily be mistaken for SNPs. Thus, the realignment stage aims to correct these artifacts by transforming those regions with misalignment due to indels into reads with a consensus indel for correct variant calling. In our analysis, Realignment was accomplished using the GATK IndelRealigner[95].

3.2.2.6 Variant Calling with HaplotypeCaller

Using GATK version 4.2, we use HaplotypeCaller, a program capable of calling SNPs and indels simultaneously through local de novo assembly of haplotypes in an active region. In other words, each time the program finds a region that shows signs of variation (it defines active regions), it discards the existing mapping

information and completely reassembles the reads in that region and identifies which possible haplotypes are present in the data. The program then realigns each haplotype against the reference haplotype using the Smith-Waterman algorithm to identify potentially variant sites. Then determine the haplotype probabilities given the read data and obtain the allele probabilities per read for each potentially variant site. The most likely genotype is then assigned to the sample. We used HaplotypeCaller to be more accurate when calling regions that are traditionally difficult to call, for example when they contain different types of variants close to each other. It also makes the HaplotypeCaller much better at calling indels than position-based callers like UnifiedGenotyper[96].

3.2.2.7 Variant Quality Score Recalibration (VQSR)

VQSR is a sophisticated filtering technique applied on the variant callset that uses machine learning to model the technical profile of variants in a training set and uses that to filter out probable artifacts from the callset. The VQSR method, uses machine learning algorithms to learn from each dataset what is the annotation profile of high-quality and low-quality variants, and does so in a way that integrates information from multiple dimensions. When VQSR is used, we can develop a continuous, covarying estimate of the relationship between variant call annotations (QD, MQ, FS etc.) and the probability that a variant call is a true genetic variant versus a sequencing or data processing artifact. This model is determined adaptively based on "true sites" provided as input from different databases. We can then apply this adaptive error model to both known and novel variation discovered in the call set of interest to evaluate the probability that each call is real.

The VQSLOD score generated, is the log odds ratio of being a true variant versus being false under the trained Gaussian mixture model [97].

3.2.2.8 Annotation

For this study we recruited patients with a clinical diagnosis of some inheritable connective tissue disease (ICTDs) and most of them did not have a molecular diagnosis for the disease. We therefore decided to annotate their exomes to identify the causal variants of the disease. To do this we used the VCF filtered from VQSR and performed annotation using ANNOVAR[98] , to annotate single nucleotide variants (SNVs) and insertions/deletions, such as examining their functional consequence on genes . It provides a wide variety of different annotation techniques, organized in the categories gene-based, region-based and filter-based annotation where are identified variants that are documented in specific databases, for example, whether a variant is reported in dbSNP, what is the allele frequency in the 1000 Genome Project, ExAC or gnomAD[98]. ANNOVAR uses six different scores to estimate deleterious impact: GERP++[99], LRT[100], MutationTaster[101], PolyPhen[102], PhyloP conservation[103], and SIFT[104]. Finally, we can identify a list of candidate genes and variants for diseases from exome data.

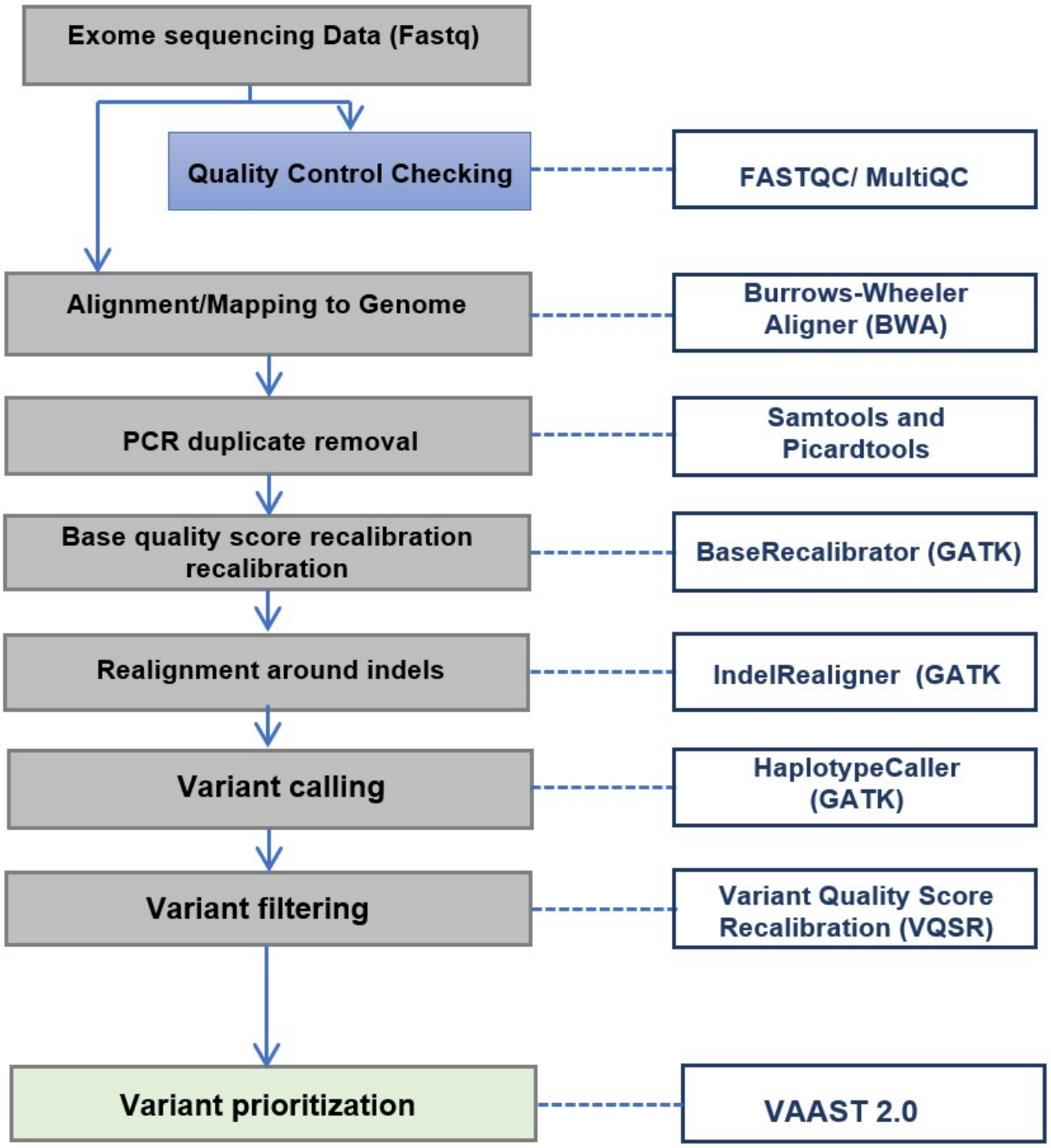


Figure 11. Analytical pipeline for ES. Left panel indicates program used for each step. Right panel indicates analysis steps

3.3 Results

One of the best genetic approaches for the discovery of phenotype modifying genes is Exome Sequencing (ES). In this study, we analyzed 5 families and 5 sporadic cases with MFS. For exome analysis we follow GATK best practices (figure 11).

In the first step, we performed the quality control of the unmapped reads (fastq) using FastQC. The outputs for all the exomes were then integrated into a single MultiQC report. Figure 12 shows a plot of the mean quality at each base position according to the phred scale and shows how the mean base quality of the exome sequences for all patients were of high quality, with values exceeding Q30 for all cases, meaning that there is a 1 in 1000 probability that the identified bases are incorrect.

Figure 12. FastQC per base sequence quality plot

However, in the multiQC graph that gives information regarding the Sequence Duplication Levels (Figure 13), we also observed that one of the sequences (CAS-01-045) has a high percentage of duplicate reads. As one of the steps in the ES analysis pipeline is to remove duplicate reads, for this sample we lost a high percentage of the exome information, however, we decided to continue with this patient for the following analyzes.

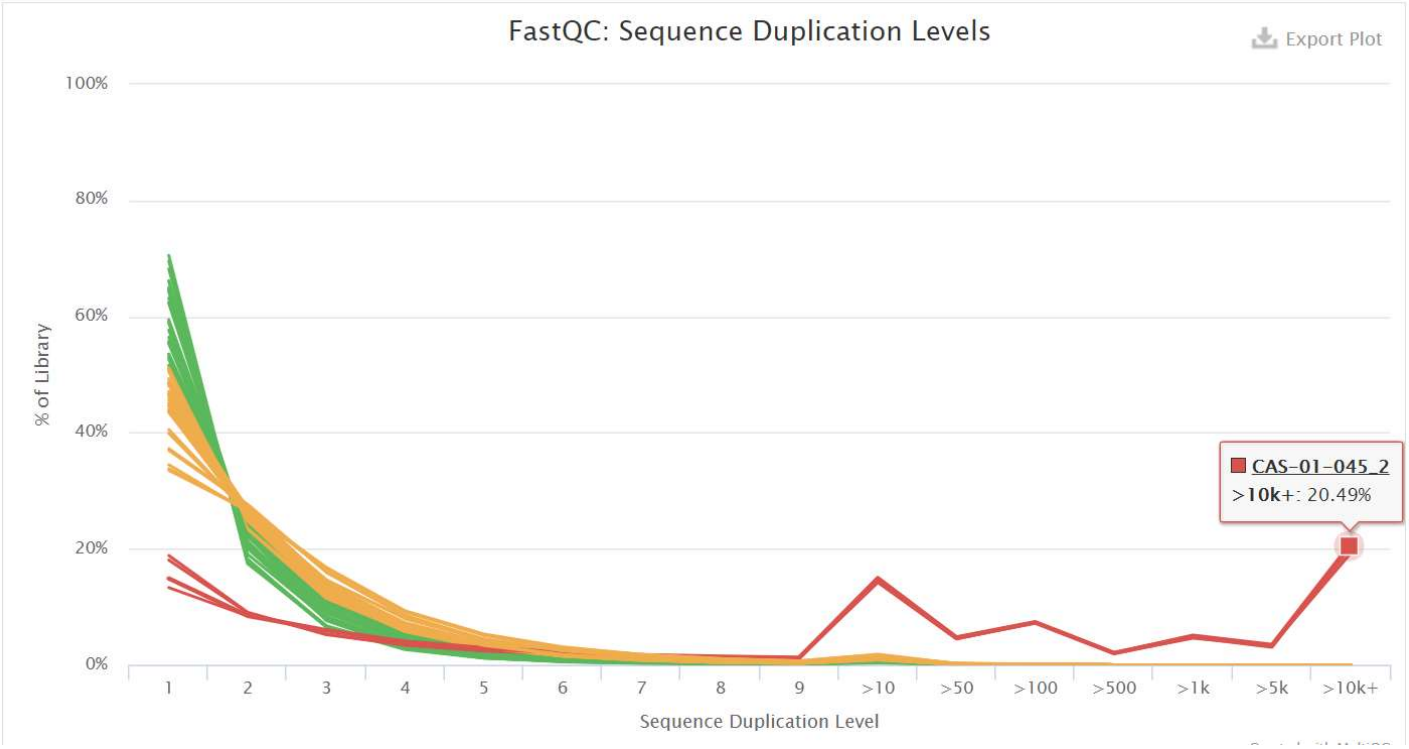


Figure 13. Sequence Duplication Levels

A very important step in the exome analysis pipeline is Variant Filtering to eliminate possible false positives that arise during Variant Discovery. GATK´s best practices

include a variant filtering step following Variant Quality Score Recalibration (VQSR). We use this "VQSR filter" that uses metrics, such as quality by depth, mapping quality, variant position, and others, from "true" variants (variants found in databases) and we generate an adaptive error model. We then apply this model to the remaining variants to calculate a probability that each variant is real. With this quality score recalibrated, we were able to filter out lower quality variants. We use a threshold that maintains a 99% sensitivity for "true" variants as recommended by GATK.

As a result, we obtained a VQSR-filtered list of variants from Haplotype Caller. Table 6 shows the number of variants obtained for each of the exomes analyzed before and after applying VQSR both for Indels and SNPs. On average, the percentage of variants removed is 13.46% for all exomes. This list is the one we used in our annotation; using ANNOVAR, and variant prioritization analysis steps discussed in the next chapter.

Table 6. Number of variants before and after of applying VQSR filtering

Patients	Number of variants unfiltered	Number of variants filtered with VQSR	Variants removed (%)
CAS-01-001	393141	330546	15,92
CAS-01-002	1366602	1229279	10,04
CAS-01-003	527945	453247	14,14
CAS-01-004	1167897	1044079	10,60
CAS-01-005	674341	585178	13,22
CAS-01-006	752165	647570	13,90
CAS-01-007	661199	568851	13,96
CAS-01-019	717899	619636	13,68
CAS-01-020	696335	592740	14,87
CAS-01-021	684503	585145	14,51
CAS-01-022	584069	493337	15,53
CAS-01-023	838723	730186	12,94
CAS-01-024	620166	523673	15,55
CAS-01-025	756476	653281	13,64
CAS-01-026	644577	549182	14,79
CAS-01-035	734369	633621	13,71
CAS-01-036	1117004	993816	11,02
CAS-01-045	201361	180839	10,19
CAS-01-046	908408	793064	12,69
CAS-01-048	827777	720239	12,99
CAS-01-016	1573815	1416103	10,02
CAS-01-027	548939	463934	15,48
CAS-01-031	624768	528365	15,43
CAS-01-041	618715	526998	14,82
CAS-01-043	956846	838384	12,38
CAS-01-044	638883	741566	13,84

We used ANNOVAR to identify the causal variant in all patients with ICTDs. In this study we were only able to recruit patients with MFS, so all patients had pathogenic variants in the *FNB1* gene. To perform variant prioritization using ANNOVAR, which estimates different scores (GERP ++, LRT, MutationTaster, PolyPhen, PhyloP conservation, and SIFT). Thus, Table 7 shows the variants found in *FBN1* for each of the patients with a clinical diagnosis of MFS analyzed in this study. These variants are classified as pathogenic or likely pathogenic, and are found in the following functional domains: TB domain (they act in specific protein-protein interactions, important for the regulatory functions associated with the fibrillin/LTBP family of proteins)[105], EGF-like calcium-binding domain (which stabilizes the interaction between the domains of the molecule, which can facilitate protein-protein interactions)[106]; and in N-terminal domain (interacts with the C-terminal domain in an important step in microfibril formation)[107].

Table 7. Causal variants of MFS in affected patients

Family or sporadic cases	Patients	FBN1 Variants HGVS nomenclature	Variant type	Clinical significance Sift	Functional domain	Cardiovascular phenotype classification
Fam 1	CAS-01-001	c.7339G>A	Missense	Pathogenic	EGF-like calcium-binding domain	Severe
	CAS-01-002					Mild
	CAS-01-003					Mild
	CAS-01-007					Mild
	CAS-01-005					Severe
Fam 2	CAS-01-019	c.1090C>T	Nonsense (stop gained)	Pathogenic	TB domain	Mild
	CAS-01-020					Mild
	CAS-01-021					Mild
	CAS-01-022					Mild
	CAS-01-023					Mild
Fam 3	CAS-01-024	c.7180C>T	Nonsense (stop gained)	Pathogenic	-	Mild
	CAS-01-026					Severe
Fam 4	CAS-01-035	c.7204+1G>A	splice region	Likely_pathogenic	-	Mild
	CAS-01-036					Mild
Fam 5	CAS-01-045	c.4196_4197insA	frameshift insertion	Pathogenic	EGF-like calcium-binding domain	Severe
	CAS-01-046					Mild
	CAS-01-048					Mild
Sporadic 1	CAS-01-016	c.8562delC	frameshift deletion	Pathogenic	-	Severe
Sporadic 2	CAS-01-027	c.1090C>T	Nonsense	Pathogenic	TB domain	Severe
Sporadic 3	CAS-01-031	c.5788+5G>T	splice region	Pathogenic	-	Severe
Sporadic 4	CAS-01-043	c.A2673G	Regulatory region	Likely_pathogenic	TB domain	Severe
Sporadic 5	CAS-01-044	c.236dupA	frameshift insertion	Pathogenic	N-terminal domain	Mild

3.4 Discussion

We performed a ES study on patients with ICTDs to obtain the list of variants from patients and their informative relatives, following the indications of GATK's best practices. The quality control showed sequences of very good quality, one of the exomes had a very high percentage of duplicate sequences that were eliminated in in the PCR duplicate removal step of the pipeline.

A very important step in the pipeline used in this analysis is the filtering of false positive variants and we run this filtering through VQSR where after the training run, in which we incorporated 30 exomes from Chilean individuals we obtained a model that was applied to each of the exomed Individuals in this study to achieve the final filtering. The result of this analysis showed that approximately 87% of the variants called in the Variant Discovery step were considered true positives, similar

results were observed in another study where the effectiveness of using VQSR as a filtering strategy was analyzed and an 87.5% variant retention was observed after filtering[82]. Our final result was a list of variants on which we performed the annotation and prioritization analysis of phenotype modifying genes in ICTDs.

Regarding the annotation of the exomes, we recruited patients with ICTDs, all with a clinical diagnosis of MFS, but we did not have the molecular diagnosis of most of the recruited patients, therefore, we decided to annotate these exomes using ANNOVAR to identify the causal variant of MFS in the *FBN1* gene. Table 7 shows the list of pathogenic variants in *FBN1* in each of the affected members of the 5 families studied and in the 5 patients with sporadic MFS.

There is notable intrafamilial variability in MFS, and the clinical status of relatives who carry an identical *FBN1* variant can have variation with respect to the onset of the disease, the involvement of different systems. This suggests that the allelic variant is not the only determinant of clinical severity. On the other hand, the available evidence suggests that the existence of some degree of genotype-phenotype correlation for at least a subset of *FBN1* variants[108–110].

With the information on the variants in *FBN1* and the stratification into mild phenotype and severe phenotype that we had already carried out, we decided to analyze these data in search of some genotype-phenotype correlations in our patients. To do this, we first discarded the variants present in families with both mild and severe patients (Families 1, 3 and 5), then we focused on family 2 with 5 patients with the same phenotype (mild) whose variant could be correlated with the phenotype. However, this variant was also found in a sporadic patient with a

severe phenotype, ruling out this possibility. Finally, as we have few individuals to perform a solid correlation analysis, we performed a review of the literature in relation to these variants and their possible correlation with the phenotype, but we did not find good results in this regard.

As in this study we focused on the identification of modifier genes in ICTDs, for this we used ES from which and following the best practices of GATK, we obtained a list of good quality variants for each exome analyzed., necessary to perform the prioritization of variants associated with extreme phenotypes in the ICTD using VAAST.

4. IDENTIFICATION OF CANDIDATE GENES THAT MODIFY THE CLINICAL PRESENTATION IN PATIENTS WITH EXTREME PRESENTATIONS OF INHERITED CONNECTIVE TISSUE DISEASES.

4.1 Introduction

Clinical manifestations of ICTDs are highly variable with significant phenotypic heterogeneity between individuals possessing the same variant, including within members of the same family[45,46]. Prognosis depends on the life-threatening complications of thoracic aortic aneurysm and dissection of aorta and is the main cause of mortality in these patients. The cardiovascular clinical course of MFS differs considerably in terms of age of onset and severity, even between individuals with the same disorder and who share the same causative variant[33], suggesting the presence of other genetic variants at other loci that play a role at determining the overall phenotype of these patients[111]. In this study we were only able to recruit patients with MFS, so we will refer only to this ICTD type.

The main mechanisms by which aortic aneurysms (AAs) occur in ICTDs, specifically in MFS are largely unknown. However, there is significant evidence that aneurysm formation is a product of changes in the ECM of the aortic wall and that they would be mediated by perturbations in the transforming growth factor-β (TGFβ) signaling pathway[52]. This perturbation occurs because fibrillin-1 is a regulator of the bioavailability of TGFβ and therefore a defective fibrillin-1 can stimulate the release of sequestered TGF-beta and increase its activity[53], which leads to inflammation, fibrosis, and activation of metalloproteinases (MMP-2 and

MMP-9), leading to loss of Vascular Smooth Muscle Cells (VSMCs). These factors, together with the decrease in collagen, reduce aortic structural integrity and lead to dilation or aneurysm[54]. In fact, it has been shown that inhibition of TGF-β signaling in mouse models of MFS prevents the development of cardiovascular phenotypes, independently of the causal variant in the *FBN1* gene[55].

The identification of genetic variants associated with or causing variability in the aortic phenotype in patients with ICTDs would facilitate decision-making regarding the clinical follow-up of each patient, the optimal time to perform surgical repair of the affected blood vessels, the design of drugs aimed at slowing the progression of aortic disease (ideally to the point of avoiding the need for surgery), among others. Currently, the understanding of the effects of genetic changes on the frequency and progression of the aortic phenotype in ICTDs is very limited, so deepening its understanding is very important for this group of patients.

Therefore, in the present study, we identify genetic variants that could modify the severity of the clinical course of heart disease in patients with MFS. All this based on the analysis of linkage in families and sporadic patients with MFS.

4.2 Methods

4.2.1 Subjects and methods

Chilean patients of all ages with a clinical diagnosis for MFS were selected. In total, we managed to recruit, 5 families and 5 sporadic cases with MFS and informative

healthy relatives. For this study, it was necessary to efficiently identify the extreme cases of MFS (mild and severe), which might allow us to confirm the presence of variants in modifier genes mitigating or exacerbating the severity of the disease.

Patients were classified as having severe (n=8) and mild phenotype (n=14) according to the age of presentation of the first cardiovascular manifestation and/or accident related to the aorta.

4.2.2 VAAST Analysis

We performed variant prioritization analysis using the Variant Annotation, Analysis, and Search Tool (VAAST 2.0) package[112]. VAAST uses an extended Composite Likelihood Ratio Test (CLRT) to determine a severity score for genomic variants[113]. VAAST provides several ways to explore genomic datasets for families and individuals and it has been used successfully to identify both causal genes and modifier genes for several genetic disorders. In this regard, VAAST has made it possible to identify candidate modifier genes for different phenotypes, including immune function in 22q11.2 deletion syndrome and chronic obstructive pulmonary disease (COPD) susceptibility[114,115]. Variants were annotated for their functional impact using Variant Annotation Tool (VAT). The annotated variants in all patients with mild and severe phenotype were combined using Variant Selection Tool (VST). VAT analysis was performed in each of the families that had mild and severe cases (families 1, 3 and 5) and in a combination that includes all mild and all severe familial and non-familial cases. Phenotype/putative modifying genes and variants for MFS were analyzed using VAAST to prioritize candidate genes under both dominant and recessive modes of inheritance.

For this study we generated ES data for small family groups as well as non-familial cases Consequently, we analyzed the data in two ways: 1) Familial approach, in which we analyzed the families that had both members affected with severe phenotype and affected with mild phenotype or; 2) Case-control approach, in which we used the exome data of all patients (familial and non-familial) classified as either severe or mild phenotype and performed a case-control analysis using VAAST[112].

In the first approach (familial) the VAAST analysis was performed in families 1, 3 and 5. Because no data of any affected individual from generation I was available that could link affected patients from generation II by ancestry, we could not do linkage analysis for the entire family 1. Therefore, family 1 was partitioned in two distinct nuclei and analyzed considering mild and severe patients with the possibility of identifying alleles shared by IBD (identical by descent). We therefore analyzed subfamily 1.1 (composed by II-1, II-2 and III-1 individuals) and subfamily 1.2 (composed by II-3, II-4, III-6 and III-8 individuals).

In the case-control approach, ES data of all MFS patients classified as severe and all MFS patients classified as mild was analyzed through the comparison of severe vs mild. Because we could not estimate the effect size of the putative modifier variants/genes, big effect size under a “monogenic-like” model or a small effect size under a “polygenic-like” model[116], we performed our analyses using the “incomplete penetrance” parameter. This allowed us to identify both genes with a strong effect present in all individuals with a given shared phenotype but also genes with a lower effect that may be present only in some of the individuals with

the shared phenotype under study. In keeping with this approach, we performed the VAAST analysis assuming incomplete penetrance and under two different schemes: a) autosomal dominant mode of inheritance that allowed selecting genes that had at least one copy of the putative variant and b) autosomal recessive mode of inheritance that selected genes with two alleles that differed from the reference for any given coordinate (this could be identical alleles or compound heterozygotes). Additionally, we allowed for the presence of locus heterogeneity in order to perform an unbiased search across the entire exome.

4.2.2.1 Variant Prioritization.

We filtered for and selected genes that met any of the following criteria: 1) Gene Ontology (GO) for biological processes distinctively associated with pathophysiology of MFS, 2) using https://string-db.org/, we selected genes whose products interact with proteins that play a role in the homeostasis of the ECM or in the TGFβ signaling pathway or 3) genes previously associated with cardiovascular phenotypes as annotated in the literature. Finally, the variants of candidate genes obtained with the VAAST analysis of the families with individuals with severe and mild phenotypes were searched for in each of the non-familial mild and severe patients and those genes with variants that did not show a clear segregation with the mild or severe phenotype were discarded. To identify candidate genes from the ES dataset, we selected both rare variants with presumably large effect sizes with a deleterious or harmful impact according to the SIFT calculation[117], as well as non-deleterious variants that, in association with the causal variant of MFS (*FBN1*),

could explain changes in the function of a molecular pathway associated with the pathophysiology of MFS.

4.3 Results

We used the Variant Annotation, Analysis, and Search Tool (VAAST) program to prioritize candidate genes that could be cardiovascular phenotype modifying genes in MFS. We performed VAAST analysis assuming dominant and recessive inheritance, incomplete penetrance, and locus heterogeneity in each of the families containing individuals with both mild and severe phenotype (1, 3 and 5). Family 1 was analyzed once with all the individuals that we had sequenced and once as separate subfamilies 1.1 and 1.2. We compared all individuals with a mild phenotype with those with a severe phenotype giving us information of both candidate genes that exacerbate the cardiovascular phenotype as well as genes that could be mitigating this phenotype. Thus, Table 8 shows the number of candidate genes obtained after analysis with VAAST in these different configurations. From the VAAST output, several hundred genes had a p-value <0.05 under each analysis set-up after performing Bonferroni correction for multiple comparisons.

Table 8. Number of candidate genes under different VAAST analysis categories

Samples	Inheritance	Analysis	Candidate Genes
Fam1_all	Dominant	1	7
	Recessive	2	8
Fam 1.1	Dominant	3	37
	Recessive	4	170
Fam 1.2	Dominant	5	14
	Recessive	6	31
Fam 3	Dominant	7	49
	Recessive	8	188
Fam 5	Dominant	9	5
	Recessive	10	6
Severe-Non_Mild	Dominant	11	12
	Recessive	12	15
Mild-Non_Severe	Dominant	13	95
	Recessive	14	97

Severe-Not_Mild: analysis to identify variants present in patients with severe phenotype but not in mild. Mild-Not_Severe: analysis to identify variants present in patients with mild phenotype but not in severe.

The primary list of genes obtained from the different analyses mentioned above is large and, in some cases, composed of hundreds of genes. In order to place these

candidate genes in a biological context, several avenues were explored. Considering all the selection criteria established in this study, we selected 8 genes that met those criteria. Among these, variants in 5 genes were associated with the severe aortic phenotype and 3 with the mild aortic phenotype (Table 9).

Table 9. Candidate genes selected under different VAAST analysis categories

Approach	Samples	Mode of Inheritance	Candidate Genes
Familial	**Fam 1.1**	Dominant	*PPARD*
		Recessive	*FBN1*
	Fam 3	Dominant	*MAP3K1*
		Recessive	*JAG1*
Non-familial	**Severe vs. Mild**	Dominant	*MAP3K1, PTPRJ*
		Recessive	*PTPRJ*
	Mild vs. Severe	Dominant	*KIAA1462, TNFSF18, TGFBR3L*
		Recessive	*TNFSF18, KIAA1462*

4.4 Discussion

We conducted a ES study using an extreme phenotype design (Severe vs. mild cardiovascular phenotype) to prioritize modifier genes that could modify the age of onset of cardiovascular complications in patients with MFS. The top-ranking candidate genes from the VAAST analysis, and subsequent filtering based on GO (biological processes), association with cardiovascular phenotypes and interaction with causal genes of aortic phenotypes, show 5 candidates genes with rare frameshift and missense variants in multiple individuals with severe phenotype that

could be conferring the susceptibility of presenting some cardiovascular complication at an early age to these patients (Figure 14) and 3 candidates genes with variants that could confer any effect that mitigate the cardiovascular phenotype in patients with MFS (Figure 15).

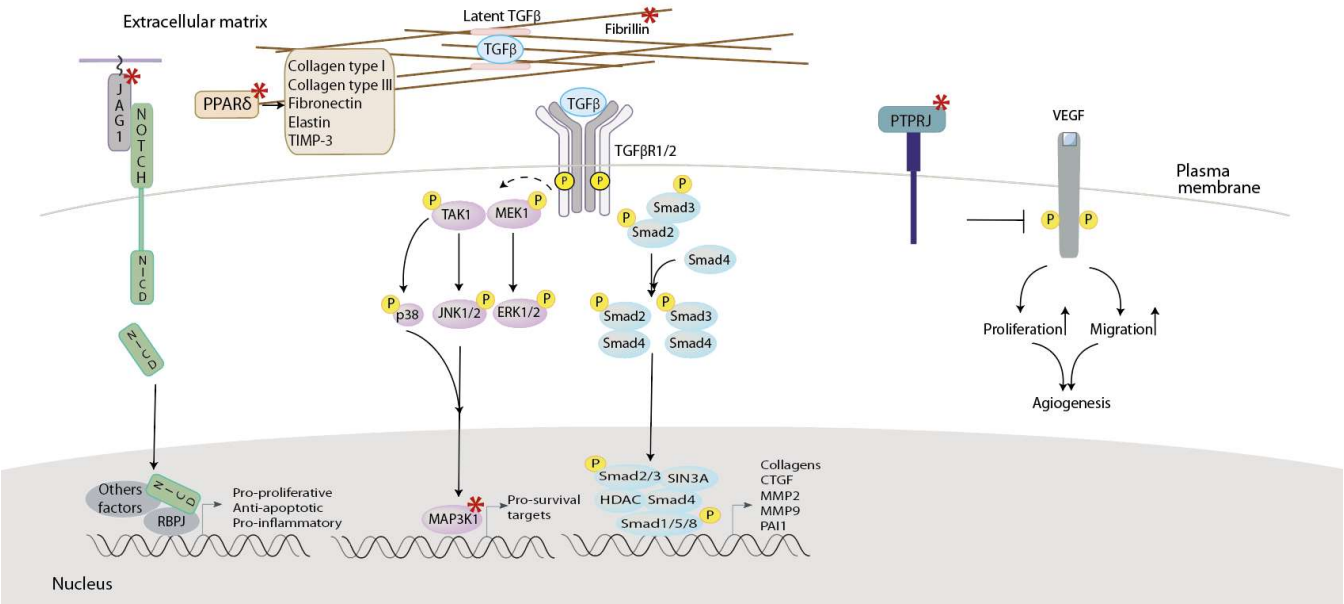


Figure 14. Genes associated with severe phenotype located in different signaling pathways. Genes associated with severe phenotype are marked with a red star.

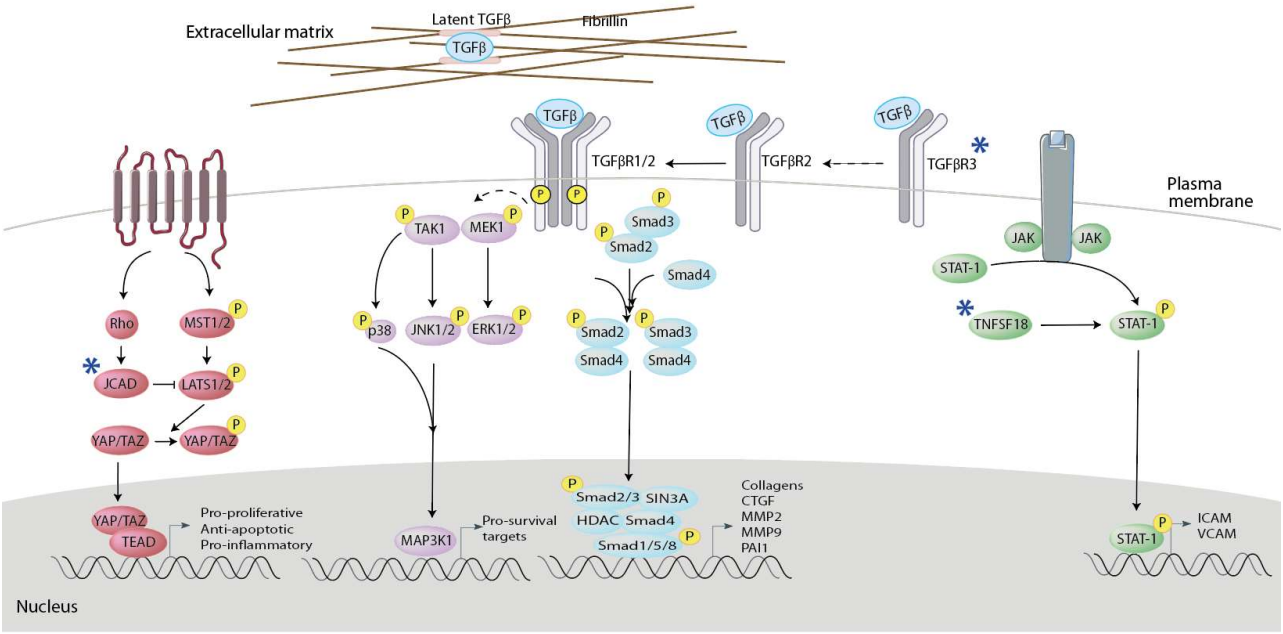


Figure 15. Genes associated with mild phenotype located in different signaling pathways. Genes associated with mild phenotype are marked with a blue star.

It has been shown that fibrillin 1 regulates the bioavailability of TGF-b and variants in this gene considerably alter intercellular communication, thus increasing the levels of TGF-β protein in the extracellular space[118]. TGF-β is a molecule that regulates several processes, including the induction of apoptosis, improves collagen production and remodeling of the extracellular matrix (ECM), among others.[13,119] Therefore, a physiological dysregulation TGF-β interferes with the remodeling and recovery of microfibrils and its effect depends, in addition to the concentration of TGF-β in the tissue, on other factors that could play a key role in tissue degradation, more specifically aortic tissue in patients with MFS. The

discovery of these other elements could be the key to explaining the difference in cardiovascular presentation in patients with MFS. In this sense, we believe that the identification of genes that could modulate active TGF-β levels should be the first to be considered in understanding the variable expressivity observed in the cardiovascular phenotype of MFS patients. Therefore, our search for possible modifier genes was aimed at identifying elements that, on the one hand, increase TGF-β activation or that interfere in the degradation process of the extracellular matrix that would favor the appearance of aortic dilation, and, on the other hand, elements that could be stopping the uncontrolled activation of TGF-β, which would finally prevent the degradation of the ECM and act as protective elements to the appearance of dilation of the aorta.

4.4.1 Candidate genes associated with severe cardiovascular phenotype in MFS

We identified 5 candidate genes associated with a severe aortic phenotype: the first is *PPARD* (Peroxisome Proliferator-Activated Receptor δ). We found a missense variant (c.140G>A) following an autosomal dominant mode of inheritance (Table 10) in exon 4 that changes the amino acid arginine into glutamine at the conserved position 47 of the protein. This change was observed in a patient with severe aortic phenotype (III1) in family 1.1, but absent in patients with mild aortic phenotype in the same family or any other patients with a mild aortic phenotype analyzed in this study. This variant is located in an intrinsically disordered protein (IDP) domain that plays important signaling and regulatory

functions and has been predicted to be involved in diseases[120,121]. It is known that *TGF-β1* is a target gene for *PPARδ* in VSMCs[122] and that *PPARδ* plays a potential role in the modulation of ECM homeostasis through the regulation of the synthesis and degradation of extracellular matrix components through the transforming growth factor-β1 and its effector Smad3[122]. To that effect, Hyo Jung Kim et al. reported that the activation of *PPARδ* increases the expression of collagen types I and III, fibronectin, elastin and TIMP-3 (tissue inhibitor of metalloproteinases 3). Furthermore, it decreases the apoptotic cell death induced by oxidized low-density lipoproteins and elastase in aortic VSMCs[122]. On the other hand, in healthy blood vessels, VSMCs reside within the lamina media and remain inactive. There is evidence that the proliferation and abnormal migration of VSMCs are a common event in the pathophysiology of many vascular diseases, where these cells proliferate and migrate from the media to the intima, leading to hyperplasia and vascular stenosis[123–125]. *PPARδ* activation represents protection against endothelial damage and dysfunction, as well as vascular cell proliferation through various mechanisms[126]. In this regard, studies suggest that PPARδ significantly inhibits PDGF-induced proliferation in VSMCs by repressing expression of cyclin D1, cyclin D3, CDK2 and CDK4[127]. In summary, *PPARδ* plays a prominent role in maintaining vascular integrity via suppressing VSMCs proliferation, inhibiting VSMCs migration with preservation of extracellular matrix (ECM), as well as inhibiting apoptosis and senescence of VSMCs by upregulating antioxidant genes and suppressing inflammation. We believe that the effect that *PPARδ* exerts in protecting the integrity of the vascular structure of the aorta could be slightly counteracting the damaging effect produced by a defective fibrillin 1 present in

patients with MFS, and that, in keeping with this hypothesis, missense variants in the *PPARD* gene could be decreasing its protective function of the vascular structure and promoting the severe aortic phenotype in these patients (Figure 14).

Table 10. Genetic variants in candidate genes selected under different VAAST analysis categories

Candidate Genes	Genomic coordinates	Variations HGVS nomenclarure	Variant type	SIFT
PPARD	Chr6:35387913	c.140G>A	missense	tolerated-low confidence
FBN1	Chr15:48808467	c.1240C>A	missense	tolerated
	Chr20:10620449	c.3355G>C	missense	tolerated
JAG1	Chr20:10620450	c.3353T>A	missense	tolerated
	Chr5:56111414	c.14C>G	missense	deleterious-low confidence
	Chr5:56177614	c.2587G>T	missense	tolerated-low confidence
	Chr5:.56111762	c.362G>A	missense	tolerated
MAP3K1	Chr5:56168548-56168649	c.1504_1505+101del	frameshift deletion	neutral
	Chr11:48166437-48166511	c.2786_2786+73del	frameshift deletion	unknown/unassessed
	Chr11:48002530-48002532	c.66_68del	in-frame deletion	damaging
PTPRJ	Chr11:48161067G	c.2182G>A	missense	deleterious
	Chr10:30316501-30316503	c.2574_2576del	frameshift deletion	unknown/unassessed
	Chr10:30318653	c.424C>T	missense	tolerated
KIAA1462	Chr10:30316499-30316500	c.2577_2578insACTGCTGCT	in-frame insertion	unknown/unassessed
	Chr1:173010746	c.361G>A	missense	tolerated
	Chr1:173010656	c.451A>G	missense	tolerated
TNFSF18	Chr1:173010651-173010652	c.455_456insTTG	frameshift insertion	unknown/unassessed
TGFBR3L	Chr19:7981648-7981650	c.418_420del	in-frame deletion	unknown/unassessed

When a recessive mode of inheritance was considered in the analysis of family 1.1 in VAAST, a second missense variant in the *FBN1* gene (c.1240C>A) (Table 10), additional to the causative variant of MFS, was found. In this case, the patient III1 of pedigree of family 1 was classified as “severe” and has one variant in *FBN1* that she shares with her mother (MFS-causing variant) and a second variant in *FBN1* that is shared with her father (Non-pathogenic variant). This variant substitutes the conserved, non-polar amino-acid Proline by the polar amino-acid Threonine in the position 414 of the protein, and it is located in a domain where it has been predicted that folding that contributes to tertiary structure of this protein occurs[128,129]. It is well known that variants that produce changes from a non-polar to a polar group significantly modify the characteristics of a protein and therefore could modify its function[130]. Previous studies have shown that variants in *FBN1* cause variation in the expression levels of the gene and in light of this, this second *FBN1* variant in this patient could be considered a modifier gene that favors a severe MFS phenotype[50,51,131] (Figure 14).

Interestingly, another candidate gene that could modulate the cardiovascular phenotype was prominent in our results. Two missense variants in *JAG1* (c.3355G>C and c.3353T>A) (Table 10) segregated with severe aortic phenotype in MFS in family 3 under a recessive inheritance model. These variants alter a well-conserved amino-acid (Gln1118Pro) in an evolutionarily conserved cytoplasmic domain of JAG1[132]. The *JAG1* gene encodes a protein called Jagged-1, which interacts with the Notch signaling pathway. Jagged-1 connects with Notch receptors (Notch1 to Notch4), resulting in the release of the Notch Intracellular

Domain (NICD) by proteolytic cleavage and subsequent translocation of NICD to the nucleus, where it interacts with the Recombinant Binding Protein Suppressor (RBPJ) and other transcription factors to regulate the activation and suppression of target genes that initiate a signaling cascade involved in vascular wall homeostasis and remodeling[133–135]. Jagged-1 normally activates Notch in VSMCs in the lamina media of the aorta, spreading a signaling pulse that is crucial to induce the differentiation of the VSMC layer towards the contractile phenotype to which VSMCs must return after undergoing a phenotypic switch to the synthetic phenotype in order to promote remodeling and, ultimately, maintain vascular homeostasis in response to hemodynamic alterations[134–136]. We hypothesize that missense variants found in *JAG1* could downregulate the Jagged1-Notch signaling and therefore impede proper restoration of vascular wall homeostasis (Figure 14).

Alterations in Notch signaling can lead to BAV, which has been observed in many patients with an aneurysm of the ascending aorta. The mechanisms by which Notch signaling is involved in the pathogenesis of the aneurysm are not clear, however, studies have shown a decrease in Notch signaling in the aorta of patients with BAV vs patients with tricuspid aortic valve, suggesting an association of Notch with pathology in the ascending aorta[137]. Furthermore, it is known that the transition from endocardial cells to mesenchymal cells is a necessary process for endocardial cells to invade the underlying extracellular matrix and become resident interstitial cells, this process can be initiated by both TGF-β signaling and by Notch signaling[138]. The association of BAV with aneurysm of the ascending aorta has led to the hypothesis that interference with Notch signaling in cells of the neural crest,

a cell line that participates in the formation of the aortic valve and the aorta, may be responsible for the abnormal morphogenesis of both structures[139].

Another candidate gene was *MAP3K1*. We identified variants associated with the severe phenotype in this gene both in the familial approach (family 3) and in the non-familial, case-control approach (Table 10). We identified three missense variants (c.14C>G, c.2587G>T, c.362G>A located in conserved regions of the MAP3K1 protein A5, V863 and G121 respectively). Interestingly, another patient with a severe aortic phenotype presented a deletion (c.1504_1505+101del) that encompasses the last two nucleotides of exon 1 and a major portion of intron 1. We hypothesize that this generates an aberrant mRNA that contains part of intron 1 and that incorporates a premature termination codon (PTC) that would undergo nonsense mediated decay (NMD) and render this allele non-functional.

MAP3K1 is the Mitogen-activated protein kinase kinase kinase 1, which normally works by inhibiting ECM cell shedding-induced apoptosis (anoikis). One consequence of ECM degradation in the pathophysiology of MFS is the increase in MMPs[140], which leads to degradation of fibronectin[141] and detachment of VSMCs and their subsequent apoptosis[142]. This apoptosis process induced by detachment of cells (anoikis) is controlled by the effect of the MAP3K1 protein and reduces cell death by activating pro-survival targets[143,144] (Figure 14).

PTPRJ was another gene that showed variants that segregated in patients with severe phenotype but not in patients with mild phenotype (frameshift variant c.2786_2786+74del, frameshift variant c.66_68del and missense variant c.2182G>A that produces the E728K substitution in the protein). Two of these

variants (c.66_68del and c.2182G>A) were predicted to have a damaging or deleterious effect that modifies the structure and function of the protein by SIFT (Table 10). c.2786_2786+74del frameshift variant is a 74 bp deletion that includes the last bases of exon 13 and part of the next intron, therefore, there is a loss of the 5' splicing donor site that generates an aberrant mRNA with 14 additional aminoacids generated from intronic sequence an also a premature termination codon (PTC). The other 48166511del frameshift variant (c.66_68del) is a 3 bp deletion that causes a loss of a single leucine residue in the protein. The variants c.2786_2786+74del and c.2182G>A are located on the non-cytoplasmatic Fibronectin type III domain which is an evolutionary conserved protein domain. We conclude that these variants could then have a major impact on the structure and function of the protein. The variant g.48002530-48002532del is localized in the signal peptides region of *PTPRJ* and its effect on translational suppression of *PRPRJ* could be assessed[145,146].

PTPTJ or protein receptor tyrosine phosphatase type J (also named *CD148* or *DEP-1*) is expressed in vascular endothelial cells. *PTPRJ* has the capacity to modulate cell proliferation, cell migration and is involved in vascular development through negative regulation of VEGF receptor 2 (*VEGFR2*) signaling[147,148] (Figure 14). Variants in *PTPRJ* have been associated with coarctation of the aorta[149]. A recent study reported that *PTPRJ* is upregulated in the aortic tissue of patients with thoracic aortic aneurysm (TAA) and could predict the presentation of TAA[150]. Although these results are still not very clear, suggest that *PTPRJ* could be exerting some function for maintenance of homeostasis of the aorta. Loss-of-

function of *PTPRJ* due to the presence of gene variants could be associated with TAA produced by the uncontrolled activation of the TGFβ signaling pathway in patients with variants in *FBN1*.

4.4.2 Candidate genes associated with the mild cardiovascular phenotype in MFS

Finally, our VAAST findings show 3 candidate genes (*KIAA1462, TNFSF18, and TGFBR3L*) (Table 10) (Figure 15), that segregated with mild aortic phenotype, suggesting they could be mitigating the severity of cardiovascular phenotype in MFS, rendering variants with a protective effect. Two of these, *KIAA1462* and *TNFSF18* are involved in degradation of the ECM through the activation of metalloproteinases. For *KIAA1462* we identified 3 variants associated with the mild phenotype. A frameshift variant (c.2574_2576del) that modifies the reading frame and generates the loss of a serine in the amino acid sequence. We also identified an in-frame variant (c.2577_2578insACTGCTGCT) which produced an insertion of the amino-acids Thr-Ala-Ala in the protein. Finally, we found a missense variant (c.424C>T) that changes the conserved amino acid Ala142 to Thr142. *KIAA1462* Also known as *JCAD* (Junctional Cadherin 5 Associated), is a cell-cell binding protein expressed primarily in endothelial cells[151] and regulates pathological angiogenesis, rather than developmental angiogenesis[152].

Recent studies have shown that depletion of *KIAA1462* decreased the expression of genes such as *VCAM1*[153], which is involved in the pathogenesis of ascending aortic aneurysm through increasing the activity of MMP-2 and MMP-9 that

promotes aortic aneurysm progression[154]. Jones et al. described a model for the effect of JCAD in conjunction with the Hippo signaling pathway in coronary artery disease (CAD). They suggest that JCAD acts downstream of RhoA to inhibit LATS1/2 and in turn there is a decrease in phosphorylation of YAP/TAZ[155]. It has been reported that in dissection and aneurysm of the aorta, downregulation of YAP/TAZ promotes the apoptosis of VSMCs[156], while Liu et al showed that downregulation of YAP/TAZ induces apoptosis of VSMCs and aortic dissection and that overexpression of YAP/TAZ is able to reverse this effect[157]. We believe that it is probable that variants in the *JCAD* gene could decrease its expression and in turn generate an increase in YAP/TAZ that, as mentioned above, could produce a protective effect for the appearance of aortic dissection (Figure 15). On the other hand, we identified 2 missense variants (c.361G>A and c.451A>G) in the gene *TNFSF18* that change aminoacids that are largely conserved: Gly121Ser and Asn151Asp respectively.

Additionally, we found a frameshift insertion (c.455_456insTTG) that incorporates a premature termination codon (PTC) and generates a shorter, aberrant mRNA. All the variants found in this gene are in an extracellular domain and using pDomTHREADER (recognition algorithm) was recognized how folk domain, therefore these variants could produce a conformational change of the protein[158]. *TNFSF18* (member 18 of the tumor necrosis factor ligand superfamily) activates the expression of MMP-2 and MMP-9 but by phosphorylation of p-STAT1 and positively regulates the expression of VCAM1 and ICAM1[159] (Figure 15).

TGFBR3L is an important paralog of the *TGFBR3* gene that is a coreceptor for TGF-β and has a role in the regulation of the transforming growth factor receptor beta signaling pathway superfamily by promoting the binding of TGFβ2 to the TGFβR1/TGFβR2 receptor complex[160]. One of the functions of TGFβR3 is to serve as a presenter of the TGFβ2 ligand to the TGFβR2 receptor that will then form a heterodimeric TGFβR1-TGFβR2 complex and when this is phosphorylated, the TGF-β signaling cascade is initiated[160,161]. In this sense, it has been shown that activation of TGFBR3 is greater in the presence of variants in *FBN1* and is involved in the over-activation observed in patients with MFS[162] (Figure 15). We believe that the in-frame variant in *TGFBR3L* found in this study (c.418_420del) generated a loss of amino acid Pro140 in the mRNA (Table 10) and is located in the zona pellucida (ZP)[163] domain of TGFBR3 whose C-region regulates both the polymerization of extracellular matrix proteins and the recognition of TGFB by the TGFBR3 receptor[164], could be decreasing the activation of TGFBR3 and therefore slowing the uncontrolled activation of the TGFB signaling pathway and protecting carriers against a severe aortic phenotype.

Our results show that there are variants in genes that regulate components of the extracellular matrix, TGFβ pathway and in genes that function in other signaling pathways that are involved in the maintenance of the ECM or angiogenesis and that could modify the cardiovascular phenotype in MFS. We found 5 candidate genes from the VAAST analysis associated with severe aortic phenotype in MFS patients and 3 genes that could be protective for this phenotype.

Fibrillin 1 plays a pivotal role in regulating the bioavailability of TGF-β[118,165,166] and MFS-causing variants in this gene considerably alter ECM homeostasis by disrupting fibrillin 1 ability to maintain TGF-β sequestered and inactive. This, in turn, pervasively increases the levels of TGF-β ligand in the extracellular space[118]. TGF-β is a molecule that regulates many biological processes, including apoptosis, collagen production and remodeling of the extracellular matrix (ECM), among others[13,119]. Therefore, dysregulation of TGF-β pathway interferes with maintenance of homeostasis of the aortic wall and its effect depends on the bioavailability of TGF-β in the aortic tissue, but also on other factors that could play a key role in patients with MFS and that have not been fully de-scribed. Discovery and characterization of these other factors could be key to explain differences in aortic phenotype in patients with MFS. In this sense, identification of genes (and gene products) that could modulate active TGF-β levels should be a top priority to understand the variable expressivity observed in the cardiovascular phenotype of MFS patients. Therefore, our search for modifier genes was aimed at identifying elements that either increase TGF-β pathway activation or that interfere in the degradation process of the ECM in such a manner that favor the appearance of an aortic phenotype. Alternatively, we aimed to identify genes that could downregulate the uncontrolled activation of TGF-β pathway, which, in keeping with our hypothesis, would render protective to the normal architecture of the aorta.

5. CONCLUSIONS

The genetic forms of TAAs (MFS, LDS and vEDS) present a high variability in the presentation of phenotype. Understanding this phenotypic variability could help to predict clinical outcomes in these patients. Therefore, we propose a new strategy to stratify patients based on mild or severe phenotypes that allow us to estimate risk profiles and finally could provide a future precision medicine model for their management.

We were able to create a registry of Chilean patients with aortic disease where we could incorporate epidemiological, familial, phenotypic (clinical) and molecular information and separate these into different etiologies, which will seek to identify truly informative patients for the study of phenotype modifying genes cardiovascular disease in ICTDs, we were finally able to classify 8 patients with severe phenotype and 14 with mild cardiovascular phenotype. This classification was made according to the age of presentation of the cardiovascular manifestations.

Patients with ICTDs present a high variability in the severity of the aortic phenotype. Understanding this phenotypic variability would allow predicting clinical outcomes in these patients, taking more effective measures for proper clinical management and avoidance of life risk situations. In this study we proposed a novel strategy to stratify ICTDs patients into mild or severe aortic phenotype, which made it possible to estimate risk profiles through the identification of variants associated with these extreme phenotypes.

Although Exome Sequencing has become an important tool for the discovery of genetic variants, its use has some limitations in relation to the loss of important information in genomic regions with insufficient coverage. The search for both pathogenic and modifier variants with ES data is therefore done at risk of the loss of information in non-coding regions of the human genome. We hope to further our studies using Whole Genome Sequencing data in the future to overcome this liability.

Our results show that there are variants in genes that regulate the component extracellular matrix or TGFβ pathway and other intracellular pathways that could modify the cardiovascular phenotype in MFS. We find 6 candidate genes from the VAAST analysis associated with severe cardiovascular phenotype in patients MFS and 3 genes that could be protective genes for this phenotype.

Finally, future studies are needed to determine the molecular mechanisms by which these genetic variants could act to modify the severity of aortic phenotype in these patients, and whether these genes are involved in the development of aortic aneurysm using in vitro and in vivo experiment models.

6. BIBLIOGRAPHIC REFERENCES:

1. Verstraeten, A., Alaerts, M., Van Laer, L. & Loeys, B. Marfan Syndrome and Related Disorders: 25 Years of Gene Discovery. *Human Mutation* **37**, 524–531 (2016).
2. Lindsay, M. E. & Dietz, H. C. Lessons on the pathogenesis of aneurysm from heritable conditions. *Nature* **473**, 308–16 (2011).
3. Dietz, H. C. *et al.* Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* **352**, 337–339 (1991).
4. Malfait, F., Wenstrup, R. J. & De Paepe, A. Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. *Genetics in Medicine* vol. 12 597–605 (2010).
5. Van Hemelrijk, C., Renard, M. & Loeys, B. The Loeys-Dietz syndrome: An update for the clinician. *Current Opinion in Cardiology* vol. 25 546–551 (2010).
6. Keane, M. G. & Pyeritz, R. E. Medical management of marfan syndrome. *Circulation* vol. 117 2802–2813 (2008).
7. Ramirez, F. & Dietz, H. C. Marfan syndrome: from molecular pathogenesis to clinical treatment [Curr. Opin. Genet. Dev. 17 (2007) 252-258] (DOI:10.1016/j.gde.2007.04.006). *Current Opinion in Genetics and Development* vol. 17 367 (2007).
8. Nadeau, J. H. Modifier genes in mice and humans. *Nature Reviews Genetics* **2**, 165–174 (2001).

9. Wrana, J. L., Attisano, L., Wieser, R., Ventura, F. & Massagué, J. Mechanism of activation of the TGF-β receptor. *Nature 1994 370:6488* **370**, 341–347 (1994).

10. Sinha, S., Nevett, C., Shuttleworth, C. A. & Kielty, C. M. Cellular and extracellular biology of the latent transforming growth factor-β binding proteins. *Matrix Biology* **17**, 529–545 (1998).

11. Nakao, A. *et al.* TGF-β receptor-mediated signalling through Smad2, Smad3 and Smad4. *The EMBO Journal* **16**, 5353–5362 (1997).

12. Massagué, J. TGFβ signalling in context. *Nature reviews. Molecular cell biology* **13**, 616 (2012).

13. Dietz, H. C. Marfan syndrome: From molecules to medicines. *American Journal of Human Genetics* **81**, 662–667 (2007).

14. Johnston, K. W. *et al.* Suggested standards for reporting on arterial aneurysms. Subcommittee on Reporting Standards for Arterial Aneurysms, Ad Hoc Committee on Reporting Standards, Society for Vascular Surgery and North American Chapter, International Society for Cardiovascular Surgery. *Journal of vascular surgery* **13**, 452–8 (1991).

15. Ruddy, J. M., Jones, J. A., Spinale, F. G. & Ikonomidis, J. S. Regional heterogeneity within the aorta: Relevance to aneurysm disease. *The Journal of Thoracic and Cardiovascular Surgery* **136**, 1123–1130 (2008).

16. Erbel, R. *et al.* 2014 ESC guidelines on the diagnosis and treatment of aortic diseases. *European Heart Journal* **35**, 2873–2926 (2014).

17. The Genetics of Aortic Aneurysms - American College of Cardiology. https://www.acc.org/latest-in-cardiology/articles/2018/05/02/12/52/the-

genetics-of-aortic-aneurysms.

18. Libby, P., Ridker, P. M. & Hansson, G. K. Progress and challenges in translating the biology of atherosclerosis. *Nature 2011 473:7347* **473**, 317–325 (2011).

19. Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. *Lancet (London, England)* **389**, 37–55 (2017).

20. S, K. *et al.* Prevalence and Risk Factors of Hypertension in Urban Areas of Cameroon: A Nationwide Population-Based Cross-Sectional Study. *Journal of clinical hypertension (Greenwich, Conn.)* **17**, 819–824 (2015).

21. E, T., F, E., S, A. & N, N. The prevalence of bicuspid aortic valve in newborns by echocardiographic screening. *American heart journal* **150**, 513–515 (2005).

22. Gould, R. A. *et al.* ROBO4 variants predispose individuals to bicuspid aortic valve and thoracic aortic aneurysm. *Nature Genetics 2018 51:1* **51**, 42–50 (2018).

23. Lindman, B. R. *et al.* Calcific aortic stenosis. *Nature reviews. Disease primers* **2**, 16006 (2016).

24. A, D. *et al.* Differences in Aortopathy in Patients with a Bicuspid Aortic Valve with or without Aortic Coarctation. *Journal of clinical medicine* **9**, (2020).

25. WAGENSEIL, J. E. & MECHAM, R. P. Vascular Extracellular Matrix and Arterial Mechanics. *Physiological reviews* **89**, 957 (2009).

26. Olsson, C., Thelin, S., Stahle, E., Ekbom, A. & Granath, F. Thoracic Aortic Aneurysm and Dissection: Increasing Prevalence and Improved Outcomes

Reported in a Nationwide Population-Based Study of More Than 14 000 Cases From 1987 to 2002. *Circulation* **114**, 2611–2618 (2006).

27. Wu, D., Shen, Y. H., Russell, L., Coselli, J. S. & LeMaire, S. A. Molecular Mechanisms of Thoracic Aortic Dissection. *The Journal of surgical research* **184**, 907 (2013).

28. Shimizu, K., Mitchell, R. N. & Libby, P. Inflammation and Cellular Immune Responses in Abdominal Aortic Aneurysms. *Arteriosclerosis, Thrombosis, and Vascular Biology* **26**, 987–994 (2006).

29. Wang, Y. *et al.* TGF-beta activity protects against inflammatory aortic aneurysm progression and complications in angiotensin II-infused mice. *The Journal of clinical investigation* **120**, 422–32 (2010).

30. Tang, P. C. Y. *et al.* Transmural inflammation by interferon-γ-producing T cells correlates with outward vascular remodeling and intimal expansion of ascending thoracic aortic aneurysms. *The FASEB Journal* **19**, 1528–1530 (2005).

31. He, R. *et al.* Characterization of the inflammatory and apoptotic cells in the aortas of patients with ascending thoracic aortic aneurysms and dissections. *The Journal of Thoracic and Cardiovascular Surgery* **131**, 671-678.e2 (2006).

32. Albornoz, G. *et al.* Familial Thoracic Aortic Aneurysms and Dissections—Incidence, Modes of Inheritance, and Phenotypic Patterns. *The Annals of Thoracic Surgery* **82**, 1400–1405 (2006).

33. Booher, A. M. & Eagle, K. A. Diagnosis and management issues in thoracic aortic aneurysm. *American Heart Journal* **162**, 38-46.e1 (2011).

34. Callewaert, B. *et al.* New insights into the pathogenesis of autosomal-dominant cutis laxa with report of five ELN mutations. *Human Mutation* **32**, 445–455 (2011).

35. Sawyer, S. L. *et al.* Longer Term Survival of a Child With Autosomal Recessive Cutis Laxa Due to a Mutation in *FBLN4*. *American Journal of Medical Genetics Part A* **161**, 1148–1153 (2013).

36. Schepers, D. *et al.* The SMAD-binding domain of SKI: a hotspot for de novo mutations causing Shprintzen–Goldberg syndrome. *European Journal of Human Genetics* **23**, 224–228 (2015).

37. Zoppi, N., Chiarelli, N., Cinquina, V., Ritelli, M. & Colombi, M. GLUT10 deficiency leads to oxidative stress and non-canonical αvβ3 integrin-mediated TGFβ signalling associated with extracellular matrix disarray in arterial tortuosity syndrome skin fibroblasts. *Human Molecular Genetics* **24**, 6769–6787 (2015).

38. Wang, L. *et al.* Mutations in Myosin Light Chain Kinase Cause Familial Aortic Dissections. *The American Journal of Human Genetics* **87**, 701–707 (2010).

39. Guo, D. *et al.* Recurrent Gain-of-Function Mutation in PRKG1 Causes Thoracic Aortic Aneurysms and Acute Aortic Dissections. *The American Journal of Human Genetics* **93**, 398–404 (2013).

40. De Backer, J. *et al.* Utility of molecular analyses in the exploration of extreme intrafamilial variability in the Marfan syndrome. *Clinical Genetics* **72**, 188–198 (2007).

41. Loeys, B. The search for genotype/phenotype correlation in Marfan syndrome: To be or not to be? *European Heart Journal* vol. 37 3291–3293

(2016).

42. Robinson, P. N. *et al.* The molecular genetics of Marfan syndrome and related disorders. *Journal of Medical Genetics* vol. 43 769–787 (2006).

43. Salik, I. & Rawla, P. *Marfan Syndrome*. *StatPearls* (StatPearls Publishing, 2020).

44. Sakai, L. Y., Keene, D. R., Renard, M. & De Backer, J. FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders. *Gene* vol. 592 (2016).

45. Li, Y. *et al.* A FBN1 mutation association with different phenotypes of Marfan syndrome in a Chinese family. *Clinica Chimica Acta* **460**, 102–106 (2016).

46. Takeda, N. *et al.* Impact of Pathogenic FBN1 Variant Types on the Progression of Aortic Disease in Patients With Marfan Syndrome. *Circulation. Genomic and precision medicine* **11**, (2018).

47. Díaz de Bustamante, A., Ruiz-Casares, E., Darnaude, M. T., Perucho, T. & Martínez-Quesada, G. Phenotypic Variability in Marfan Syndrome in a Family With a Novel Nonsense FBN1 Gene Mutation. *Revista Española de Cardiología (English Edition)* **65**, (2012).

48. Hutchinson, S. *et al.* Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: A potential modifier of phenotype? *Human Molecular Genetics* **12**, (2003).

49. Fernandes, G. R., Massironi, S. M. G. & Pereira, L. V. Identification of Loci Modulating the Cardiovascular and Skeletal Phenotypes of Marfan Syndrome in Mice. *Scientific Reports* **6**, (2016).

50. Gerdes Gyuricza, I., Barbosa de Souza, R., Farinha-Arcieri, L. E., Ribeiro

Fernandes, G. & Veiga Pereira, L. Is HSPG2 a modifier gene for Marfan syndrome? *European Journal of Human Genetics* **28**, (2020).

51. Lima, B. L. *et al.* A new mouse model for marfan syndrome presents phenotypic variability associated with the genetic background and overall levels of Fbn1 expression. *PLoS ONE* **5**, (2010).
52. Michel, J.-B., Jondeau, G. & Milewicz, D. M. From genetics to response to injury: vascular smooth muscle cells in aneurysms and dissections of the ascending aorta. *Cardiovascular Research* **114**, 578–589 (2018).
53. Bolar, N., Van Laer, L. & Loeys, B. L. Marfan syndrome. *Current Opinion in Pediatrics* **24**, 498–504 (2012).
54. Castellano, J. M., Kovacic, J. C., Sanz, J. & Fuster, V. Are we ignoring the dilated thoracic aorta? *Annals of the New York Academy of Sciences* **1254**, 164–174 (2012).
55. Habashi, J. P. *et al.* Losartan, an AT1 Antagonist, Prevents Aortic Aneurysm in a Mouse Model of Marfan Syndrome. *Science* **312**, 117–121 (2006).
56. Schepers, D. *et al.* A mutation update on the LDS-associated genes TGFB2/3 and SMAD2/3. *Human mutation* **39**, 621–634 (2018).
57. MacCarrick, G. *et al.* Loeys–Dietz syndrome: a primer for diagnosis and management. *Genetics in Medicine* **16**, 576–587 (2014).
58. Loeys, B. L. *et al.* Aneurysm Syndromes Caused by Mutations in the TGF-β Receptor. *New England Journal of Medicine* **355**, 788–798 (2006).
59. Eagleton, M. J. Arterial complications of vascular Ehlers-Danlos syndrome. *Journal of vascular surgery* **64**, 1869–1880 (2016).
60. Bradley, T. J., Bowdin, S. C., Morel, C. F. J. & Pyeritz, R. E. The Expanding

Clinical Spectrum of Extracardiovascular and Cardiovascular Manifestations of Heritable Thoracic Aortic Aneurysm and Dissection. *Canadian Journal of Cardiology* **32**, 86–99 (2016).

61. Papagiannis, J. Sudden death due to aortic pathology. *Cardiology in the Young* **27**, S36–S42 (2017).

62. Malfait, F. *et al.* The 2017 international classification of the Ehlers-Danlos syndromes. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* **175**, 8–26 (2017).

63. Germain, D. P. Ehlers-Danlos syndrome type IV. *Orphanet Journal of Rare Diseases* **2**, 32 (2007).

64. Verstraeten, A., Luyckx, I. & Loeys, B. Aetiology and management of hereditary aortopathy. *Nature Reviews Cardiology* **14**, 197–208 (2017).

65. Dang, H. *et al.* Novel variation at chr11p13 associated with cystic fibrosis lung disease severity. *Human Genome Variation* **3**, 16020 (2016).

66. Sampson, U. K. A. *et al.* Global and regional burden of aortic dissection and aneurysms: Mortality trends in 21 world regions, 1990 to 2010. *Global Heart* **9**, 171-180.e10 (2014).

67. Sampson, U. K. A. *et al.* Estimation of global and regional incidence and prevalence of abdominal aortic aneurysms 1990 to 2010. *Global Heart* **9**, 159–170 (2014).

68. Kroner, B. L. *et al.* The National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC): Results from phase I and scientific opportunities in phase II. *American Heart Journal* **162**, 627-632.e1 (2011).

69. Eagle, K. A. Rationale and design of the National Registry of Genetically Triggered Thoracic Aortic Aneurysms and Cardiovascular Conditions (GenTAC). *American Heart Journal* **157**, 319–326 (2009).

70. FONASA. Caracterización sociodemográfica y socioeconómica en la población asegurada inscrita. 1–13 (2020).

71. Loeys, B. L. *et al.* The revised Ghent nosology for the Marfan syndrome. *Journal of medical genetics* **47**, 476–85 (2010).

72. Kodra, Y. *et al.* Data quality in rare diseases registries. in *Advances in Experimental Medicine and Biology* vol. 1031 149–164 (Springer New York LLC, 2017).

73. Maccarthy, J., Guerin, S., Wilson, A. G. & Dorris, E. R. Facilitating public and patient involvement in basic and preclinical health research. *PLoS ONE* **14**, e0216600 (2019).

74. Fink, A. K., Loeffler, D. R., Marshall, B. C., Goss, C. H. & Morgan, W. J. Data that empower: The success and promise of CF patient registries. *Pediatric Pulmonology* vol. 52 S44–S51 (2017).

75. Powell, J. T. & Wanhainen, A. Analysis of the Differences Between the ESVS 2019 and NICE 2020 Guidelines for Abdominal Aortic Aneurysm. *European Journal of Vascular and Endovascular Surgery* **60**, 7–15 (2020).

76. Hiratzka, L. F. *et al.* 2010 ACCF/AHA/AATS/ACR/ASA/SCA/SCAI/SIR/STS/SVM guidelines for the diagnosis and management of patients with thoracic aortic disease: Executive summary: A report of the american college of cardiology foundation/american heart association task force on pra. *Circulation* **121**,

266–369 (2010).

77. Mehta, R. H. *et al.* Acute type A aortic dissection in the elderly: Clinical characteristics, management, and outcomes in the current era. *Journal of the American College of Cardiology* **40**, 685–692 (2002).

78. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature 2012 493:7431* **493**, 216–220 (2012).

79. Ilinca, A. *et al.* Whole-Exome Sequencing in 22 Young Ischemic Stroke Patients With Familial Clustering of Stroke. *Stroke* 1056–1063 (2020) doi:10.1161/STROKEAHA.119.027474.

80. Ultra-Rare Genetic Variation in the Epilepsies: A Whole-Exome Sequencing Study of 17,606 Individuals. *American journal of human genetics* **105**, 267–282 (2019).

81. Seo, G. H. *et al.* Identification of extremely rare mitochondrial disorders by whole exome sequencing. *Journal of Human Genetics 2019 64:11* **64**, 1117–1125 (2019).

82. Bruse, S. *et al.* Whole exome sequencing identifies novel candidate genes that modify chronic obstructive pulmonary disease susceptibility. *Human Genomics* **10**, 1–13 (2016).

83. Teer, J. K. & Mullikin, J. C. Exome sequencing: the sweet spot before whole genomes. *Human molecular genetics* **19**, (2010).

84. Petersen, B. S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D. & Franke, A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC genetics* **18**, (2017).

85. Majewski, J., Schwartzentruber, J., Lalonde, E., Montpetit, A. & Jabado, N.

What can exome sequencing do for you? *Journal of medical genetics* **48**, 580–589 (2011).

86. Rabbani, B., Tekin, M. & Mahdieh, N. The promise of whole-exome sequencing in medical genetics. *Journal of human genetics* **59**, 5–15 (2014).

87. Altmann, A. *et al.* A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Human genetics* **131**, 1541–1554 (2012).

88. Ewing, B., Hillier, L. D., Wendl, M. C. & Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Research* **8**, 175–185 (1998).

89. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

90. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

91. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, (2009).

92. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

93. Picard Tools - Docs CSC. https://docs.csc.fi/apps/picard/.

94. Base Quality Score Recalibration (BQSR) – GATK. https://gatk.broadinstitute.org/hc/en-us/articles/360035890531-Base-Quality-Score-Recalibration-BQSR-.

95. gatk-docs/(howto)_Perform_local_realignment_around_indels.md at master ·

broadinstitute/gatk-docs · GitHub. https://github.com/broadinstitute/gatk-docs/blob/master/gatk3-tutorials/(howto)_Perform_local_realignment_around_indels.md.

96. HaplotypeCaller – GATK. https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller.

97. Variant Quality Score Recalibration (VQSR) – GATK. https://gatk.broadinstitute.org/hc/en-us/articles/360035531612-Variant-Quality-Score-Recalibration-VQSR-).

98. ANNOVAR Documentation. https://annovar.openbioinformatics.org/en/latest/.

99. Davydov, E. V. *et al.* Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++. *PLOS Computational Biology* **6**, e1001025 (2010).

100. S, C. & JC, F. Identification of deleterious mutations within three human genomes. *Genome research* **19**, 1553–1561 (2009).

101. JM, S., C, R., M, S. & D, S. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature methods* **7**, 575–576 (2010).

102. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods 2010 7:4* **7**, 248–249 (2010).

103. KS, P., MJ, H., KR, R. & A, S. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* **20**, 110–121 (2010).

104. Kumar, P., Henikoff, S. & Ng, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols 2009 4:7* **4**, 1073–1081 (2009).

105. Robertson, I., Jensen, S. & Handford, P. TB domain proteins: evolutionary

insights into the multifaceted roles of fibrillins and LTBPs. *The Biochemical journal* **433**, 263–276 (2011).

106. Handford, P. A. Fibrillin-1, a calcium binding protein of extracellular matrix. *Biochimica et biophysica acta* **1498**, 84–90 (2000).
107. Yadin, D. A. *et al.* Structure of the Fibrillin-1 N-Terminal Domains Suggests that Heparan Sulfate Regulates the Early Stages of Microfibril Assembly. *Structure(London, England:1993)* **21**, 1743 (2013).
108. Franken, R. *et al.* Relationship between fibrillin-1 genotype and severity of cardiovascular involvement in Marfan syndrome. *Heart* **103**, 1795–1799 (2017).
109. Arnaud, P. *et al.* Clinical relevance of genotype–phenotype correlations beyond vascular events in a cohort study of 1500 Marfan syndrome patients with FBN1 pathogenic variants. *Genetics in Medicine 2021 23:7* **23**, 1296–1304 (2021).
110. Faivre, L. *et al.* Effect of Mutation Type and Location on Clinical Outcome in 1,013 Probands with Marfan Syndrome or Related Phenotypes and FBN1 Mutations: An International Study. *The American Journal of Human Genetics* **81**, 454–466 (2007).
111. Aubart, M. *et al.* Association of modifiers and other genetic factors explain Marfan syndrome clinical variability. *European Journal of Human Genetics* **26**, 1759–1772 (2018).
112. B, K. *et al.* Using VAAST to Identify Disease-Associated Variants in Next-Generation Sequencing Data. *Current protocols in human genetics* **81**, (2014).

113. Yandell, M. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Research* **21**, 1529–1542 (2011).

114. S, B. *et al.* Whole exome sequencing identifies novel candidate genes that modify chronic obstructive pulmonary disease susceptibility. *Human genomics* **10**, (2016).

115. Pinnaro, C. T. *et al.* Candidate modifier genes for immune function in 22q11.2 deletion syndrome. *Molecular Genetics and Genomic Medicine* **8**, 1–10 (2020).

116. Génin, E., Feingold, J. & Clerget-Darpoux, F. Identifying modifier genes of monogenic disease: Strategies and difficulties. *Human Genetics* **124**, 357–368 (2008).

117. SIFT - Predict effects of nonsynonmous / missense variants. https://sift.bii.a-star.edu.sg/.

118. Chaudhry, S. S. *et al.* Fibrillin-1 regulates the bioavailability of TGFβ1. *Journal of Cell Biology* **176**, 355–367 (2007).

119. Huang, F. & Chen, Y. G. Regulation of TGF-β receptor activity. *Cell and Bioscience* **2**, 9 (2012).

120. Wright, P. E. & Jane Dyson, H. Intrinsically disordered proteins in cellular signalling and regulation. (2015) doi:10.1038/nrm3920.

121. F1D8S7 (F1D8S7) - protein - InterPro. http://www.ebi.ac.uk/interpro/protein/UniProt/F1D8S7/.

122. Kim, H. J. *et al.* Peroxisome proliferator-Activated receptor δ regulates extracellular matrix and apoptosis of vascular smooth muscle cells through the activation of transforming growth factor-β1/Smad3. *Circulation Research*

105, 16–24 (2009).

123. Orr, A. W., Hastings, N. E., Blackman, B. R. & Wamhoff, B. R. Complex regulation and function of the inflammatory smooth muscle cell phenotype in atherosclerosis. *Journal of Vascular Research* **47**, 168–180 (2010).
124. Yu, H., Payne, T. J. & Mohanty, D. K. Effects of Slow, Sustained, and Rate-Tunable Nitric Oxide Donors on Human Aortic Smooth Muscle Cells Proliferation. *Chemical Biology and Drug Design* **78**, 527–534 (2011).
125. Petit, C. *et al.* Regulation of SMC traction forces in human aortic thoracic aneurysms. *Biomechanics and Modeling in Mechanobiology* **20**, 717–731 (2021).
126. Jia, Z. & Li, S. PPARβ/δ in the Pathogenesis of Vascular Diseases. *Journal of Cardiovascular Diseases and Cardiac Surgery* 3–9 (2017) doi:10.29199/cdcs.101012.
127. Liu, G. *et al.* PPAR δ agonist GW501516 inhibits PDGF-stimulated pulmonary arterial smooth muscle cell function related to pathological vascular remodeling. *BioMed Research International* **2013**, (2013).
128. Lobley, A., Sadowski, M. I. & Jones, D. T. pGenTHREADER and pDomTHREADER: new methods for improved protein fold recognition and superfamily discrimination. *Bioinformatics* **25**, 1761–1767 (2009).
129. P35555 (P35555) - protein - InterPro. http://www.ebi.ac.uk/interpro/protein/UniProt/P35555/.
130. Volkenstein, M. V. Coding of Polar and Non-polar Amino-acids. *Nature 1965 207:4994* **207**, 294–295 (1965).
131. Sawyer, S. L. *et al.* Utility of whole-exome sequencing for those near the end

of the diagnostic odyssey: Time to address gaps in care. *Clinical Genetics* vol. 89 275–284 (2016).

132. P78504 (P78504) - protein - InterPro. http://www.ebi.ac.uk/interpro/protein/UniProt/P78504/.

133. Fouillade, C., Monet-Leprêtre, M., Baron-Menguy, C. & Joutel, A. Notch signalling in smooth muscle cells during development and disease. *Cardiovascular Research* **95**, 138–146 (2012).

134. Manderfield, L. J. *et al.* Notch activation of Jagged1 contributes to the assembly of the arterial wall. *Circulation* **125**, 314–323 (2012).

135. De la Pompa JL, E. J. C. tissue interactions: N. signaling in cardiac development and disease. D. C. 2012;22(2):244-254. doi:10. 1016/j. devcel. 2012. 01. 014. Nihms353774. **22**, 244–254 (2013).

136. Loerakker, S. *et al.* Erratum: Mechanosensitivity of Jagged-Notch signaling can induce a switch-type behavior in vascular homeostasis (Proc. Natl. Acad. Sci. U.S.A. (2020) 115 (E3682-E3691) DOI: 10.1073/pnas.1715277115). *Proceedings of the National Academy of Sciences of the United States of America* **118**, (2021).

137. Balistreri, C. R. *et al.* Deregulation of Notch1 pathway and circulating endothelial progenitor cell (EPC) number in patients with bicuspid aortic valve with and without ascending aorta aneurysm. *Scientific Reports* vol. 8 (2018).

138. Monaghan, M. G. *et al.* Endocardial-to-mesenchymal transformation and mesenchymal cell colonization at the onset of human cardiac valve development. *Development (Cambridge)* **143**, 473–482 (2016).

139. Phillips, H. M. *et al.* Neural crest cells are required for correct positioning of the developing outflow cushions and pattern the arterial valve leaflets. *Cardiovascular Research* **99**, 452–460 (2013).

140. Saarinen, J., Kalkkinen, N., Welgus, H. G. & Kovanen, P. T. Activation of human interstitial procollagenase through direct cleavage of the Leu83-Thr84 bond by mast cell chymase. *Journal of Biological Chemistry* **269**, 18134–18140 (1994).

141. Tchougounova, E., Forsberg, E., Angelborg, G., Kjellén, L. & Pejler, G. Altered processing of fibronectin in mice lacking heparin. A role for heparin-dependent mast cell chymase in fibronectin degradation. *Journal of Biological Chemistry* **276**, 3772–3777 (2001).

142. Leskinen, M. J., Lindstedt, K. A., Wang, Y. & Kovanen, P. T. Mast cell chymase induces smooth muscle cell apoptosis by a mechanism involving fibronectin degradation and disruption of focal adhesions. *Arteriosclerosis, Thrombosis, and Vascular Biology* **23**, 238–243 (2003).

143. Raman, M., Chen, W. & Cobb, M. H. Differential regulation and properties of MAPKs. *Oncogene* **26**, 3100–3112 (2007).

144. Karin, M. & Gallagher, E. From JNK to pay dirt: Jun kinases, their biochemistry, physiology and clinical importance. *IUBMB Life* **57**, 283–295 (2005).

145. Q12913 (Q12913) - protein - InterPro. http://www.ebi.ac.uk/interpro/protein/UniProt/Q12913/.

146. Karagyozov, L., Grozdanov, P. N. & Böhmer, F. D. The translation attenuating arginine-rich sequence in the extended signal peptide of the

protein-tyrosine phosphatase PTPRJ/DEP1 is conserved in mammals. *PLOS ONE* **15**, e0240498 (2020).

147. Oshikawa, J. *et al.* Extracellular SOD-derived H2O2 promotes VEGF signaling in caveolae/lipid rafts and post-ischemic angiogenesis in mice. *PLoS ONE* **5**, (2010).

148. Fournier, P., Dussault, S., Fusco, A., Rivard, A. & Royal, I. Tyrosine phosphatase PTPRJ/DEP-1 Is an essential promoter of vascular permeability, angiogenesis, and tumor progression. *Cancer Research* **76**, 5080–5091 (2016).

149. Sanchez-Castro, M. *et al.* Search for Rare Copy-Number Variants in Congenital Heart Defects Identifies Novel Candidate Genes and a Potential Role for FOXC1 in Patients with Coarctation of the Aorta. *Circulation: Cardiovascular Genetics* **9**, 86–94 (2016).

150. Magouliotis, D. E. *et al.* In-depth bioinformatic study of the cadherin 5 interactome in patients with thoracic aortic aneurysm unveils 8 novel biomarkers. **00**, 1–8 (2021).

151. Akashi, M., Higashi, T., Masuda, S., Komori, T. & Furuse, M. A coronary artery disease-associated gene product, JCAD/KIAA1462, is a novel component of endothelial cell-cell junctions. *Biochemical and Biophysical Research Communications* **413**, 224–229 (2011).

152. Hara, T. *et al.* Targeted Disruption of JCAD (Junctional Protein Associated with Coronary Artery Disease)/KIAA1462, a Coronary Artery Disease-Associated Gene Product, Inhibits Angiogenic Processes in Vitro and in Vivo. *Arteriosclerosis, Thrombosis, and Vascular Biology* **37**, 1667–1673 (2017).

153. Xu, S. *et al.* The novel coronary artery disease risk gene JCAD/KIAA1462 promotes endothelial dysfunction and atherosclerosis. *European Heart Journal* **40**, 2398–2408 (2019).

154. Fan, J. *et al.* MCP-1, ICAM-1 and VCAM-1 are present in early aneurysmal dilatation in experimental rats. *Folia Histochemica et Cytobiologica* **48**, 455–461 (2010).

155. Jones, P. D. *et al. JCAD* , a Gene at the 10p11 Coronary Artery Disease Locus, Regulates Hippo Signaling in Endothelial Cells. *Arteriosclerosis, Thrombosis, and Vascular Biology* **38**, (2018).

156. Jiang, W. J. *et al.* Disruption of mechanical stress in extracellular matrix is related to Stanford type A aortic dissection through down-regulation of Yesassociated protein. *Aging* **8**, 1923–1939 (2016).

157. Liu, T. *et al.* YAP1 up-regulation inhibits apoptosis of aortic dissection vascular smooth muscle cells. *European review for medical and pharmacological sciences* **21**, 4632–4639 (2017).

158. Q9UNG2 (Q9UNG2) - protein - InterPro. http://www.ebi.ac.uk/interpro/protein/UniProt/Q9UNG2/.

159. Gao, J., Wang, S. & Liu, S. The involvement of protein TNFSF18 in promoting p-STAT1 phosphorylation to induce coronary microcirculation disturbance in atherosclerotic mouse model. *Drug Development Research* **82**, 115–122 (2021).

160. Blobe, G. C. *et al.* Functional Roles for the Cytoplasmic Domain of the Type III Transforming Growth Factor β Receptor in Regulating Transforming Growth Factor β Signaling. *Journal of Biological Chemistry* **276**, (2001).

161. Villarreal, M. M. *et al.* Binding Properties of the Transforming Growth Factor-β Coreceptor Betaglycan: Proposed Mechanism for Potentiation of Receptor Complex Assembly and Signaling. *Biochemistry* **55**, 6880–6896 (2016).

162. Groeneveld, M. E. *et al.* Betaglycan (TGFBR3) up-regulation correlates with increased TGF-β signaling in Marfan patient fibroblasts in vitro. *Cardiovascular Pathology* **32**, 44–49 (2018).

163. H3BV60 (H3BV60) - protein - InterPro. http://www.ebi.ac.uk/interpro/protein/UniProt/H3BV60/.

164. Diestel, U. *et al.* Identification of a Novel TGF-β-Binding Site in the Zona Pellucida C-terminal (ZP-C) Domain of TGF-β-Receptor-3 (TGFR-3). *PLOS ONE* **8**, e67214 (2013).

165. Neptune, E. R. *et al.* Dysregulation of TGF-β activation contributes to pathogenesis in Marfan syndrome. *Nature Genetics* **33**, 407–411 (2003).

166. Judge, D. P. *et al.* Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome. *Journal of Clinical Investigation* **114**, 172–181 (2004).

7. APPENDIX

 # Alignment/Mapping to Genome Using Burrows-Wheeler Aligner (BWA)

 # PCR duplicate removal Using Samtools and Picardtools

 # Base quality score recalibration Using BaseRecalibrator (GATK)

```
#! /bin/bash

#$ -N bamprep
#$ -M yjimenezb@udd.cl
#$ -m beas
```

```
#$ -cwd
source $HOME/.bash_profile
source $HOME/.bashrc

set -e
set -u
set -o pipefail

if [ $HOSTNAME == 'sofia.udd.cl' ] || [[ $HOSTNAME == computecpu-
1-*.local ]] || [[ $HOSTNAME == computegpus-1-*.local ]]
then
    genomes="/hpcudd/ICIM/shared/genomes"
    bundle="/hpcudd/ICIM/shared/gatk-bundle"
elif [ $HOSTNAME == 'mendel' ]
then
    genomes="/storage/shared/references"
    bundle="/storage/shared/gatk-bundle"
else
    echo "Unrecognized host $HOSTNAME"
    echo "can't locate genome references"
    exit 1
fi

while getopts '1:2:i:l:h' ARGS; do
       case "$ARGS" in
        1)
          read1="$OPTARG"
          ;;
        2)
          read2="$OPTARG"
          ;;
        i)
          RGID="$OPTARG"
          ;;
        l)
          RGLB="$OPTARG"
          ;;
        h)
          echo "script usage: $(basename $0) [-1 read1.fq] [-2
read2.fq] [-i readgroup_ID] [-l readgroup_LIB]" >&2
          exit 0
          ;;
        ?)
          echo "script usage: $(basename $0) [-1 read1.fq] [-2
read2.fq] [-i readgroup_ID] [-l readgroup_LIB]" >&2
          exit 1
```

```
          ;;
    esac
done

shift "$(($OPTIND - 1))"
mkdir -p ${RGID}_tmpdir

index="${genomes}/Homo_sapiens/Ensembl/GRCh37/Sequence/BWAIndex/ge
nome.fa"
refdata="${genomes}/Homo_sapiens/Ensembl/GRCh37"
genome="${refdata}/Sequence/WholeGenomeFasta/genome.fa"
dbsnp="${bundle}/b37/dbsnp_138.b37.vcf.gz"
indels="${bundle}/b37/Mills_and_1000G_gold_standard.indels.b37.vcf
.gz"

bwa mem                                                          \
    -t 8                                                         \
    -M                                                           \
    -R "@RG\tID:${RGID}\tLB:${RGLB}\tSM:${RGID}\tPL:ILLUMINA" \
    ${index}                                                     \
    ${read1}                                                     \
    ${read2}| samtools view -@ 2 -Sb -o ${RGID}_tmpdir/${RGID}.bam
- 2>/dev/null

exit_bwa=$?

cd ${RGID}_tmpdir

if [ $exit_bwa -eq 0 ] && [ -s ${RGID}.bam ]
then
       picard SortSam                    \
        I=${RGID}.bam               \
        O=sorted.${RGID}.bam        \
        SO=coordinate               \
        VALIDATION_STRINGENCY=SILENT \
        TMP_DIR=tmpdir              \
        VERBOSITY=ERROR             \
        QUIET=true

    exit_sort=$?
fi


if [ $exit_sort -eq 0 ] && [ -s sorted.${RGID}.bam ]
then
    rm -f ${RGID}.bam
```

```
    picard MarkDuplicates               \
        I=sorted.${RGID}.bam            \
        O=markDups.sorted.${RGID}.bam   \
        M=${RGID}.metrics.txt           \
        ASO=coordinate                  \
        VALIDATION_STRINGENCY=SILENT    \
        TMP_DIR=tmpdir                  \
        VERBOSITY=ERROR                 \
        QUIET=true                      \
        CREATE_INDEX=true

    exit_mkd=$?
fi


if [ $exit_mkd -eq 0 ] && [ -s markDups.sorted.${RGID}.bam ]
then
    rm -f sorted.${RGID}.bam
    rm -fr tmpdir

    gatk4 BaseRecalibrator            \
        -R $genome                     \
        -I markDups.sorted.${RGID}.bam \
        --knownSites $dbsnp             \
        --knownSites $indels            \
        -o markDups.sorted.${RGID}_ALLchr_recal_data.table

    exit_bqsr1=$?
fi

if [ $exit_bqsr1 -eq 0 ] && [ -s
markDups.sorted.${RGID}_ALLchr_recal_data.table ]
then
    gatk4 PrintReads                                          \
        -R $genome                                             \
        -I markDups.sorted.${RGID}.bam                         \
        -BQSR markDups.sorted.${RGID}_ALLchr_recal_data.table \
        -o bqsr_markDups.sorted.${RGID}.bam

    exit_bqsr2=$?
fi

if [ $exit_bqsr2 -eq 0 ] && [ -s bqsr_markDups.sorted.${RGID}.bam
]
then
```

```
    rm -f markDups.sorted.${RGID}.bam
    rm -f markDups.sorted.${RGID}.bai
fi
```

 # Realignment around indels Usign IndelRealigner (GATK

 # Variant Calling Using HaplotypeCaller (GATK)

```
#! /bin/bash

#$ -N lf2vc
#$ -M yjimenezb@udd.cl
#$ -m beas
#$ -cwd
source $HOME/.bash_profile
source $HOME/.bashrc

set -e
set -u
set -o pipefail

sample_name=${1}
genome="/hpcudd/ICIM/shared/genomes/Homo_sapiens/Ensembl/GRCh37/Se
quence/WholeGenomeFasta/genome.fa"

#Indels left align, final step of preprocessing raw data
samtools view -b bqsr_markDups.sorted.${sample_name}.bam |
bamleftalign -f $genome | samtools view -b - >
${sample_name}_preprocessed.bam
picard BuildBamIndex I=${sample_name}_preprocessed.bam
O=${sample_name}_preprocessed.bai VALIDATION_STRINGENCY=SILENT
TMP_DIR=/hpcudd/ICIM/yjimenez

#Variant Calling with GATK version 4.2 HaplotypeCaller
gatk HaplotypeCaller \
    -R $genome \
    -I ${sample_name}_preprocessed.bam \
    -O ${sample_name}.22XY.g.vcf \
    -ERC GVCF \
```

Variant filtering Using Variant Quality Score Recalibration (VQSR)

```
#! /bin/bash

#$ -N VQSR_SNP.sh
#$ -M yjimenezb@udd.cl
#$ -m beas
#$ -cwd
source $HOME/.bash_profile
source $HOME/.bashrc

set -e
set -u
set -o pipefail


bundle="/hpcudd/ICIM/shared/gatk-bundle/b37"
hapmap="${bundle}/hapmap_3.3.b37.vcf.gz"
omni="${bundle}/1000G_omni2.5.b37.vcf.gz"
g1k="${bundle}/1000G_phase1.snps.high_confidence.b37.vcf.gz"
dbsnp="${bundle}/dbsnp_138.b37.vcf.gz"
mills="${bundle}/Mills_and_1000G_gold_standard.indels.b37.vcf.gz"

#Variant Recalibrator VQSR
~/gatk-4.1.9.0/gatk VariantRecalibrator \
   --variant CAS-01-001.22XY.g.vcf.raw.vcf --variant CAS-01-
002.22XY.g.vcf.raw.vcf --variant CAS-01-003.22XY.g.vcf.raw.vcf --
variant CAS-01-004.22XY.g.vcf.raw.vcf --variant CAS-01-
005.22XY.g.vcf.raw.vcf --variant CAS-01-006.22XY.g.vcf.raw.vcf --
variant CAS-01-007.22XY.g.vcf.raw.vcf --variant CAS-01-
015.22XY.g.vcf.raw.vcf --variant CAS-01-016.22XY.g.vcf.raw.vcf --
variant CAS-01-019.22XY.g.vcf.raw.vcf --variant CAS-01-
020.22XY.g.vcf.raw.vcf --variant CAS-01-021.22XY.g.vcf.raw.vcf --
variant CAS-01-022.22XY.g.vcf.raw.vcf --variant CAS-01-
023.22XY.g.vcf.raw.vcf --variant CAS-01-024.22XY.g.vcf.raw.vcf --
variant CAS-01-025.22XY.g.vcf.raw.vcf --variant CAS-01-
026.22XY.g.vcf.raw.vcf --variant CAS-01-029.22XY.g.vcf.raw.vcf --
variant CAS-01-030.22XY.g.vcf.raw.vcf --variant CAS-01-
031.22XY.g.vcf.raw.vcf --variant CAS-01-010.22XY.g.vcf.raw.vcf --
```

```
variant CAS-01-011.22XY.g.vcf.raw.vcf --variant CAS-01-
027.22XY.g.vcf.raw.vcf --variant CAS-01-034.22XY.g.vcf.raw.vcf --
variant CAS-01-035.22XY.g.vcf.raw.vcf --variant CAS-01-
036.22XY.g.vcf.raw.vcf --variant CAS-01-041.22XY.g.vcf.raw.vcf --
variant CAS-01-043.22XY.g.vcf.raw.vcf --variant CAS-01-
044.22XY.g.vcf.raw.vcf --variant CAS-01-045.22XY.g.vcf.raw.vcf --
variant CAS-01-046.22XY.g.vcf.raw.vcf --variant CAS-01-
047.22XY.g.vcf.raw.vcf --variant CAS-01-048.22XY.g.vcf.raw.vcf --
variant CAS-01-049.22XY.g.vcf.raw.vcf \
   -tranche 100.0 -tranche 99.95 -tranche 99.9 -tranche 99.5 -
tranche 99.0 -tranche 97.0 -tranche 96.0 -tranche 95.0 -tranche
94.0 -tranche 93.5 -tranche 93.0 -tranche 92.0 -tranche 91.0 -
tranche 90.0 \
   --
resource:hapmap,known=false,training=true,truth=true,prior=15.0
${hapmap} \
   --
resource:omni,known=false,training=true,truth=false,prior=12.0
${omni} \
   --
resource:1000G,known=false,training=true,truth=false,prior=10.0
${g1k} \
   --
resource:dbsnp,known=true,training=false,truth=false,prior=2.0
${dbsnp} \
   --use-annotation QD \
   --use-annotation FS \
   --use-annotation SOR \
   --use-annotation MQ \
   --use-annotation MQRankSum \
   --use-annotation ReadPosRankSum \
   --mode SNP \
   -O output_SNP.recal \
   --tranches-file output_SNP.tranches \
   --rscript-file output_SNP_plots.R



#! /bin/bash

#$ -N VQSR_INDEL.sh
#$ -M yjimenezb@udd.cl
#$ -m beas
#$ -cwd
source $HOME/.bash_profile
```

```
source $HOME/.bashrc

set -e
set -u
set -o pipefail


bundle="/hpcudd/ICIM/shared/gatk-bundle/b37"
hapmap="${bundle}/hapmap_3.3.b37.vcf.gz"
omni="${bundle}/1000G_omni2.5.b37.vcf.gz"
g1k="${bundle}/1000G_phase1.snps.high_confidence.b37.vcf.gz"
dbsnp="${bundle}/dbsnp_138.b37.vcf.gz"
mills="${bundle}/Mills_and_1000G_gold_standard.indels.b37.vcf.gz"

#Variant Recalibrator VQSR
~/gatk-4.1.9.0/gatk VariantRecalibrator \
   --variant CAS-01-001.22XY.g.vcf.raw.vcf --variant CAS-01-
002.22XY.g.vcf.raw.vcf --variant CAS-01-003.22XY.g.vcf.raw.vcf --
variant CAS-01-004.22XY.g.vcf.raw.vcf --variant CAS-01-
005.22XY.g.vcf.raw.vcf --variant CAS-01-006.22XY.g.vcf.raw.vcf --
variant CAS-01-007.22XY.g.vcf.raw.vcf --variant CAS-01-
015.22XY.g.vcf.raw.vcf --variant CAS-01-016.22XY.g.vcf.raw.vcf --
variant CAS-01-019.22XY.g.vcf.raw.vcf --variant CAS-01-
020.22XY.g.vcf.raw.vcf --variant CAS-01-021.22XY.g.vcf.raw.vcf --
variant CAS-01-022.22XY.g.vcf.raw.vcf --variant CAS-01-
023.22XY.g.vcf.raw.vcf --variant CAS-01-024.22XY.g.vcf.raw.vcf --
variant CAS-01-025.22XY.g.vcf.raw.vcf --variant CAS-01-
026.22XY.g.vcf.raw.vcf --variant CAS-01-029.22XY.g.vcf.raw.vcf --
variant CAS-01-030.22XY.g.vcf.raw.vcf --variant CAS-01-
031.22XY.g.vcf.raw.vcf --variant CAS-01-010.22XY.g.vcf.raw.vcf --
variant CAS-01-011.22XY.g.vcf.raw.vcf --variant CAS-01-
027.22XY.g.vcf.raw.vcf --variant CAS-01-034.22XY.g.vcf.raw.vcf --
variant CAS-01-035.22XY.g.vcf.raw.vcf --variant CAS-01-
036.22XY.g.vcf.raw.vcf --variant CAS-01-041.22XY.g.vcf.raw.vcf --
variant CAS-01-043.22XY.g.vcf.raw.vcf --variant CAS-01-
044.22XY.g.vcf.raw.vcf --variant CAS-01-045.22XY.g.vcf.raw.vcf --
variant CAS-01-046.22XY.g.vcf.raw.vcf --variant CAS-01-
047.22XY.g.vcf.raw.vcf --variant CAS-01-048.22XY.g.vcf.raw.vcf --
variant CAS-01-049.22XY.g.vcf.raw.vcf \
   -tranche 100.0 -tranche 99.95 -tranche 99.9 -tranche 99.5 -
tranche 99.0 -tranche 97.0 -tranche 96.0 -tranche 95.0 -tranche
94.0 -tranche 93.5 -tranche 93.0 -tranche 92.0 -tranche 91.0 -
tranche 90.0 \
   --resource:mills,known=false,training=true,truth=true,prior=12
${mills} \
```

```
   --
resource:dbsnp,known=true,training=false,truth=false,prior=2.0
${dbsnp} \
   --use-annotation QD \
   --use-annotation FS \
   --use-annotation SOR \
   --use-annotation MQ \
   --use-annotation MQRankSum \
   --use-annotation ReadPosRankSum \
   --mode INDEL \
   -O output_INDEL.recal \
   --tranches-file output_INDEL.tranches \
   --rscript-file output_INDEL_plots.R \


#$ -m beas
#$ -cwd
source $HOME/.bash_profile
source $HOME/.bashrc

set -e
set -u
set -o pipefail

sample_name=${1}
~/gatk-4.1.9.0/gatk ApplyVQSR \
   -R
/hpcudd/ICIM/shared/genomes/Homo_sapiens/Ensembl/GRCh37/Sequence/W
holeGenomeFasta/genome.fa \
   -V ${sample_name} \
   -O ${sample_name}.SNP_Filtered.vcf \
   --tranches-file output_SNP.tranches \
   --recal-file output_SNP.recal \
   --mode SNP


#$ -m beas
#$ -cwd
source $HOME/.bash_profile
source $HOME/.bashrc

set -e
set -u
set -o pipefail

sample_name=${1}
```

```
~/gatk-4.1.9.0/gatk ApplyVQSR \
   -R
/hpcudd/ICIM/shared/genomes/Homo_sapiens/Ensembl/GRCh37/Sequence/W
holeGenomeFasta/genome.fa \
   -V ${sample_name} \
   -O ${sample_name}.SNP_INDEL_Filtered.vcf \
   --tranches-file output_INDEL.tranches \
   --recal-file output_INDEL.recal \
   --mode INDEL
```

 # Variant Annotation Using ANNOVAR

```
#! /bin/bash
#
# Usage: annovar
#

# -- our name ---
#$ -N annovar
#$ -S /bin/sh
# Make sure that the .e and .o file arrive in the
# working directory
#$ -cwd
#Merge the standard out and standard error to one file
#$ -j y
/bin/echo Running on host: `hostname`.
/bin/echo In directory: `pwd`
/bin/echo Starting on: `date`
#$ -m beas
#$ -M yjimenez@udd.cl
#Para activar los requerimientos del sistema
source $HOME/.bash_profile
source $HOME/.bashrc

input=${1}
name=$(echo $input|cut -d "." -f 1)

perl ../annovar/table_annovar.pl \
```

```
     ${input} \
     ../annovar/humandb  \
    -buildver hg19 -remove     \
    -out ${input}.annovar.tmp  \
    -protocol
refGene,1000g2015aug_all,exac03,esp6500siv2_all,avsnp147,clinvar_2
0190305,dbnsfp33a,gnomad_exome,gnomad_genome  \
    -operation g,f,f,f,f,f,f,f,f -nastring . -vcfinput

n=$(head -1 ${input}.annovar.tmp.hg19_multianno.txt|datamash
transpose|wc -l)
t=$(tail -n +2 ${input}.annovar.tmp.hg19_multianno.txt|head -
1|datamash transpose|wc -l)
x=$(expr ${t} - ${n})
paste <(head -1 ${input}.annovar.tmp.hg19_multianno.txt) <(for i
in $(seq 1 $x);do echo "Otherinfo";done|datamash transpose) >
${input}.annovar.tmp.header
cat ${input}.annovar.tmp.header <(tail -n +2
${input}.annovar.tmp.hg19_multianno.txt) > ${input}_annovar.txt
cp ${input}_annovar.txt ${name}_annovar.txt

rm -f ${input}.annovar.tmp.*
```

 # VAAST 2.0

 # vaast_converter

```
VAAST/bin/vaast_tools/vaast_converter --build hg19 name.vcf
```

 # Variant Annotation Tool (VAT)

```
VAT --features genes.gff3 --fasta assembly.fasta
variants.gvf > variants.vat.gvf
```

Variant Selection Tool (VST)

 # I: Intersection (Variants shared by all files)

 # U: Union (All variants)

 # C: Complement (Variants unique to the first file

 # Fam 1_all

D: C(U(II-1, II-3,III-6), U(III-1, III-8)

R: C(U(II-1, II-3,III-6), U(III-1, III-8)

 # Fam 1.1

D: C(I(III-1, II-2), II-1)

R: I(III-1, U(II-1, II-2))

 # Fam 1.2

D: C(I(II-4, III-8) , U(II-3, III-6))

R: I(III-8, U(II-3, II-4) , C(III-8, III-6))

 # Fam 3

D: C(I(II-1, I-1) , I-2)

R: I(II-1, U(I-1, I-2))

 # Fam 5

D: C(II-8, I(I-10, II-7))

R: C(II-8, I(I-10, II-7))

 # Severe

D: C(U(Severe) , U(Mild))

R: C(U(Severe) , U(Mild))

 # Mild

D: C(U(Mild), U(Severe))

R: C(U(Mild), U(Severe))

 # VAAST: Dom or Rec

COMMAND: VAAST -m lrt -iht d or r -j 0.0000026 -o name_output -d 1e7 -p 20 RefSeq_hg19.p10_VAAST.gff3 1KGv3_CG_Div_NHLBI_dbSNP_RefSeq_hg19.cdr name.cdr