ANALYSIS OF THE *TGFBR1* GENE AS A CANDIDATE GENE IN MARFAN SYNDROME AND RELATED DISORDERS PATIENTS, NEGATIVE FOR *FBN1* AND *TGFBR2* MUTATIONS

(ANALISIS GEN TGFBR1 SEBAGAI GEN KANDIDAT PADA PASIEN SINDROMA MARFAN DAN KELAINAN TERKAIT LAINNYA, TANPA MUTASI PADA GEN FBN1 DAN TGFBR2)

Thesis
Submitted to fulfil the assignment and fit-out requisite
in passing Post-graduate Program Majoring Genetics Counseling
Diponegoro University Semarang

Nani Maharani

G4A006011

Biomedical Science Post Graduate Program
Majoring Genetics Counseling
Diponegoro University Semarang
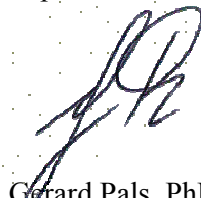**2009**

APPROVAL SHEET

THESIS

ANALYSIS OF THE *TGFBR1* GENE AS A CANDIDATE GENE IN MARFAN SYNDROME AND RELATED DISORDERS PATIENTS, NEGATIVE FOR *FBN1* AND *TGFBR2* MUTATIONS
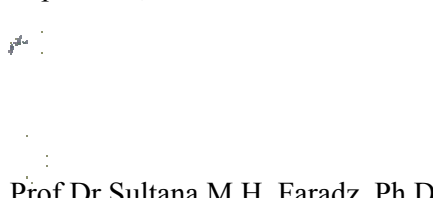
By
Nani Maharani
G4A006011

Has been defended in front of the defence committee
On January 6th, 2009
and has been approved by

Supervisor   Supervisor,   Supervisor

Gerard Pals, PhD[1]   Prof.Dr.Sultana M.H. Faradz, Ph.D   Prof. Ben CJ Hamel, PhD[2]
                      NIP: 130 701 415

Recognition,
Head of Master's degree program in
Biomedical Sciences

DR. dr. Winarto, DMM, SpMK, SpM(K)
NIP. 130 675 157

1) Head of Connective Tissue Research, Department of Clinical Genetics Vrije Universiteit Medisch Centrum, The Netherlands
2) Head of Division of Clinical Genetics, Department of Human Genetics Radboud University Medical Centre, The Netherlands

DECLARATION

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, exept where due acknowledgement is made in the text

Nani Maharani

January, 2009

CURRICULUM VITAE

Personal Data

Name	:	Nani Maharani, dr
Address	:	Jl. Dewi Sartika Raya No. 35 Semarang 50221
Cell phone	:	+6281325780633
Place & Date of Birth	:	Semarang / November 12th, 1981
Sex / marital status	:	Female / single

Educational Background

2007 – present	Post Graduate Program Diponegoro University, Master in Biomedical Science Majoring Genetic Counseling (Twinning Program with Vrije Universiteit Amsterdam, The Netherlands)
2004 – 2006	Diponegoro University, Medical Faculty (Medical Doctor)
2000 – 2004	Diponegoro University, Medical Faculty (Bachelor Degree)
1997 – 2000	High School at SMU N 3 Semarang majoring Natural Science
1994 – 1997	Junior High School at SMP N 3 Semarang
1988 – 1994	Elementary School at SD Petompon I Semarang

Training and Course

2007, Sept 1st	Workshop Early Detection on Neurodevelopmental Disorders (Certificate from Ikatan Dokter Anak Indonesia/Bagian Ilmu Kesehatan Anak FK UNDIP-RSUP Dr KARIADI – Pusat Riset Biomedik FK UNDIP)
2007, Jan 26th	Medical Genetic Course : From Basic to Clinic (Certificate from Medical Faculty Diponegoro University Semarang – Radboud University Medical Centre The Netherland)
2006, Nov 3rd-5th	Advanced Cardiac Life Support Course (Certificate from Indonesian Heart Association

Working Experience and Internship

2008 – present	Secretary of Working Group on Sexual Ambiguity Center for Biomedical Research (CEBIOR) Medical Faculty of Diponegoro University
2006 - present	Education staff in Pharmacology and Therapeutics Department Medical Faculty of Diponegoro University
2007 – 2008	Student Assistant in Parasitology Department Medical Faculty Diponegoro University

ACKNOWLEDGEMENT

I would also like to gratefully acknowledge the guidance and tuition of all my teachers and advisors in Genetic Counseling (Master Program of Biomedical Sciences Diponegoro University). Particularly I would like to acknowledge with appreciation to Dr. Tri Indah Winarni, MsiMed, Dr. Asri Purwanti, SpA(K), Dr. MA Sungkar SpPD SpJP, for the guidance to help me build the basic in Marfan Syndrome research and the research in general.

My deep and sincere thanks also go to the Rector of Diponegoro University Prof. Dr. dr. Susilo Wibowo, MSi.Med, SpAnd, the former Head of Biomedical Science Post Graduate Program of Diponegoro University Prof. dr. Soebowo, SpPA(K), for the opportunity to join this master degree; the present Head of Biomedical Science Post Graduate Program of Diponegoro University DR. dr. Winarto, SpM(K), and the Dean of Medical Faculty Diponegoro dr. Soejoto, SpKK(K) for the recommendation, the opportunity and great support in this study.

I will always be grateful to all the staf of DNA Laboratory VUMC Amsterdam for their kindly help, cooperation and discussions on lab works. Especially to people in connective tissue disorders group Eline Zwikstra, Margriet Smith, Marian Muijs, Eric van den Akker and Linda. My thanks also go to my colleagues Meredith Kressenberg, Youssef Moutouakil, Umit Baylan and Rob van Andel for the guidance and friendly discussion during the working hours.

I am also grateful to all the staf of Centre for Biomedical Research, Semarang, Indonesia, particularly to Mrs. Wiwik Lestari, Mrs. Lusi Suwarsi, Mrs.

ABBREVIATIONS

AA	Amino acid
Po	Polar
NPo	Non-polar
N	Neutral
B	Basic
A	Acidic
ACTA2	Actin alpha 2
ALK1	Activin receptor-like kinase type 1
ALK5	Activin receptor-like kinase type 5
CT-scanning	Computed tomography scanning
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
ECM	Extracellular matrix
FBN1	Fibrillin 1
FBN2	Fibrillin 2
FH	Family history
MFS	Marfan Syndrome
LDS	Loeys-Dietz Syndrome
LLC	Large Latent Complex
LTBP	Latent TGFβ binding protein
LTBP4	Latent TGFβ binding protein type 4
MASS phenotype	Mitral valve prolaps, aortic root diameter at upper limits of normal for body size, stretch marks of the skin and skeletal conditions similar to Marfan Syndrome phenotype
MRI	Magnetic Resonance Imaging
MYH11	Myosin heavy chain 11
PCR	Polymerase Chain Reaction
PolyPhen	Polymorphism Phenotyping

PSIC score	Position-specific Independent Counts
SIFT	Sorting Intolerance From Tolerance
M	Median sequence conservation
S	Sequences represented at this position
SLC	Small latent complex
TGF-β	Transforming growth factor beta
TGFBR1	Transforming growth factor beta receptor type 1
TGFBR2	Transforming growth factor beta receptor type 2
TAAD	Thoracic aortic aneurysms and dissections

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ATTACHMENTS

ABSTRACT

Background

Marfan Syndrome (MFS) and related disorders involves particularly skeletal, ocular and cardiovascular. Aortic aneurysms and dissections is the commonest feature of MFS leading to death. MFS caused by mutation in FBN1, and recently, also in TGFBR2 and TGFBR1. Mutation analysis in TGFBR1 gene is needed to know if the mutation is present in patient with MFS and related disorders.

Methods

One hundred and ninety four patients with MFS and related disorders, who have at least one major criteria of MFS and found to be negative for FBN1 and TGFBR2 mutation, are included. The DNA of the patients were then analyzed for TGFBR1 mutation by direct sequencing of the whole gene. The potency of pathogenicity of the mutation was predicted by referring to previous publication, amino acid changes, multiple alignment analysis and with the help of internet-based software, PolyPhen and SIFT.

Results

Ten patients were found to carry TGFBR1 missense mutation. Each of them carried a different mutation, except 2 patients carried the same mutation. Seven out of nine of the mutations are considered pathogenic and 2 are not pathogenic. Aortic aneurysm is present in most patients with the mutation. None of the patient with classic MFS has mutation in TGFBR1 gene.

Conclusion

Despite of mutation analysis on FBN1 and TGFBR2, mutation analysis on TGFBR1 in patient with MFS and related disorders is needed, especially on those who have aortic aneurysm. Knowledge of the presence of a mutation in an individual or in a family, may give a better guidance for comprehensive treatment including genetic counseling

Keywords : Marfan Syndrome and related disorders, TGFBR1 mutation

ABSTRAK

Latar Belakang
Sindroma Marfan *(MFS)* dan kelainan terkait bermanifestasi di beberapa organ, terutama skeletal, okular dan kardiovaskular. Aneurysma dan diseksi aorta merupakan manifestasi yang paling sering mengakibatkan kematian pada *MFS*. *MFS* disebabkan oleh mutasi pada *FBN1*, dan akhir-akhir ini ditemukan juga disebabkan mutasi pada *TGFBR2* dan *TGFBR1*. Analisis pada gen *TGFBR1* diperlukan untuk mengetahui apakah pada pasien Marfan Syndrome dan kelainan terkait lainnya terdapat mutasi pada gen *TGFBR1*.

Metode
Sebanyak 194 pasien dengan *MFS* dan kelainan terkait yang memiliki paling tidak satu kelainan mayor diikutsertakan dalam penelitian ini. Sebelumnya, pasien telah terbukti tidak memiliki mutasi pada *FBN1* dan *TGFBR2*. Sekuensing pada gen *TGFBR1* dilakukan untuk mengetahui adanya mutasi. Potensi patogenisitas mutasi dianalisis dengan mengacu pada publikasi-publikasi sebelumnya, melihat perubahan asam amino, melakukan *multiple alignment analysis* dan menggunakan *software* PolyPhen dan SIFT.

Hasil
Didapatkan 10 pasien dengan mutasi pada *TGFBR1,* dari keseluruhan pasien yang diperiksa. Setiap pasien memiliki 1 m*issense mutation* yang berbeda, kecuali 2 pasien dengan mutasi yang sama. Dari 9 *missense mutations* pada *TGFBR1,* 7 diantaranya patogenik dan 2 nonpatogenik. Aneurisma aorta merupakan manifestasi klinik yang muncul pada hampir semua pasien dengan mutasi. Mutasi pada *TGFBR1* tidak ditemukan pada pasien dengan *MFS* klasik.

Kesimpulan
Analisis mutasi *TGFBR1* pada *MFS* dan kelainan terkait tanpa mutasi di *FBN1* dan *TGFBR2* perlu dilakukan, terutama pada pasien dengan aneurisma aorta. Pengetahuan tentang keberadaan mutasi pada individu dalam keluarga dapat menjadi petujuk penting untuk penanganan yang komprehensif termasuk konseling genetika.

Kata kunci : Sindroma Marfan dan kelainan terkait, mutasi *TGFBR1*

Chapter I

INTRODUCTION

I.1 Background

Marfan Syndrome (MFS), a common autosomal dominant inherited disorder of fibrous connective tissue, has an estimated incidence of 1 : 5,000.[1,2] This syndrome involves many organ systems, particularly the skeletal, ocular and cardiovascular system. The most important life-threatening complication in MFS is the occurrence of thoracic aortic aneurysms leading to aortic dissection, rupture, or both.[3]

MFS is known to be one of the diseases in the spectrum of type-1 fibrillinopathies, which constitute a range of clinical phenotypes that are caused by mutation in the gene for fibrillin-1 (*FBN1* gene).[1,2,4] In many cases, a diagnosis of MFS can be established by the Ghent criteria.[5] However, the interpretation of these criteria is not always easy, due to the large clinical range of fibrillinopathies that overlap with MFS, and to age-dependent manifestations.

The initial idea from previous publications about the pathogenesis of MFS concentrated on a static dominant negative model based on the concept of fibrillin-rich micro fibrils as purely architectural elements in the extra cellular matrix. Mutations in the fibrillin-1 gene (*FBN1* gene), known to cause MFS, however, have not always been found in MFS patients. Recent

findings of the pathogenesis of MFS demonstrate changes in growth factor signaling and other changes in matrix-cell interactions.[4]

A connection of Marfan syndrome with the TGFβ signalling pathway was initially found in a study on mouse model of Marfan Syndrome with *FBN1* mutation, and having lung emphysema as phenotypic manifestation. This mouse model showed increased TGFβ signalling.[6] The involvement of TGFβ-receptor gene mutation in MFS has been shown in a Japanese patient with MFS who had a balanced chromosomal translocation involving chromosome 3p24. This locus had been found to show genetic linkage with MFS in a large French pedigree. The breakpoint in the Japanese patient disrupted the *TGFBR2* gene. The same gene had a point mutation in the French Marfan family.[7] Later research on TGFβ showed that the use of TGFβ antagonists such as TGFβ neutralizing antibody or the angiotensin II type 1 receptor blocker, Losartan, reduce the growth of aortic aneurysm in a mouse model.[8]

The proteins fibrillin-1, TGFBR1 and TGFBR2 take part in transforming growth factors β (TGFβ) signaling, thus mutations in one of these gene could cause similar phenotypes. TGFβ is stored in the extra cellular matrix in a latent form, bound to fibrillin 1 to form a complex. The complex is released by proteases, and the active TGFβ binds to its receptors on the cell surface (TGFβR1 and TGFβR2), leading to dimerization of the receptor. The kinase domain of the receptor is then activated and starting a signaling cascade in the cell regulating a number of cellular processes such as

apoptosis, inflammation, proliferation and growth.[9] Thus, TGFβ signaling will depend on the amount of latent TGFβ present in the tissue, strength of the binding of the complex and activity of TGFβ receptors.

Mutations in the *TGFBR1* and *TGFBR2* genes have also been reported in individuals with Loeys-Dietz aortic aneurysms syndrome, a syndrome characterized by hypertelorism, bifid uvula and/or cleft palate, generalized arterial tortuosity with ascending aortic aneurysm, and worse cardiovascular risk profile than classic MFS.[10] Another study reported *TGFBR1* and *TGFBR2* mutations in individuals with MFS-like phenotypes who previously tested negative for mutations in *FBN1* gene.[11] Mutations in *TGFBR1* have been found in other syndromes related with MFS, e.g. Sphrintzen-Goldberg Syndrome, and in patients with Thoracic Aortic Aneurysms and Dissection (TAAD).[6,11,12] So far, in total 22 different mutations have been found in the *TGFBR1* gene.[13] The phenotypes of patients having the mutations in *TGFBR* genes could not be clearly differentiated from each other.

In this descriptive research we looked for and analyzed mutations in the *TGFBR1* gene in patients referred to the DNA laboratory of Vrije Universiteit Medisch Centrum Amsterdam (VUmc), The Netherlands, with a clinical suspicion of MFS or related disorders, who did not have a *FBN1* or *TGFBR2* mutation.

I.2 Research Questions

I.2.1 General research question :

What kind of mutations can be found in the *TGFBR1* gene in people with clinical Marfan Syndrome, and other related disorders with negative *FBN1* and *TGFBR2* mutations?

I.2.2 Specific research question

1. Is there any mutation in the *TGFBR1* gene as a candidate gene for Marfan Syndrome and related disorders with negative *FBN1* and *TGFBR2* mutations, and if yes, what kind of mutation is it?
2. How is the prediction of pathogenicity of the mutation?
3. How is the distribution of clinical phenotype on genotype?
4. Is there any clinical characteristic that may lead to *TGFBR1* gene mutation analysis?

I.3 Research purposes

I.3.1 General purposes :

To identify and analyze the kind of mutations in the *TGFBR1* gene as candidate gene for Marfan Syndrome and related disorders with negative *FBN1* and *TGFBR2* mutations, and to see the distribution of clinical phenotype on the genotype .

I.3.2 Specific purposes :

1. To detect the mutations in the *TGFBR1* gene in a person with Marfan Syndrome and related disorders with negative *FBN1* and *TGFBR2* mutations.

2. To analyze the kind of mutations and the potential pathogenic effect of the mutations.

3. To see the distribution of clinical phenotype on the genotype.

4. To see whether there is a clinical characteristic that may lead to *TGFBR1* mutation analysis.

Chapter II

LITERATURE REVIEW

II.1 MARFAN SYNDROME AND RELATED DISORDERS

Patients with Marfan Syndrome (MFS) may have abnormalities in several different organ systems, but mostly in skeletal, ocular and cardiovascular systems.[1] Skeletal features of MFS are increased height, disproportionately long limbs and digits, elbow contracture, anterior chest deformity, mild to moderate joint laxity, vertebral column deformity (scoliosis and thoracic lordosis) and a narrow, high palate with crowding of the teeth. Ocular findings in MFS include increased axial globe length, corneal flatness and (sub) luxation of the lenses (ectopia lentis). Mitral valve prolaps, mitral regurgitation, dilatation of the aortic root and aortic regurgitation are cardiovascular features. Aneurysm of the aorta and aortic dissection are the major life-threatening cardiovascular complications. Mostly, this feature brings MFS into special attention. Other common features are striae distensae, pulmonary blebs, which predispose to spontaneous pneumothorax and spinal arachnoid cysts or diverticula. By CT or MRI scanning also dural ectasia can be found. The early-onset severe MFS, neonatal MFS, presents with serious cardiovascular abnormalities as well as congenital contractures. MFS is also associated with a high prevalence of obstructive sleep apneu.[1,2,14,15]

The diagnosis of MFS is based on a set of clinical diagnostic criteria, termed The Ghent Criteria.[5] In clinical practice, these criteria are not always

obvious, since there are many conditions overlapping with MFS and because of age-dependent manifestation. The overlapping conditions are Familial Aortic Aneurysm, Bicuspid Aortic Valve with Aortic Dilatation, Familial Ectopia Lentis, MASS phenotype, Marfan Body Type, Mitral Valve Prolapse Syndrome, Congenital Contractural Arachnodactily (Beals syndrome), Stickler syndrome, Shprintzen-Goldberg Syndrome, Loeys-Dietz Syndrome and Ehlers-Danlos syndrome.[1,2,4,14,15] The clinical features of those overlap disorders are described in the table below :

Table 1. Clinical features of some overlapping disorders

No.	Disorders	Clinical Features
1.	Familial Aortic Aneurysm	Aortic aneurysms, aortic dissection, familial
2.	Loeys-Dietz Syndrome	Widely-spaced eyes (hypertelorism), bifid uvula, generalized arterial tortuosity with widespread arterial aneurysms and dissection
3.	Ehlers-Danlos syndrome	Skin hyperextensibility, joint hypermobility, easy bruising, tissue fragility, mitral valve prolapse, aortic dilatation (uncommon)
4.	MASS phenotype	Mitral valve prolapse, aortic root diameter at the upper limit of normal, stretch mark (striae), skeletal features of Marfan (joint hypermobility, pectus excavatum/carinatum, scoliosis)
5.	Marfan Body Type	Tall, long-thin arms & leg, long-thin fingers, scoliosis, hypermobility of the joint
6.	Mitral Valve Prolapse Syndrome	Mitral valve prolapse
7.	Congenital Contractural Arachnodactily	Joints contracture, crumpled ears, arachnodactily, scoliosis, kyphoscoliosis, osteopenia, dolichostenomelia, pectus excavatum or carinatum, muscular hypoplasia, micrognathia, high-arched palate
8.	Shprintzen-Goldberg Syndrome	Omphalocele, scoliosis, laryngeal/pharyngeal hypoplasia, mild dysmorphic face, learning disabilities
9.	Familial Ectopia Lentis	Ectopia lentis, with the signs of myopia, astigmatisms, and blur vision

This table shows some of the disorders that have overlapping phenotypes with Marfan Syndrome.

II.2 TGFβR1, *TGFBR1* GENE AND CONTROL OF TGFβ SIGNALLING

Fibrillin and TGFβR are taking part in the TGFβ signalling pathway. Fibrillin is the major constitutive element of the extracellular microfibrils which has a crucial role in regulating TGFβ bioavailability in the vascular system.[4] The bioavailability of active TGFβ is regulated at multiple levels, including secretion and interaction with extra cellular matrix components.
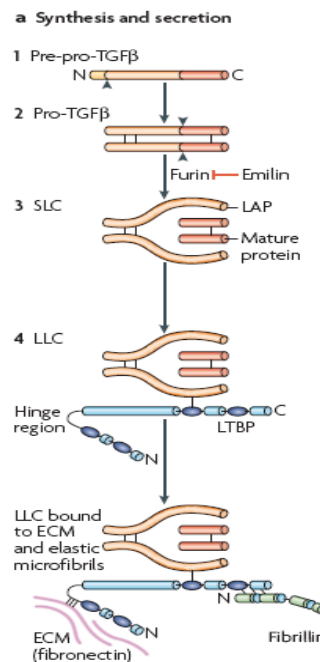
Figure 1. Regulation of TGFβ bioavailability *(taken from Nature Reviews on Molecular Cell Biology 2007)*

Synthesis and secretion (a) : *TGFβ is synthesized as a pre-pro-protein, which undergoes proteolytic processing in the rough endoplasmic reticulum (1). Two monomers of TGFβ dimerize through disulfide bridges (2). The pro- TGFβ dimer is then cleaved by furin convertase to yield the small latent TGFβ complex (SLC), in which the latency-associated*

peptide (LAP) and the mature peptide are connected (3). This processing step is inhibited by emilin-1. The large latent TGFβ binding protein (LTBP) is attached, and form the large latent TGFβ complex (LLC) (4). The N-terminal and hinge region of LTBP interact covalently with extra cellular matrix component, such as fibronectin. The C-terminal region of LTBP interacts non covalently with the N-terminal region of fibrillin-1.[9]
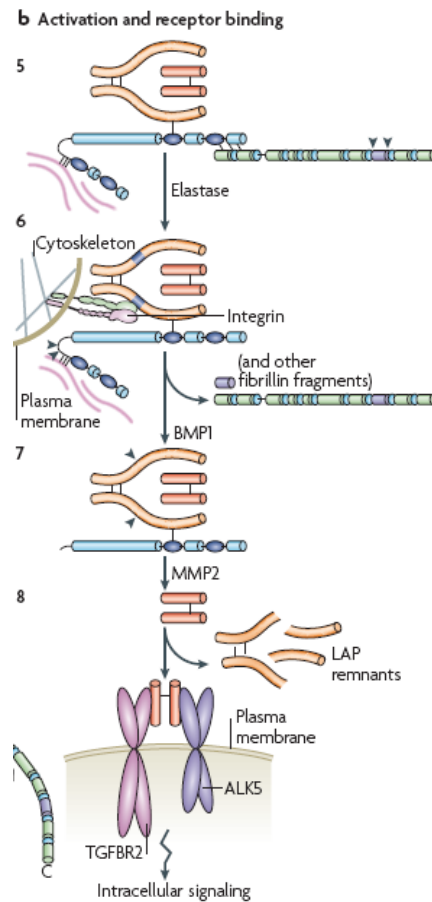
Figure 2. Regulation of TGFβ bioavailability--continued *(taken from Nature Reviews on Molecular Cell Biology 2007)*

Activation and receptor binding (b) : *An internal fragment of fibrillin-1 released by proteolysis mediated by elastases at sites (indicated with black arrowheads) (5), interacts with N-terminal region of fibrillin-1 to displace LTBP and release LLC (6). The LLC can be targeted to the cell surface by binding to integrins via RGD sequence (blue regions) in LAP. Bone morphogenetic protein-1 (BMP1) can cleave two sites in the hinge region of LTBP, which results in the release of LLC (7). Matrix metalloprotease-2 (MMP2) and other proteases can cleave LAP to release mature TGFβ (red). Mature TGFβ can then bind to its receptors, TGFβR2 and TGFβR1.[9]*

Transforming growth factor-β plays a pivotal role in vascular remodeling and the resolution process of angiogenesis. TGFβ regulates cellular processes by binding to a heterodimeric complex of the type I and type II serine/threonine kinases receptors (TGFβR1 and TGFβR2). Once the active TGFβ family member is released from the extra cellular matrix, it signals via the receptors, the TGFβR2 and TGFβR1 (also known as ALK5; a type I receptor).

The type I receptor acts downstream of the type II receptor and propagates the signal to the nucleus by phosphorylating specific members of the SMAD family, receptor-regulated(R)-Smads.[9]
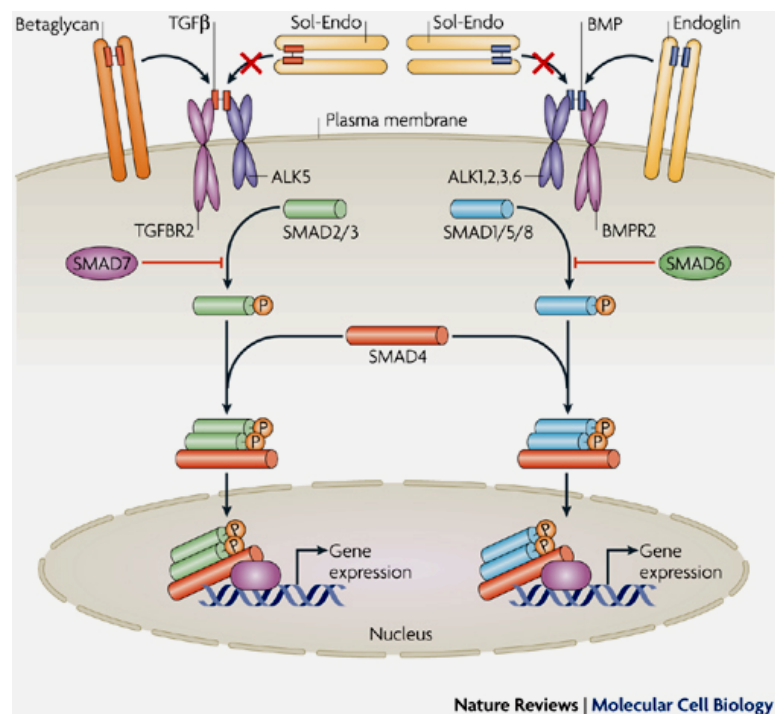
Figure 3. Signal transduction by TGFβ family members *(taken from Nature Reviews on Molecular Cell Biology 2007)*

The type I receptor acts downstream of the type II receptor and propagates the signal to the nucleus by phosphorylating specific members of the SMAD family, receptor-regulated(R)-Smads.[9] The phosphorylated SMADs will then gives signal to the nucleus, and regulates the transcription steps of the genes which play roles in differentiation, growth inhibition, deposition of extra cellular matrix and apoptosis.

TGFβR1 (ALK5) is required for TGFβ-ALK1 activation, whereas ALK1 inhibits intracellular ALK5-SMAD signaling. The differential activation of these two distinct type-I receptor pathways by TGFβ provides the endothelial cells with an intricate mechanism to precisely regulate, and even switch between, TGFβ-induced biological responses. For example, TGFβ-ALK1 activation leads to stimulation of endothelial cell proliferation and migration, whereas TGFβ-ALK5 activation inhibits these responses.[9]

The *TGFBR1* gene is also known as activin A receptor like kinase, or serine/threonine-protein kinase receptor R4 gene. The DNA size is approximately 45kb long, the mRNA size is 2308bp, contains of 9 exons and is located on chromosome 9q22.33.[16,17] The schematic diagram of The *TGFBR1* gene with its exons and introns is presented in the figure below : contains of 9 exons and is located on chromosome 9q22.33.[16,17]
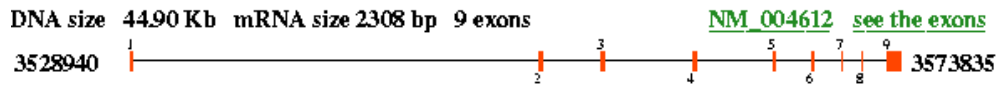
Figure 4. Schematic diagram of *TGFBR1* gene with its exons and introns.

The gene starts from base 3528940 until 3573835, the size is 44.90 Kb, there are 9 exons, with the transcript size 2308 bp. The NCBI code for this gene is NM_004612.

The gene contains 14 different gt-ag introns. Transcription produces 12 different mRNAs, 9 alternatively spliced variants and 3 unspliced forms. There are 4 probable alternative promoters, 2 non overlapping alternative last exons and 10 validated alternative polyadenylation sites.[18]

The protein domains of TGFBR1 consist of : extra cellular domain, transmembrane domain, cytoplasmic domain, glycine-serine rich domain, and serine-threonine kinase domain. These domains are highly conserved across species.[16] The schematic diagram of TGFBR1 domains is described in figure below :

Figure 5. The schematic diagram of TGFBR1 domains, exons and domain organization

Schematic figure of TGFBR1 showing extracellular domain (yellow), transmembrane domain (blue), serine-threonine kinase domain (red), intracellular domain without specific function (grey) and glycine-serine-rich domain (green)[10,16]

Mutations in the genes encoding transforming growth factor-β receptor have been found in patients with MFS and Marfan-like connective tissue disorders. Some syndromes are associated with such mutations including Marfan Syndrome itself,[11,19] Loeys-Dietz Syndrome (LDS) (*TGFBR2* and *TGFBR1*[10,20]) and Sphrintzen-Goldberg Syndrome (*TGFBR2*.[12,21]). Mutations in *TGFBR2* and *TGFBR1* were also found in patients with Familial Thoracic Aneurysms and Dissections.[11]

II.3 ANALYSIS OF DNA SEQUENCE TO DECIDE PATHOGENICITY

Some steps are needed to decide whether the variation in DNA sequence is necessarily pathogenic or not.

The databases of mutation, such as LSDBs (Locus-specific databases), HGMD (Human Gene Mutation Database), UMD (Universal Mutation Database), OMIM (Online Mendelian Inheritance in Man), dbSNP (database of Single Nucleotide Polymorphisms) / Ensembl database can be used for reference. For *TGFBR1* gene, we can look for the previous mutations that have been found, in UMD (Universal Mutation Database : www.umd.be). The dbSNP/Ensembl database (www.ensembl.org) can be used to check whether the point mutation we found is a polymorphism or not.[13,22]

By looking at the type of DNA sequence changes, we can predict their significancy in affecting gene function.

Deletions of the whole gene, nonsense mutation (a form of nonsynonymous substitution where a codon specifying an amino acid is

replaced by a stop codon) and frameshift mutation (a mutation that alters the normal translational reading frame of an mRNA by adding or deleting a number of bases that is not a multiple of three), are almost certain to destroy gene function.[23]

Mutation that change the conserved splice site (GT…AG nucleotides) affects splicing, and will usually abolish the function of the gene. In silico predictions for splice site are available, for example Splice Sequence Finder (Montpelier) www.umd.be/SSF, GeneSplicer Web Interface www.tigr.org/tdb/GeneSplicer/gene_spl.html, etc.[23]

A missense mutation is more likely to be pathogenic if it affects a part of the protein domain known to be functionally important.[23]

Changing of an amino acid is more likely to affect function if that amino acid is conserved in related genes, orthologs (genes present in different genomes which are directly related through descent from a common ancestor) or paralogs (genes present in a single genome as a result of gene duplication). If two or more sequences show sufficient degree of similarity (sequence homology), they can be assumed to be derived from the same ancestor. The higher the degree of similarity, the gene are more conserved, means that the gene has very important role through evolution. The mutation at that point will be strongly suspicious to be pathogenic. Multiple alignment analysis is comparing the amino acid sequence of certain protein with the closest similar sequences from some species. By looking at the position and the presence of

amino acid, we can decide whether the amino acid is conserved across the species or not.[23,24]

Amino acid substitutions are more likely to affect function if they are nonconservative. Nonconservative substitutions result in replacement of one amino acid by another that is chemically not similar. For example, the change from a polar to a non polar amino acid, or an acidic to a basic.[23]

Another way to predict the potential pathogenicity of a mutation is by using in silico prediction analysis. There are some software available in the internet, that can be used to do the prediction, for example PolyPhen (Polymorphisms Phenotyping) and SIFT (Sorting Intolerance From Tolerance). PolyPhen (http://coot.embl.de/PolyPhen/)[25], is an automatic tool for prediction of possible impact of an amino acid substitution on the structure and function of human protein. This prediction is based on empirical rules which are applied to the sequence, phylogenetic and structural information characterizing the substitution. A protein identifier from proteins database, such as SWALL is needed before entering the amino acid substitution. This program will then identify the sites in which the new amino acid replaced, do multiple alignment, and calculate the so-called profile matrix by Position-Specific Independent Counts (PSIC). The PSIC score will be used as one of prediction parameter. A Protein Quarternary Structure (PQS) database is also used as another consideration. The results of PolyPhen can be : probably damaging (it is with high confidence supposed to affect protein function or structure), possibly damaging (supposed to affect protein function or

structure), benign (most likely lacking any phenotypic effect) and unknown (in some rare cases, when the lack of data do not allow PolyPhen to make a prediction).[26] The detail guideline to interpreting PolyPhen result is attached in the attachment.

SIFT BLink is a sequence-homology-based tool that sorts intolerant from tolerant amino acid substitutions and predict whether an amino acid substitution at a particular position in a protein will have a phenotypic effects. SIFT BLink bases its prediction on sequence data alone and does not depend on knowledge of protein structure and function. The results of SIFT BLink prediction are affect protein function and tolerated (means that the substitution can be tolerated, thus does not affect protein function). The sequence data for specific protein is inputted, and will be followed by some steps in which SIFT BLink process the data input to prediction. Substitutions at each position with normalized probabilities less than a chosen cutoff are predicted to be deleterious, while those greater than or equal to the cutoff are predicted to be tolerated.[24]

II.4. THEORETICAL SCHEME

```
                              ┌──────────────┐
                              │  TGFβ genes  │
                              └──────┬───────┘
                                     ↓
┌──────────────┐              ┌──────────────┐              ┌──────────────┐
│ EMILIN1      │─────┐        │ Synthesis and│        ┌─────│ Fibulin genes│
│ gene         │     └───────→│ secretion of │←───────┘     └──────────────┘
└──────────────┘              │    TGFβ      │              ┌──────────────┐
┌──────────────┐     ┌───────→│              │←───────┐     │ Fibrillin    │
│ LTBP genes   │─────┘        └──────┬───────┘        └─────│ genes        │
└──────────────┘                     ↓                      └──────────────┘
┌──────────────┐              ┌──────────────┐              ┌──────────────┐
│ ENG (endoglin)│────┐        │ Altered      │        ┌─────│ BMP genes    │
│ gene         │     └───────→│ activation   │←───────┘     └──────────────┘
└──────────────┘              │ and receptor │              ┌──────────────┐
┌──────────────┐     ┌───────→│ binding      │←───────┐     │ SMAD genes   │
│ TGFBR2       │─────┘        └──────┬───────┘        └─────└──────────────┘
│ gene         │                     ↓                      ┌──────────────┐
└──────────────┘                                      ┌─────│ TGFBR1 gene  │
                                                      └─────│ mutation     │
                                                            └──────────────┘
```

Synthesis and secretion of **TGFβ**

Altered activation and receptor binding

TGFBR1 gene mutation

Altered intracellular signal transduction

Altered intranuclear regulation of genes involved in cell differentiation, growth inhibition, deposisition of extracellular matrix, apoptosis

| Blood vessel | Heart | Skin and integumen | Bone / skeletal system | Eyes | Lungs | Others |

Aneurysm

Hyperextensibility / contracture, etc

Myopia, ectopia lentis

Dural ectasia, etc

Heart valve defect

Tall, imperfect osteogenesis, etc

Spontan pneumothorax

Marfan Syndrome and related disorders phenotype

Notes :
- TGFβ   = Transforming growth factor beta
- LTBP   = Latent transforming growth factor binding protein
- BMP   = Bone morphogenetic protein
- TGFBR1   = Transforming growth factor beta receptor type 1
- TGFBR2   = Transforming growth factor beta receptor type 2

II.5 CONCEPTUAL SCHEME

$$\boxed{\textit{TGFBR1} \text{ gene}}$$

$$\downarrow$$

$$\boxed{\begin{array}{c}\text{Mutations at}\\\text{any specific}\\\text{sites}\end{array}}$$

$$\downarrow$$

$$\boxed{\begin{array}{c}\text{Marfan Syndrome}\\\text{and related}\\\text{disorders}\\\text{phenotypes}\end{array}}$$
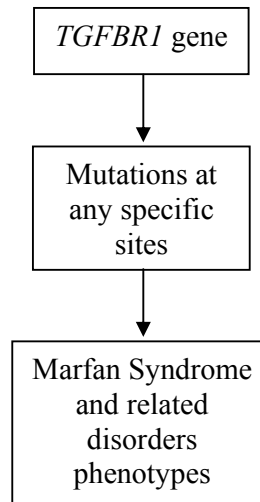
The conseptual scheme of this research

Chapter III

RESEARCH METHODOLOGY

III.1. Research field

This research is in the field of medical genetics.

III.2. Research location

This research was held in the DNA Diagnostic Laboratory of Vrije Universiteit Medisch Centrum (VUmc), Amsterdam, The Netherlands for DNA analysis.

III.3. Research period

This research has been conducted in one year.

III.4. Research design

This is a descriptive study.

III.5. Research methods

III.5.1. Population

The population of this research is the DNA samples of patients with Marfan Syndrome and related disorders which have been reffered to DNA Diagnostic Laboratory of VUmc Hospital Amsterdam, The Netherlands from the year 1998-2008.

III.5.2. Samples

The DNA samples were donation with permission from Gerard Pals, PhD as the principal investigator of Connective Tissue Disorders research in the DNA Diagnostic Laboratory of VUmc Hospital Amsterdam, The Netherlands. All of the samples used in this research are part of Connective Tissue Disorders research project, and have been consent to be included in research (informed consent form attached).

We selected the first 194 unrelated patient's from VUmc's DNA Diagnostic Laboratory database by their registration numbers, which have been referred as Marfan Syndrome, suspected Marfan Syndrome, or related disorders. The phenotypic characteristics of the patients were then traced from their laboratory request form.

III.5.2.1. Inclusion criteria :

1.   Having at least one major criterion of MFS

2.   Found to be negative for *FBN1* and *TGFBR2* mutations.

III.5.2.2. Exclusion criteria :

1.   Not enough amount of DNA available for complete examination.

III.5.2.3. Minimum sample requirement :

This is the first research on *TGFBR1* gene in Marfan Syndrome and related disorders patients in The Netherlands. Another research on *TGFBR1* revealed a frequency of 4% among numbers of patients.[19] Sample amount determination for estimation of proportion in population is as below [28] :

$$n = \frac{Z\alpha^2 \, PQ}{d^2}$$

P=0.04; Zα= 1.96; d=0.10

$$n = \frac{(1.96)^2 \times 0.04 \times (1-0.04)}{(0.10)^2} = 153$$

Notes :
P = the proportion of *TGFBR1* mutations found in previous study = 0.04[19]
d = precision level = 0.10
α = significancy level = 0.95, Z_α = 1.96

The minimum sample which is required is 153 samples.

III.6. Research Variables :

The variables of this research are :

1. Clinical phenotypes of Marfan Syndrome and related disorders

   Scale : nominal

2. Mutation in *TGFBR1* gene

   Scale : nominal

3. Pathogenicity of mutation

   Scale : nominal

III.7. Operational Definitions

1. Marfan Syndrome : a group of clinical signs, fulfilling the Revised Criteria of Marfan Syndrome (Ghent Criteria).

2. Ghent Criteria of Marfan Syndrome : clinical criteria for diagnosing Marfan Syndrome (details attached in the attachment). For the index cases, major criteria in at least 2 different organ systems and involvement in third organ is needed, if the family/genetic history is not contributory. For a relative of an index case, one major criterion in

an organ system and an involvement of second organ is needed if a major criterion in family history is present.

3. Suspected MFS : incomplete Ghent criteria with more than one signs which are mentioned in the criteria.

4. Related disorders of Marfan Syndrome : disorders that share several symptoms with Marfan Syndrome, including Loeys-Dietz Syndrome, Ehler-Danlos Syndrome vascular type, Aortic aneurysms and dissection, Bicuspid Aortic Valve with Aortic Dilatation, Familial Ectopia Lentis, MASS phenotype, Mitral Valve Prolapse Syndrome, Congenital Contractural Arachnodactily (Beals syndrome), Stickler syndrome, Shprintzen-Goldberg Syndrome, joint hypermobility, etc. The clinical features of these disorders are in the attachment.

5. Phenotype : all the clinical signs found in Marfan Syndrome and related disorders patients

6. Mutation : an alteration in DNA sequence.

7. Pathogenicity : the condition in which the mutation will results in protein changes thus causing disease, predicted by the type of amino acid changes, domain localization, multiple alignment, and prediction results of internet-based software, with consideration to literature.

III.8. Mutation Detection

III.8.1 Amplification

In order to get enough amount of DNA fragment to be visible in the gel and have strong enough signal in sequencing, the *TGFBR1*

gen in the DNA need to be amplified. The genomic DNA reference sequence for *TGFBR1* gene amplification is ENSG00000106799.[29] PCR was done for 9 exons of *TGFBR1*, on genomic DNA, with the primers below :

Table 2 : Primers sequence for amplifying the *TGFBR1* gene exon 1-9

No.	Exons	Forward/Reverse	Primer's sequence (5'>3')	Product length (base pairs)
1.	1	Forward	AGTTACAAAGGGCCGGAGCGAGG	302
2.	1	Reverse	TTTGAGAAAGAGCAGGAGCGAGCCA	
3.	2	Forward	TTGGGCTTCCACGTGTATGTG	576
4.	2	Reverse	GTCACTTCTTGCCTCTAAACG	
5.	3	Forward	GCCACCTACAGTGTTTTTGTCGT	530
6.	3	Reverse	TTATACCACCATGGAGCTGACTTAT	
7.	4	Forward	GTATCAGTTTTCTGGGTCACTCA	462
8.	4	Reverse	ATTCGACTTAATGGGTCTAATCTAC	
9.	5	Forward	CAGTGTGTGACTCAGGATTG	340
10	5	Reverse	CCACCTTCTATTTTCATAGACATT	
11.	6	Forward	AATGCCGTAAGTATTGTAGGTCAT	426
12.	6	Reverse	TCTTCTTACCTGTTGGCAATCTA	
13.	7	Forward	TTTTGTGGGATTTAGTTGACATCA	448
14.	7	Reverse	TTTCTCTGGCACTCGGTGA	
15.	8	Forward	AAGGTGTGGGTGGAATATCAACTC	512
16.	8	Reverse	GGCCCTTTCAATGTGCTTACAAT	
17.	9	Forward	TCGGCCTTTTCAGGTTTGCTAA	584
18.	9	Reverse	CCTGGGAAAGAAGCGTTCATAG	
19.	9	Forward2	TTGTAGGCCTTGAGAGTAATGGCTA	383

Table shows the sequence of each primer (forward and reverse) which is used to amplified exon 1 to 9 of *TGFBR1* gene, and the product size.

Notes : For exon 9 we used 2 forward primers, because of a long T-stretch in the DNA sequence. A long T-strech is vulnerable to deletion, so that the sequence output will be messy, and the mutation after the deleted-T will not be detected. The other forward primer which start after T-stretch will prevent the undetected mutation.

At the 5' end of each primer an M13 tail primer sequence forward or reverse (M13 primer, INVITROGEN, Cat.No.N520-02 (F) and N530-02 (R)) was added, in order to simplify the sequencing procedure. With M13 tail primer attached in the PCR primer, the amplified fragment will start from M13 sequence, so that in cycle-sequencing reaction we will need only M13 tail primer to amplify all exon, and not different primer for different exons. The sequences of M13 tails primers are as below :

Table 3 : M13 primers sequence

No.	Primer's name	Sequences
1.	M13 forward	GTAAAACGACGGCCAG
2.	M13 reverse	CAGGAAACAGCTATGA

The table shows the sequences of M13 tail primer, which is attached to PCR primer and used in cycle-sequencing reaction.

Five microliter DNA solution (DNA concentration : 20 ng/µl) was added into 25 µl PCR mixture, which contained 0.2 µl of 25 mM dNTPs, 0.75 µl of 50 mM MgCl$_2$ (Invitrogen), 1µl of 10pmol/µl each primer (Invitrogen), 2.5 µl of 10x PCR buffer (Invitrogen), 0.2 µl of 5U/µl Taq DNA polymerase (Platinum Taq DNA Polymerase, Invitrogen,

Cat.No.10966-034) and 15.35 µl H2O. The thermal profile included initial denaturation for 5 minutes at 94^0C, followed by 35 cycles of denaturation (1 minute at 94^0C), annealing (1 minute at 65^0C), and extension (1 minute at 72^0C), in PE9700 Applied Biosystem thermocycler. Five microliters of each sample was then runned on an 2% agarose gel with 100V for 30 minutes and stained with ethidium bromide, to confirm PCR amplification product (the size of PCR product as described in table 1).

III.8.2 DNA sequencing

The purpose of sequencing is to determine the order of the nucleotides of a gene.

Prior to sequencing, the PCR products were purified from excess primers and dNTPs molecule by a mixture of Exo 1 (Exonuclease 1, USB Corp. Cleveland, Ohio, Cat.No.70073X) and SAP enzyme (Shrimp Alkaline Phosphatase, USB Corp. Cleveland, Ohio, Cat.No.70092Y). Five microliters of PCR were taken into the reaction, together with 0.25 µL SAP, 0.25 µL Exo1 and 1.5 µL HPLC H2O. The mixture was then incubated in a thermocycler with a program of 30 minutes in 37^0C followed by 15 minutes in 80^0C. Then diluted with 15 µL HPLC H2O, and was added sequencing primer (forward or reverse) as much as 1 µL.

Sequencing reactions used the BigDye Terminator Cycle Sequencing kit (version 3 Applied Biosystems. Foster City, CA, USA, Cat.No.4737458) on an ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, Ca,

USA). Seven microliter of BigDye mix, which contained of 0.5 μL BigDye V3.1 reaction mix, 1.75 μL BigDye V3.1 5x sequencing buffer and 4.75 μL HPLC H2O, were taken into reaction together with 3 μL of SAP-Exo1-PCR product mixture. Then runned in a thermocycler with a program of 96^0C 10 minutes denaturation, 55^0C 5 seconds annealing, 60^0C 4 minutes elongation, 25 cycles.

The products of sequencing reaction were then precipitated using ethanol precipitation method in order to remove unincorporated dye terminators. The product would then be added with 20 μL formamide, heated in thermocycler on 94^0C for 2 minutes and cooled down to 4^0C, and put in the sequencer (ABI 3730 Genetic Analyzer, Applied Biosystem)

III.9. Mutation Analysis

We compared the sequence of patients with the reference sequence. The variant numbering is based on the cDNA sequence (ENST00000374994),[30] where +1 corresponds to the nucleotide A of ATG, the translation initial codon.

The UMD database of *TGFBR1* mutations, the Ensemble SNPs database of polymorphisms and the previous reports on *TGFBR1*, were used to confirm the DNA sequence variants. Whenever the variant was not mentioned as polymorphism in one of those references, we did the analysis based on the changes in amino acid types, domain conservation in some

species, protein structure and previous publications on the mutations. Internet-based software programs to predict the possible impact of amino acid substitutions were also used to help the analysis.

The first program was PolyPhen (http://coot.embl.de/PolyPhen/)[25], a web-based tool to predict the possible impact of amino acid substitution on the structure and function of the protein. The data query needs a protein identifier which codes specific protein in the protein database. The protein identifier in SWALL-protein database for TGFBR1 is P36897.[31] We use default query parameters for protein quaternary structure (PQS) databases and performing calculations for all hits. The second program we used was SIFT Blink[27], a sequence homology-based amino acid substitution prediction method (available at http://blocks.fhcrc.org/sift/SIFT_BLink_submit.html). We applied gi:4759226 protein sequence of TGFBR1 by using parameter "best BLAST hit to each organism" and omitting sequences 100% identical to query. Results were reported as "affects protein function" or "tolerated" according to this analysis.

To help predict the affect of mutation on the splice site, we used web-based tool Human Splicing Finder[32] (available at www.umd.be/HSF/) which analyzed the sequence towards the presence of enhancer motifs, silencer motifs, exonic splicing regulatory sequences, potential branch points and potential splice sites.

III.10. Research Flow

```
┌──────────────────┐  ┌──────────┐  ┌──────────┐  ┌──────────────┐
│ Ehler-Danlos     │  │  Aortic  │  │  Dural   │  │  Loeys-Dietz │
│ Syndr            │  │ aneurysms│  │  ectasia │  │   Syndrome   │
│ vascular type    │  │          │  │          │  │              │
└──────────────────┘  └──────────┘  └──────────┘  └──────────────┘
        ┌──────────┐      ┌──────────┐     ┌──────────────┐
        │ Ectopia  │      │ Classic  │     │  Suspected   │
        │ lentis   │      │   MFS    │     │     MFS      │
        └──────────┘      └──────────┘     └──────────────┘
           ┌──────────────┐                   ┌──────────────┐
           │ Sphrintzen-  │                   │    Other     │
           │  Goldberg    │                   │   related    │
           │  Syndrome    │                   │  disorders   │
           └──────────────┘                   └──────────────┘

                      ┌──────────────┐
                      │   Patients   │
                      │   included   │
                      └──────────────┘
                            │
                      ┌──────────────┐
                      │ Amplification│
                      │ of TGFBR1    │
                      │ gene on gDNA │
                      │ by PCR       │
                      └──────────────┘
                            │
                      ┌──────────────┐
                      │ Confirmation │
                      │ of PCR       │
                      │ product on   │
                      │ agarose gel  │
                      └──────────────┘
                            │
                      ┌──────────────┐
                      │Pre-sequencing│
                      │ preparation  │
                      └──────────────┘
                            │
                      ┌──────────────┐
                      │  Sequencing  │
                      └──────────────┘
                            ┆
```
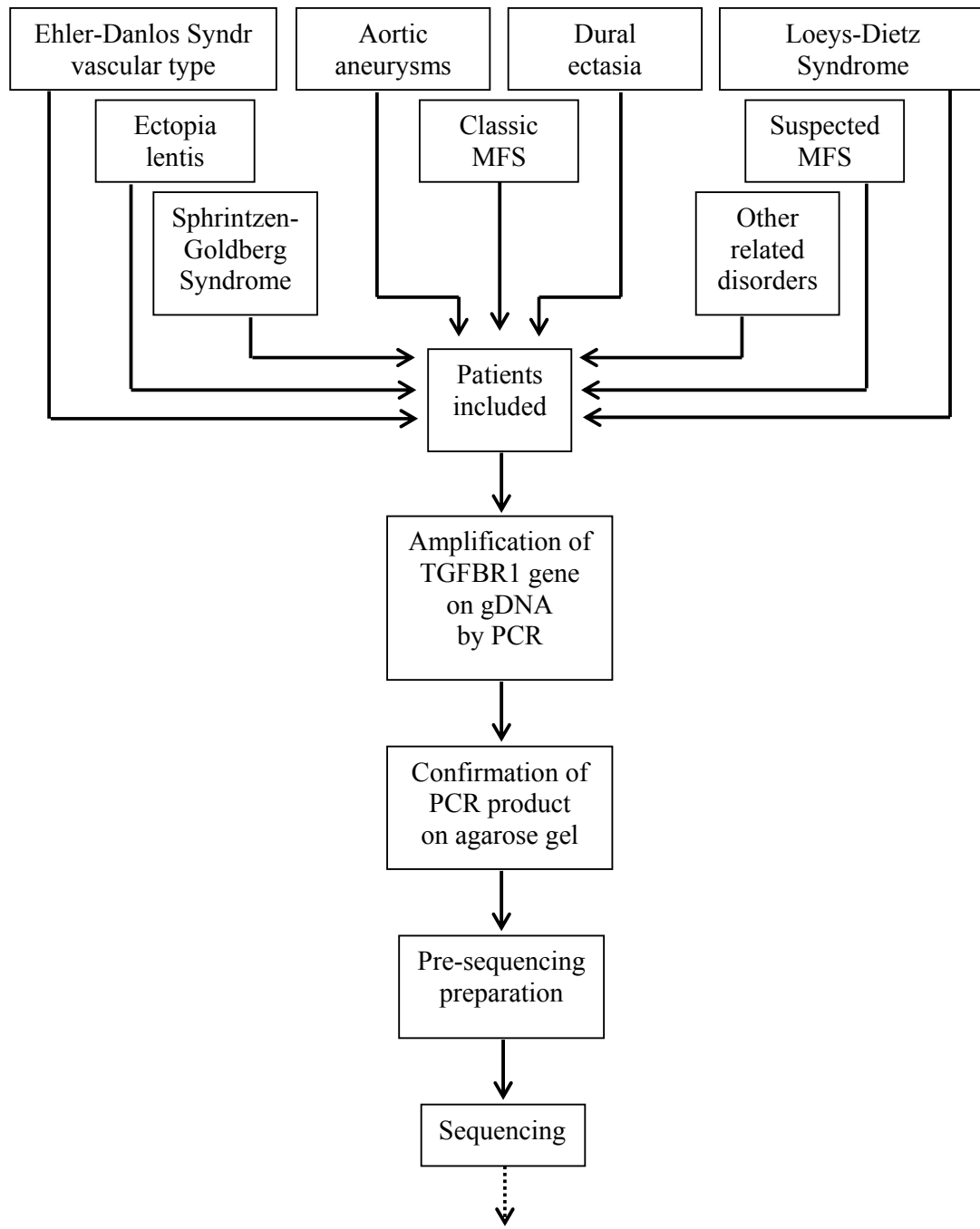
to be continued in next chart

III.10. Research Flow (continued)
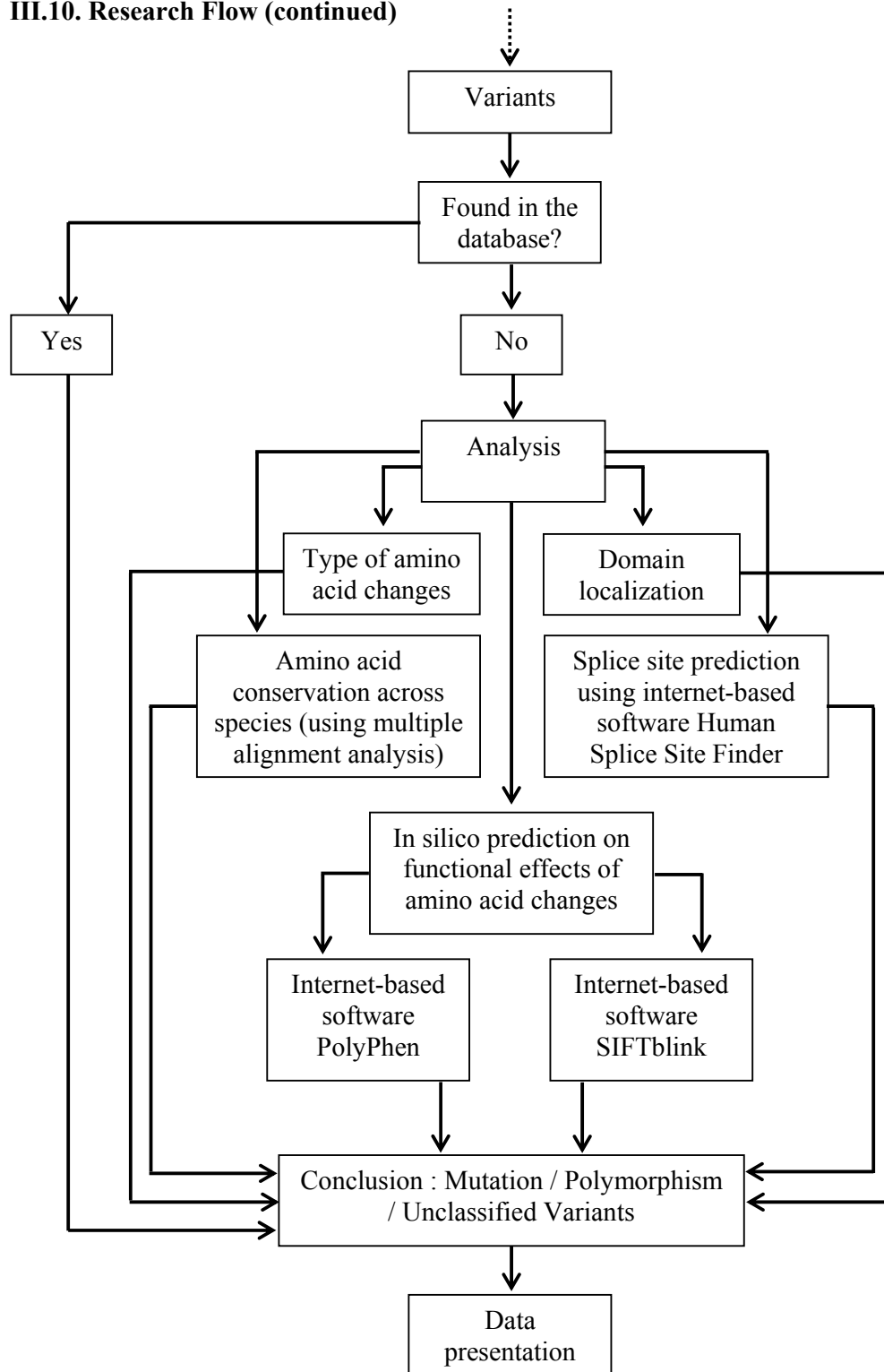
III.11. Data Analysis

The data will be analyzed descriptively for the clinical features of the patients, the number of patients in each diagnosis group, the mutations that have been found and the distribution in each exon and domain, the amino acid type changes and the prediction of pathogenicity with their multiple sequence alignment, and the polymorphisms and unclassified variants. The details are as below :

1. The clinical features of the patients; the list of mutations, amino acid-type changes and the prediction results from PolyPhen and SIFT; the list of polymorphisms and unclassified variants and the distribution of *TGFBR1* mutations on clinical diagnosis will be presented in tables.

2. The number of patients in each diagnosis will be presented in graph.

3. The distribution of mutations in each exon and domain will be presented in schematic figure.

4. The multiple sequence alignment will be presented in figure.

Chapter IV

RESULTS

IV.1 Clinical diagnosis of patients

The patient samples included in this study came from many centers, inside and outside The Netherlands, such as Belgium and United Kingdom. All the DNA samples included are a donation with permission from DNA diagnostic laboratory of Vrije Universiteit Amsterdam, The Netherlands.

The clinical information of patients described here has been collected from clinical observations that have been mentioned in the laboratory request form.

Tabel 4. Detail Clinical Features of Marfan Syndrome and Related Disorders Patients based on Organ System presented in Percentage

No.	Clinical Features	Number of Patients	Percentage from total 194 patients
	Skeletal		
1	Marfanoid habitus	18	9.28%
2	Joint hypermobility	16	8.25%
3	Pectus excavatum/carinatum	11	5.67%
4	Increased span-height ratio	11	5.67%
5	Tall and thin	9	4.64%
6	Scoliosis	9	4.64%
7	Arachnodactily	7	3.61%
8	High and narrow palate	5	2.58%
9	Positive fingers signs (thumb sign & wrist sign)	3	1.54%
10	Kyfosis	2	1.03%
11	Flat foot	2	1.03%
12	Shoulder luxation	2	1.03%
13	Spondilolisthesis	1	0.52%
14	Contracture of the hands	1	0.52%
15	Palatoschizis	1	0.52%

No.	Clinical Features	Number of Patients	Percentage from total 194 patients
16	Crowded teeth	1	0.52%
17	Skeletal abnormalities (unspecified)	26	13.40%
	Cardiovascular		
1	Aortic aneurysms	137	70.62%
2	Aortic dissection	24	12.37%
3	Mitral Valve Prolaps	5	2.58%
4	Aortic valve insufficiency	4	2.06%
5	Pulmonary stenosis	2	1.03%
6	Dissections of artery coronaria	2	1.03%
7	Aneurysms of other big vessel	1	0.52%
8	Persisten Ductus Arteriosus	1	0.52%
9	Mitral Insufficiency	1	0.52%
10	Varices	1	0.52%
11	Heart problem (unspecified)	1	0.52%
	Occular		
1	Myopia	4	2.06%
2	Lens subluxation	4	2.06%
3	Ectopia Lentis	3	1.54%
4	Flat cornea	1	0.52%
5	Retinal detachment	1	0.52%
6	Eye abnormality (unspecified)	7	3.61%
	Lung		
1	Spontaneous pneumothorax	3	1.54%
2	Lung abnormality (unspecified)	2	1.03%
	Dura		
1	Dural ectasia	6	3.09%
	Skin & Integumen		
1	Striae	5	2.58%
2	Thin skin	2	1.03%
3	Uterus & Bladder prolaps	2	1.03%
4	Hernia inguinalis	1	0.52%
5	Skin abnormality unspecified	4	2.06%
	Others		
1	Uvula bifida	2	1.03%
2	Mental retardation	1	0.52%

Notes : one patient may have more than one clinical features.

The clinical diagnoses of the patients were based on clinical findings

and matched with Ghent Criteria. A diagnosis of MFS was based on Ghent

Criteria. Incomplete Ghent Criteria, or having at least one major criterion in an organ system with minor criterion of another organ, or more than one minor criterion, would be considered as Suspected MFS. The patients with only specific clinical features (such as only has aortic aneurysm, ectopia lentis, dural ectasia or joint hypermobility) would be grouped as the clinical findings, recognized as Marfan Syndrome, Suspected MFS, Aortic Aneurysms and/ Dissections, Familial Aortic Aneurysms and/ Dissections, Ectopia Lentis, Dural Ectasia, Joint Hypermobility.

The summary of patients based on clinical diagnosis are presented in the graph below :
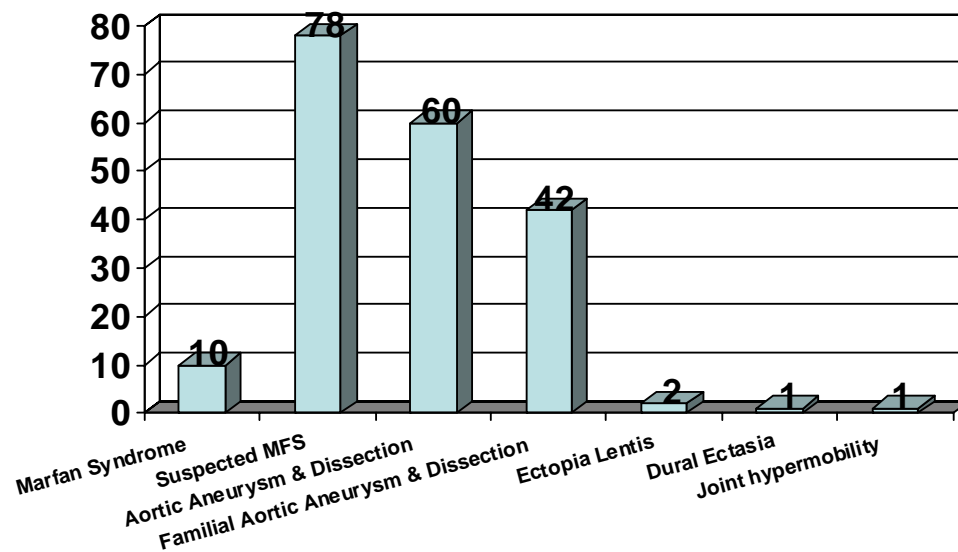
Figure 6. Bar graph showing the number of patients in each group

Most of the patients included in this research were diagnosed as suspected MFS (78 patients), followed by aortic aneurysms and dissection (60 patients) and familial cases of aortic aneurysms and dissections (42 patients).

IV.2 *TGFBR1* mutation detection results

On sequencing all 9 exons of *TGFBR1*, a total of 9 mutations, 7 different polymorphisms and 3 unclassified variants in *TGFBR1* were found. The mutations were found in 10 patients. The 9 mutations, occured in 7 different exons (see table 5).

We did analysis on mutations by observing the amino acid changes, looking at the conservation in 11 different species and the domain localization, and using internet-based software to predict the pathogenicity of amino acid changes.

The list of mutations, amino acid-type changes and the prediction results from PolyPhen and SIFT are presented in table 3 below :

Table 5. Mutations, amino acid type changes and Predicted Functional Effects of amino acid changes

No	Location	Mutation	Mutation Type	AA changes	PolyPhen	SIFT	Diagnosa
1	Exon 2	c.113G>A; p.C38Y	Missense mutation	Po N > Po N	Probably damaging	Tolerated	Familial aortic aneurysm &/ dissection
2	Exon 3	c.451C>T; p.R151C	Missense mutation	Po B > Po N	Benign	Tolerated	Suspected Marfan Syndrome
3	Exon 4	c.605C>T; p.A202V	Missense mutation	NPo N > NPo N	Benign	Affects protein function	1. Suspected Marfan Syndrome 2. Familial aortic aneurysm &/ dissection
4	Exon 5	c.839C>T; p.S280L	Missense mutation	Po N > NPo N	Possibly damaging	Tolerated	Suspected Marfan Syndrome
5	Exon 5	c.958A>G; p.I320V	Missense mutation	NPo N > NPo N	Possibly damaging	Tolerated	Suspected Marfan Syndrome
6	Exon 5	c.965G>A; p.G322D	Missense mutation	NPo N > Po A	Benign	Tolerated	Aortic aneurysm &/ dissection
7	Exon 6	c.980C>T; p.P327L	Missense mutation	NPo N > NPo N	Probably damaging	Affects protein function	Suspected Marfan Syndrome
8	Exon 8	c.1282T>G; p.Y428D	Missense mutation	Po N > Po A	Probably damaging	Affects protein function	Familial aortic aneurysm &/ dissection
9	Exon 9	c.1460G>A; p.R487Q	Missense mutation	Po B > Po N	Probably damaging	Tolerated	Suspected Marfan Syndrome

Notes :
AA = Amino Acid
PolyPhen = Polymorphisms Phenotyping
SIFT = Sorting Intolerance from Tolerance
PolyPhen and SIFT are prediction tools for predicting the functional effects of amino acid substitution

Explanation of the table and sequencing results :

All of the mutations are missense mutations, in which a nucleotide substitution results in an amino acid change :

1. The mutation is located in exon 2 of *TGFBR1* gene, at the position 113 of cDNA, in which guanine is replaced by adenine, resulted in the change of amino acid 38 from cysteine (a polar-neutral amino acid) to tyrosine (a polar-neutral). This mutation is predicted to be probably damaging by PolyPhen and tolerated by SIFT.

    The position of mutation in gene sequence is shown below :
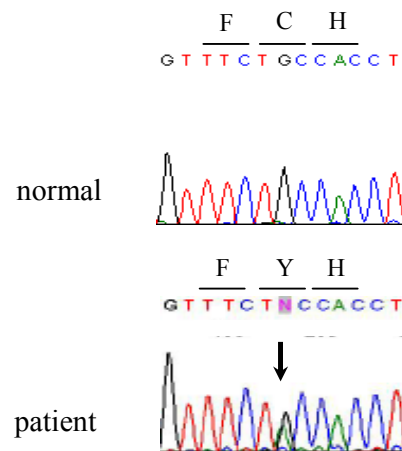
Figure 7. Mutation c.113G>A; p.C38Y in *TGFBR1* (forward sequence)
Mutation in exon 2, showed a Cysteine (TGC) change to Tyrosine (TAC).

2. The mutation is located in exon 3 of *TGFBR1* gene, at the position 451 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 151 from arginine (a polar-basic amino acid) to cysteine (a polar-neutral amino acid). This mutation is predicted to be benign by PolyPhen and tolerated by SIFT.

The position of mutation in gene sequence is shown below :

Figure 8. Mutation c.451C>T; p.R151C in *TGFBR1* (forward sequence)
Mutation in exon 3, showed an Arginine (CGC) change to Cysteine (TGC).

3. The mutation is located in exon 4 of *TGFBR1* gene, at the position 605 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 202 from alanine (a nonpolar-neutral amino acid) to valine (a nonpolar-neutral amino acid). This mutation is predicted to be benign by PolyPhen and affects protein function by SIFT.

The position of mutation in gene sequence is shown below :

Figure 9. Mutation c.605C>T; p.A202V in *TGFBR1* (forward sequence)
Mutation in exon 4, showed an Alanine (GCG) change to Valine (GTG).

4. The mutation is located in exon 5 of *TGFBR1* gene, at the position 839 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 280 from serine (a polar-neutral amino acid) to leucine (a nonpolar-neutral amino acid). This mutation is predicted to be possibly damaging by PolyPhen and tolerated SIFT.

The position of mutation in gene sequence is shown below :
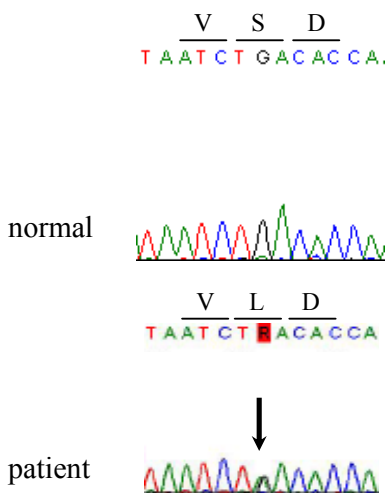
Figure 10. Mutation c.839C>T; p.S280L in TGFBR1 (reverse sequence)
Mutation in exon 5, showed an Serine (TCA) change to Leucine (TGA), sequence shown in reverse.

5. The mutation is located in exon 5 of *TGFBR1* gene, at the position 958 of cDNA, in which adenine is replaced by guanine, resulted in the change of amino acid 320 from isoleucine (a nonpolar-neutral amino acid) to valine (a nonpolar-neutral amino acid). This mutation is predicted to be possibly damaging by PolyPhen and tolerated SIFT.

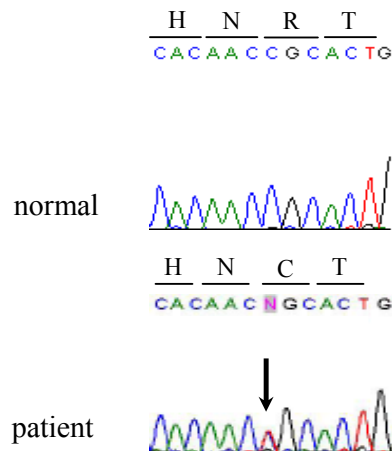The position of mutation in gene sequence is shown below :
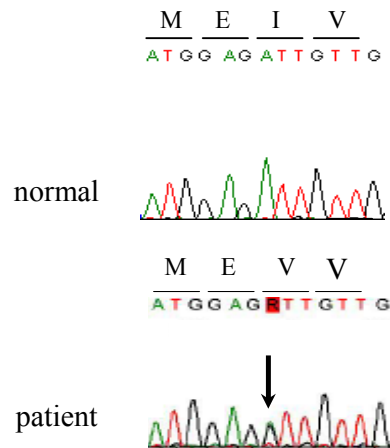
Figure 11. Mutation c.958A>G; p.I320V in *TGFBR1* (forward sequence)
Mutation in exon 5, showed an Isoleucine (ATT) change to Valine (GTT).

6. The mutation is located in exon 5 of *TGFBR1* gene, at the position 965 of

    cDNA, in which guanine is replaced by adenine, resulted in the change of

    amino acid 322 from glycine (a nonpolar-neutral amino acid) to aspartic

    acid (a polar-acidic amino acid). This mutation is predicted to be benign

    by PolyPhen and tolerated by SIFT.
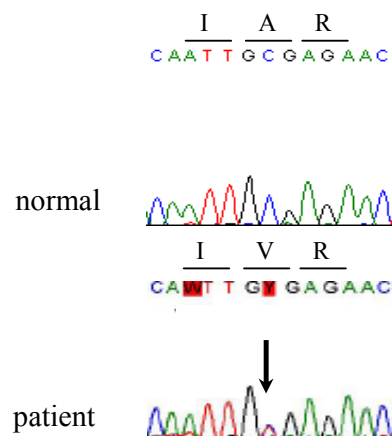
The position of mutation in gene sequence is shown below :
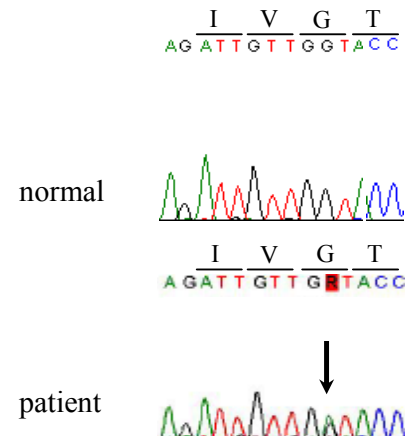
Figure 12. Mutation c.965G>A; p.G322D in *TGFBR1* (forward sequence)
Mutation in exon 5, showed a Glycine (GGT) change to Aspartic acid (GAT).

7. The mutation is located in exon 6 of *TGFBR1* gene, at the position 980 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 327 from proline (a nonpolar-neutral amino acid) to leucine (a nonpolar-neutral amino acid). This mutation is predicted to be probably damaging by PolyPhen and affects protein function by SIFT.

The position of mutation in gene sequence is shown below :
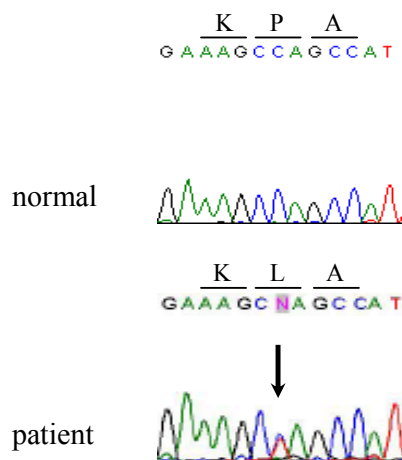
Figure 13. Mutation c.980C>T; p.P327L in *TGFBR1* (forward sequence)
Mutation in exon 6, showed a Proline (CCA) change to Leucine (CTA).

8. The mutation is located in exon 8 of *TGFBR1* gene, at the position 1282 of cDNA, in which timidine is replaced by guanine, resulted in the change of amino acid 428 from tyrosine (a polar-neutral amino acid) to aspartic acid (a polar-acidic amino acid). This mutation is predicted to be probably damaging by PolyPhen and tolerated by SIFT.

The position of mutation in gene sequence is shown below :

L   P   Y   Y
C T G C C T T AT T AT G

normal

L   P   D   Y
C T G C C T N AT T AT G

patient
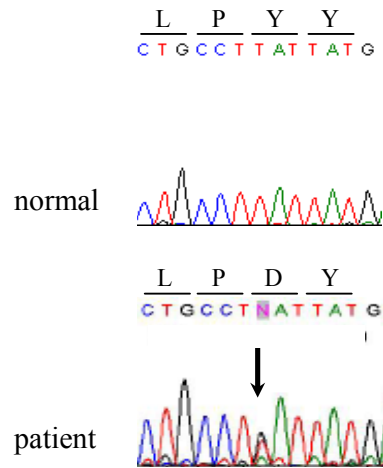
Figure 14. Mutation c.1282T>G; p.Y428D in *TGFBR1* (forward sequence)
Mutation in exon 8, showed a Tyrosine (TAT) change to Aspartic acid (GAT).

9. The mutation is located in exon 9 of *TGFBR1* gene, at the position 1460 of cDNA, in which guanine is replaced by adenine, resulted in the change of amino acid 487 from arginine (a polar-basic amino acid) to glutamine (a polar-neutral amino acid). This mutation is predicted to be probably damaging by PolyPhen and tolerated by SIFT.

The position of mutation in gene sequence is shown below :
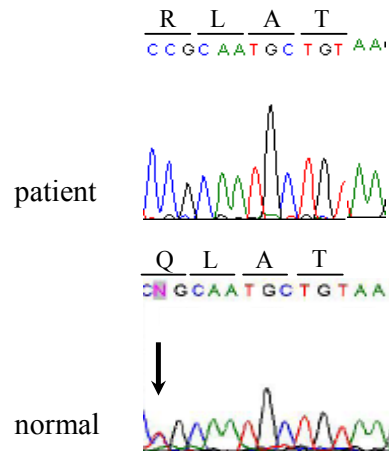
Figure 15. Mutation c.1460G>A; p.R487Q in TGFBR1 (reverse sequence)
Mutation in exon 9, showed an Arginine (CGG) change to Glutamine (CAG).

Seven out of nine mutations occured at a well-conserved amino acid of the kinase domain. Mutations in exon 2 and exon 3 occured in the extracellular domain and cytoplasmic, intracellular domain, respectively. The distribution of mutations in each exon and domain are shown in figure 10 below :
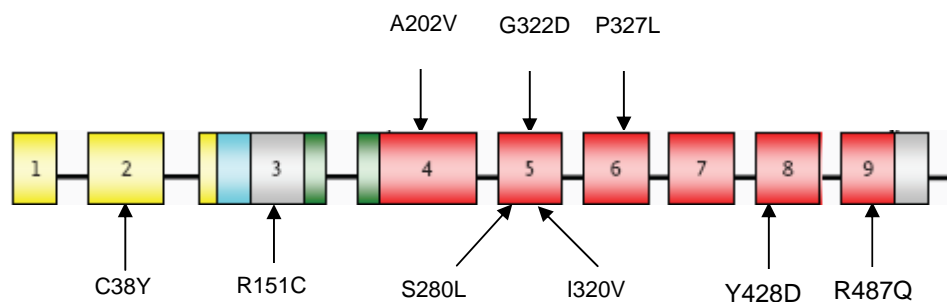
Figure 16. Exons, domain organization and location of the mutations
Note : extracellular domain (yellow), transmembrane domain (blue), serine-threonine kinase domain (red), intracellular domain without specific function (grey) and glycine-serine-rich domain (green)[7,17]

From the picture above we can see that most of the mutations are located in exon 5 (3 out of 9 different mutations).

We did the multiple alignment to see the conservation of TGFBR1 across the species. The homologs for TGFBR1 are TGFBR1 in *P.troglodytes*, TGFBR1 in *C.familiaris*, Tgfbr1 in *B.taurus*, tgfbr1 in *M.musculus*, TGFBR1 in *R.norvegicus*, tgfbr1 in *D.rerio*, babo in *D.melanogaster*, AgaP_AGAP008247 in *A.gambia*, daf-1 in *C.elegans*. The multiple alignments are shown in Table 6 :

Table 6. Multiple Sequence Alignment

| Protein	Species	Alignment		
		C38	R151	A202
TGFBR1	*H. sapiens*	LQCF**C**HLCT	ICHN**R**TVI	RTIA**R**TIV
TGFBR1	*P. troglodytes*	LQCF**C**HLCT	ICHN**R**TVI	RTIA**R**TIV
TGFBR1	*C. familiaris*	LQCF**C**HLCT	ICHN**R**TVI	RTIA**R**TIV
TGFBR1	*B. taurus*	LQCF**C**HLCT	ICHN**R**TVI	RTIA**R**TIV
Tgfbr1	*M. musculus*	LQCF**C**HLCT	ICHN**R**TVI	RTIA**R**TIV
Tgfbr1	*R. norvegicus*	LQCF**C**HLCT	ICHN**R**TVI	RTIA**R**TIV
TGFBR1	*G. gallus*	LQCF**C**HLCT	LCHN**R**TVI	RTIA**R**TIV
tgfbr1	*D. rerio*	LLCY**C**ERCV	MCHN**R**SII	RTIA**R**TIV
babo	*D. melanogaster*	IKCH**C**DTCK	YCQR**R**ARM	RSIA**R**QVQ
AgaP_AGAP008247	*A. gambiae*	LKCH**C**DICK	***R**R**KRNS	RSIA**R**QIQ
daf-1	*C. elegans*	EFLN**E**TDRS	DWYI**R**FKP	LTIG**G**QIR
Protein	Species	Alignment		
		S280	I320	G322
TGFBR1	*H. sapiens*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
TGFBR1	*P. troglodytes*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
TGFBR1	*C. familiaris*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
TGFBR1	*B. taurus*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
Tgfbr1	*M. musculus*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
Tgfbr1	*R. norvegicus*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
TGFBR1	*G. gallus*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
tgfbr1	*D. rerio*	LWLV**S**DYHE	LHME**I**VGTQ	LHMEIV**G**TQ
babo	*D. melanogaster*	LWLV**T**DYHE	LHMD**I**VGTR	LHMDIV**G**TR
AgaP_AGAP008247	*A. gambiae*	LWLV**T**DYHE	LHMD**I**VGTR	LHMDIV**G**TR
daf-1	*C. elegans*	LWLV**T**EYHP	LHNQ**I**GGSK	LHNQIG**G**SK

Protein	Species	Alignment		
		P327	R487	Y428
TGFBR1	*H. sapiens*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDLVPSDP
TGFBR1	*P. troglodytes*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDLVPSDP
TGFBR1	*C. familiaris*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDLVPSDP
TGFBR1	*B. taurus*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDLVPSDP
Tgfbr1	*M. musculus*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDLVPSDP
Tgfbr1	*R. norvegicus*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDLVPSDP
TGFBR1	*G. gallus*	GK**P**AIAHRD	LTAL**R**VKKS	**Y**YDLVPSDP
tgfbr1	*D. rerio*	GK**P**AIAHRD	LTAL**R**IKKT	**Y**YDVVQPDP
babo	*D. melanogaster*	GK**P**AIAHRD	LSSL**R**IKKT	**F**YDVVQPDP
AgaP_AGAP008247	*A. gambiae*	GK**P**AIAHRD	LSSL**R**IKKT	**F**YDVVQPDP
daf-1	*C. elegans*	NK**P**AMAHRD	FTSY**I**CRKR	**Y**IEWTDRDP

multiple sequence alignment taken from

[*http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=homologene&dopt=MultipleAlignment&list_uids=3177*](http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=homologene&dopt=MultipleAlignment&list_uids=3177) *)33* Notes : letters in red indicated the mutated amino acids.

From this multiple alignment, it is shown that the mutated amino acid is highly conserved, and that *TGFBR1* mutations occurred at evolutionary-conserved domains.

Furthermore, 7 different polymorphisms in 6 different exons were found. and listed as below :

Table 7. Polymorphisms found in this study

No.	Exon / Intron	Polymorphism	Found in
1	1	c.70_78delGCGGCGGCG	63 patients
2	4	c.805+39A>G	2 patients
3	6	c.1125A>C p.T375T	2 patients
4	7	c.1255+24G>A	72 patients
5	7	c.1237C>A p.R413R	1 patient
6	8	c.1386+90_94delTCTTT	64 patients
7	9	g.45245A>G	66 patients

Explanation of the table, starts from the first polymorphism :

1. Polymorphism 1 is located in exon 1, it is an in-frame deletion of 9 bases starts from position 70 of cDNA until 78 (GCGGCGGCG), and causes deletion of 3 amino acid Alanin.

2. Polymorphism 2 is located in intron 4, in the position of cDNA 805+39, where adenine is replaced by guanine.

3. Polymorphism 3 is located in exon 6, in the position of cDNA 1125, where adenine is replaced by cytosine. It is a silent mutation, because the nucleotide change results in the same amino acid, Threonine.

4. Polymorphism 4 is located in intron 7, in the position of cDNA 1255+24 where guanine is replaced by adenine.

5. Polymorphism 5 is located in exon 7, in the position of cDNA 1237, where cytosine is replaced by adenine. It is a silent mutation, because the nucleotide change results in the same amino acid, Arginine.

6. Polymorphism 6 is located in intron 8. It is a deletion of 5 bases (TCTTT) in the position of cDNA 1386+90 to 1386+94.

7. Polymorphism 7 is located in intron 9, in the position of gDNA 45245, where adenine is replaced by guanine.

These 7 polymorphisms have been previously reported in the Ensembl database of polymorphisms.[29]

Three variants are left as unclassified (see table 5). They neither have been reported as polymorphisms, nor as mutations.

Table 8. Unclassified Variants (UV)

No.	Exon / Intron	Unclassified Variants	Found in	Clinical Features
1	2	c.343+46T>G	1 patient	Skeletal abnormality and aortic aneurysm
2	7	c.1255+103G>A	4 patients	1. Joint hypermobility, aortic aneurysm, pneumothorax 2. Familial aortic aneurysm 3. Scoliosis, pectus excavatum, arachnodactily, narrow & high palate 4. Pectus carinatum, flexible shoulder, tall & skinny
3	8	c.1386+87_91delTTTTC	1 patient	Joint hypermobility, aortic aneurysm

Explanation of the table, starts from the first UV :

1. UV 1 is located in intron 2, in the position of cDNA 343+46, where timidine is replaced by guanine. This UV presents in patient with skeletal abnormality and aortic aneurysm.

2. UV 2 is located in intron 7, in the position of cDNA 1255+103, where guanine is replaced by adenine. This UV presents in 4 patients with :

   1. Joint hypermobility, aortic aneurysm and pneumothorax.

   2. Familial aortic aneurysm.

   3. Scoliosis, pectus excavatum, arachnodactily, narrow and high palate.

   4. Pectus carinatum, flexible shoulder, tall and skinny.

3. UV 3 is located in intron 8. It is a deletion of 5 bases (TTTTC), starts from c.1386+87 to 1386+91. This UV presents in patient with joint hypermobility and aortic aneurysm.

All of the Uvs are non-coding variants (located in intron, which are not code the amino acid). To decide the pathogenicity, they need to be analyzed on cDNA to see whether this UV affecting splice site, therefore tissue biopsies of these patients are needed to perform the analyses. The DNA of parents are unfortunately unavailable.

IV.3 Distribution of mutations on clinical diagnosis

Table 9. *TGFBR1* mutations on clinical diagnosis

NO.	CLINICAL DIAGNOSIS	NUMBER OF PATIENTS	*TGFBR1* MUTATIONS			
			Total	Detail		
				Exon	Mutation	Pathogenicity status
1.	Marfan Syndrome	10	0	-	-	-
2.	Suspected Marfan Syndrome	78	6	1. Exon 3 2. Exon 4 3. Exon 5 4. Exon 5 5. Exon 6 6. Exon 9	c.451C>T; p.R151C c.605C>T; p.A202V c.839C>T; p.S280L c.958A>G; p.I320V c.980C>T; p.P327L c.1460G>A; p.R487Q	Pathogenic Unlikely to be pathogenic Pathogenic Unlikely to be pathogenic Pathogenic Pathogenic
3.	Aortic aneurysm and or dissection	60	1	1. Exon 5	c.965G>A; p.G322D	Pathogenic
4.	Familial aortic aneurysm and or dissection	42	3	1. Exon 2 2. Exon 4 3. Exon 8	c.113G>A; p.C38Y c.605C>T; p.A202V c.1282T>G; p.Y428D	Pathogenic Unlikely to be pathogenic Pathogenic
5.	Ectopia lentis	2	0	-	-	-
6.	Dural ectasia	1	0	-	-	-
7.	Joint hypermobility	1	0	-	-	-

Most of the mutations occured in patients with suspected Marfan Syndrome, followed by familial cases of aortic aneurysms. None of the patient with classic MFS, ectopia lentis, dural ectasia or joint hypermobility has *TGFBR1* mutation.

IV.4 Clinical characteristics of patients carrying the mutations

The clinical information has been collected from clinical phenotypes that have been mentioned in laboratory request.

The first patient (II.4), who has the mutation c.113G>A, p.C38Y is a male, having a type A thoracic aorta dissection at age 46. No other features related to MFS, LDS, EDS Vascular type or other syndrome had been found. One of his brothers has a history of aortic dissection.

Pedigree :
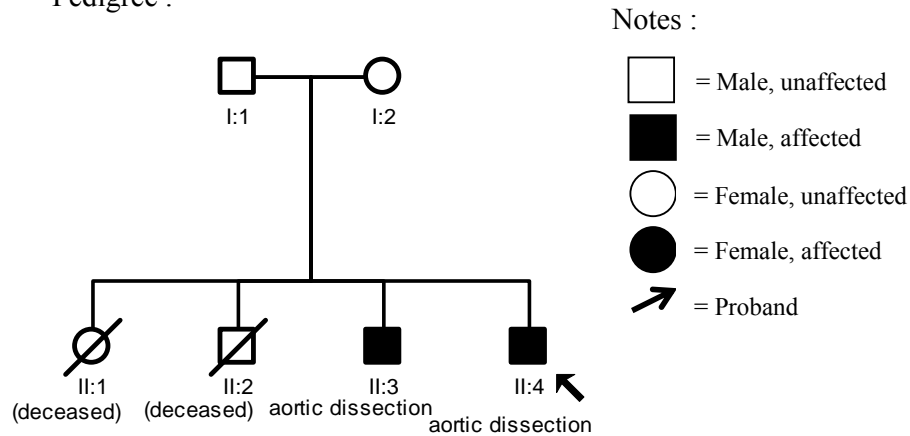
Figure 17. Pedigree of family 1
Familial case of Thoracic aortic aneurysm, in which two members of the family have the same clinical feature

Patient 2 (c.451C>T, p.R151C), male, 50 years old, has an thoracic aortic aneurysms and minor signs of MFS. His mother has valvular heart disease and his father died suddenly at the age of 62 without any known cause.

Patient 3 (I.2) who has mutation c.605C>T, p.A202V, is a female, 57 years old with thoracic aortic dissection. No other feature related to MFS, LDS, EDS vascular type or other syndrome has been found. She had a son who died earlier because of thoracic aortic dissection at age 23. Her daughter is healthy. The same mutation did not appear in her daughter's DNA. The presence or absence of this mutation in her affected son will provide more information with regard to pathogenicity. Unfortunately, the DNA of her son is not available.

Pedigree of patient 3 :

Figure 18. Pedigree of patient 3
An autosomal dominant pattern of inheritance in which proband has child with the same features

Patient 4 (III.1), female, 31 years old, has also the mutation c.605C>T, p.A202V. She was diagnosed as suspected MFS, unfortunately her clinical detail is not available. Her mother, maternal uncle and maternal grandmother have MFS. Unfortunately, they have passed away and there is no DNA available to perform further analysis.

Figure 19. Pedigree of patient 4
An autosomal dominant pattern of inheritance in which proband and her previous generation have the same features

Patient 5 (c.839C>T, p.S280L), male, 24 years old, has skeletal features of MFS. He is tall with thin and long extremities, contractures of the hands, recurrent shoulder luxations and arachnodactyly. No other features of MFS in other organ system have been found. No other member of his family is found to have the same features.

Patient 6 (c.958A>G p.I320V) is a male 53 years old who was diagnosed as suspected MFS, unfortunately the clinical detail is not available. There is no other family member known to have the same features. Unfortunately, the detailed clinical information can not be provided.

Patient 7 (c.965G>A, p.G322D), female, 43 years old, has a mild dilatation of the ascending aorta. In 2002, she was diagnosed with an abdominal aortic aneurysm requiring surgery and in the same year she had a

type B aortic dissection. No other features of MFS, LDS, EDS vascular type or other syndrome have been found. None of her family has the same features as hers.

Patient 8 (III.1) who has mutation c.980C>T, p.P327L is a 39 years old man with aortic root aneurysm and dissection, who had undergone replacement of aortic root. He has an increased arm span-height ration, his Beighton score is 2/9, and he has flat feet, positive left thumb sign, positive right wrist sign, myopia and few striae near axilla. His mother (II.2) has an aortic root dilatation, iris diaphania on temporal side, Beighton score 1/9 and positive right wrist sign. He has a maternal uncle (II.3) with aortic root dilatation (had undergone aortic root replacement) and maternal grandfather (I.1) who died because of aortic dissection. The DNA of mother and uncle showed the same mutation. The pedigree of this patient is shown below :

Figure 20. Pedigree of Patient 8
An autosomal dominant pattern of inheritance in which proband and her previous generation have the same features

Patient 9 (II.1) who has mutation c.1282T>G, p.Y428D is male, 45 years old diagnosed as having thoracic aortic aneurysm and dissection at the age of 35 years. His mother has a descending aortic aneurysm. No other features of MFS, LDS, EDS vascular type or other syndromes have been found.
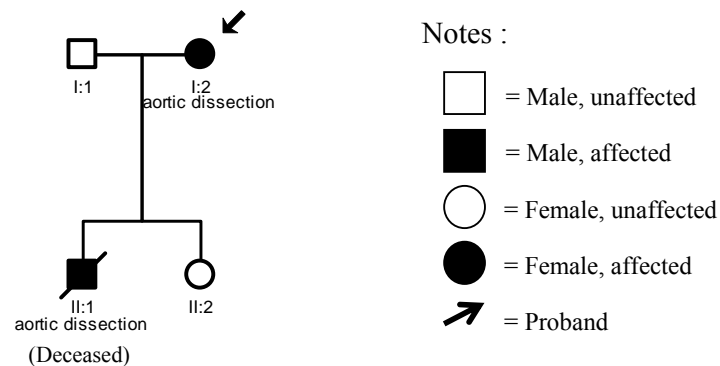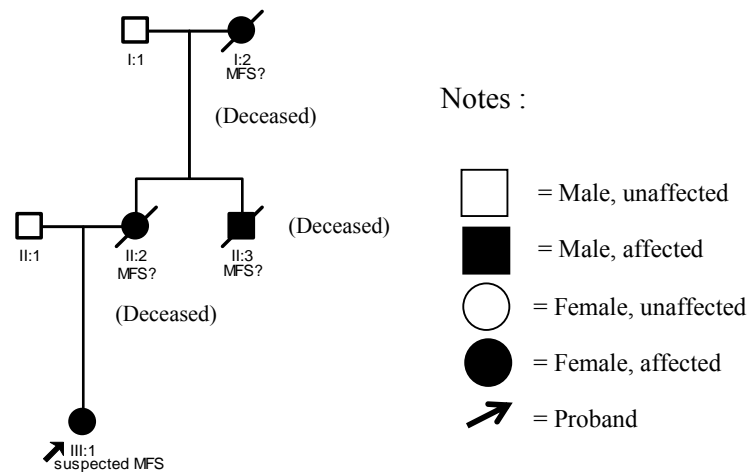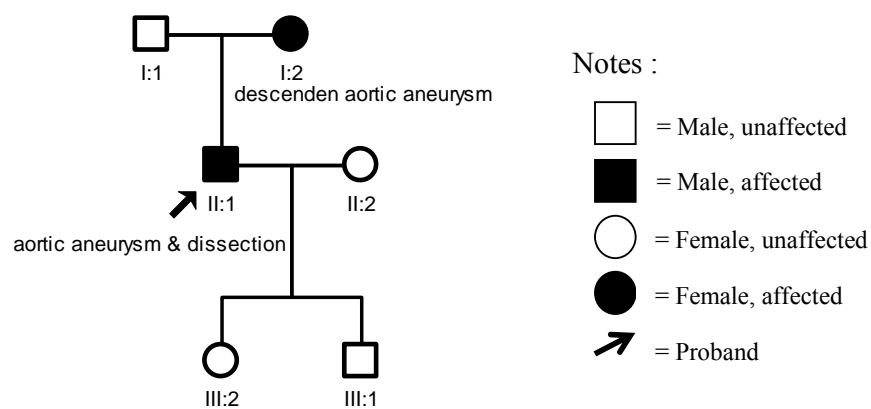
Figure 21. Pedigree of patient 9
An autosomal dominant pattern of inheritance in which proband and have the same clinical feature

Patient 10, (c.1460G>A, p.R487Q), female, 17 years old (IV.2), was diagnosed as having an aortic aneurysm and hypermobility of the joints at 10 years old. Her brother (IV.1) carries the same mutation and also having aortic aneurysms. No other features of MFS were apparent in these two patients. The pedigree is depicted in figure 7. I.2 died of aortic dissection at the age of 36 years, II-2 at the age of 50 years, III-2 at the age of 21 years
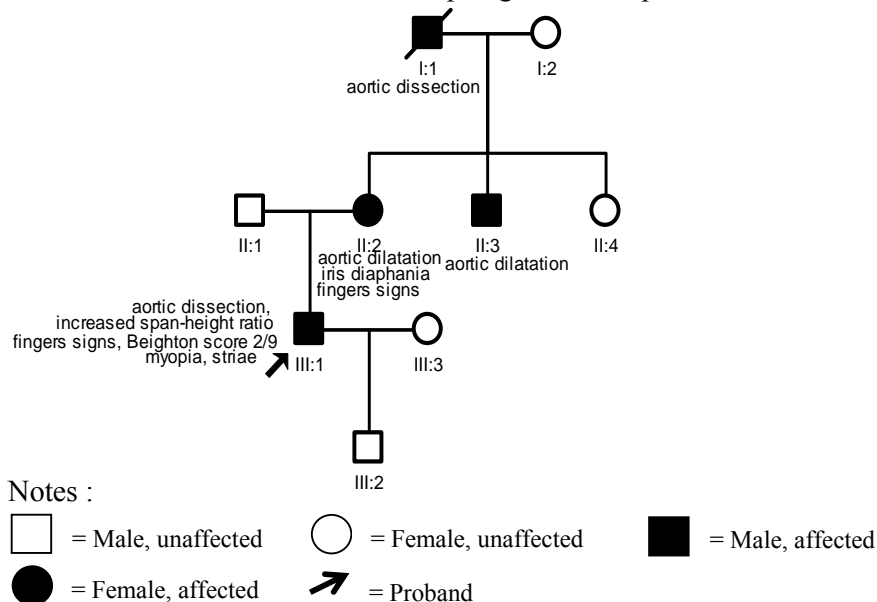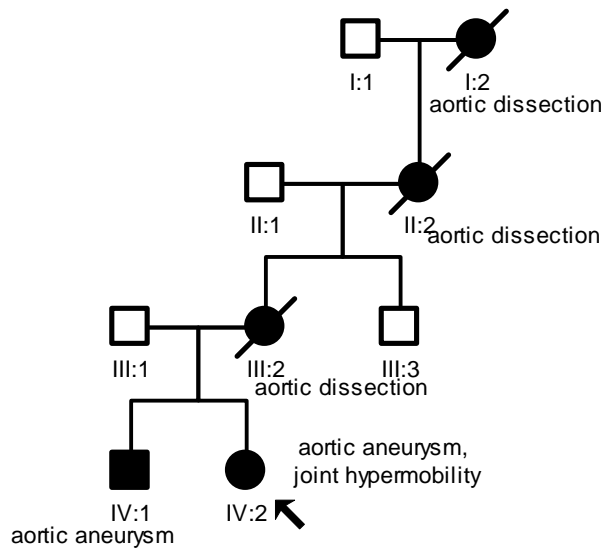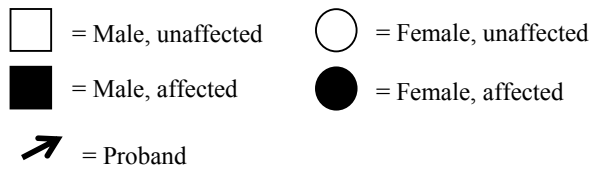
Figure 22. Pedigree of patient 10
An autosomal dominant pattern of inheritance in which proband, her sibling and her previous generation have the same clinical features

The summary of clinical and molecular findings in patients carrying the mutations is shown in Table 9.

Table 10. Clinical Findings of Patients with *TGFBR1* Mutations

Patient	Age (years)	Nucleotide Change	Clinical features	FH
1	58	c.113G>A; p.C38Y	thoracic aortic aneurysm	+
2	51	c.451C>T; p.R151C	thoracic aortic aneurysm	-
3	57	c.605C>T; p.A202V	thoracic aortic aneurysm and dissection	+
4	31	c.605C>T; p.A202V	Suspected Marfan Syndrome (details unknown)	+
5	25	c.839C>T; p.S280L	Tall, thin and long extremities, contracture of the hands, shoulder luxation habitualis, arachnodactily	-
6	53	c.958C>T; p.I320V	Suspected Marfan Syndrome (details unknown)	-
7	43	c.965G>A; p.G322D	abdominal aortic aneurysm requiring surgery, aortic dissection type B, ascendance aortic aneurysm	-
8	57	c.980C>T; p.P327L	Aortic root aneurysms and minor signs of MFS	+
9	45	c.1282T>G; p.Y428D	Aortic aneurysms	+
10	17	c.1460G>A; p.R487Q	Aortic aneurysms and joint hypermobility	+

Note : FH = Family History
(+) = present
(-) = absent

From the table above, it is shown that 7 out of 10 patients have aortic aneurysms as clinical features, and 6 patients have positive family history.

Chapter V

DISCUSSION

Most of the patients included in this study were those with suspected MFS with or without the presence of aortic aneurysm and/or dissection (78/194), aortic aneurysm and/or dissection (60/194), and familial cases of aortic aneurysm and/or dissection (42/194). The mutations found in this study is 10 mutations in 194 samples (10/194). Previously published results by Singh et al (2006) and Loeys et al (2006) on MFS patients without *FBN1* mutations, as reviewed by Mizuguchi T and Matsumoto N, yield 2/41 and 1/22, respectively.[34] The study by Matyas et al (2006) has found 4.0% (3/70) *TGFBR1* mutations on MFS-related patients without *FBN1* involvement.[11] Furthermore, combining the findings of *TGFBR1* mutations in LDS and Thoracic Aortic Aneurysms and Dissections (TAAD), it is very likely that *TGFBR1* mutations do play role in the pathogenesis of MFS and related disorders through TGFβ signaling, although the frequency is not significant. Mutation analysis on *TGFBR1* should be considered in MFS and its related disorders, without the presence of mutations in *FBN1* and *TGFBR2*.

All the mutations that have been found in this study are novel, except one mutation in exon 9 p.R487Q, so there is no previous publication about the pathogenicity of these mutations. The pathogenicity of mutations were analyzed based on the changes in amino acid, the amino acid conservation across the species, the protein domain where the mutation was occurred, in some patients we looked for the presence or absence of the mutation in affected or unaffected

family members respectively, and use internet-based prediction tools software PolyPhen and SIFT. The mutation p.C38Y is considered pathogenic from the multiple alignment analysis and PolyPhen analysis. The result from PolyPhen analysis predicted that the changes will disturb the formation of disulfide bond of the protein, thus will disturb the structure of the protein. The mutation p.R151C is predicted to be benign and tolerated by Polyphen and SIFT analysis, respectively. However, considering big changes in amino acid type and the domain conservation across multiple species, this mutation is considered as pathogenic. The mutation c.605C>T, p.A202V has been predicted to affect protein function based on SIFT analysis, but predicted as benign on PolyPhen. It occurred at a highly conserved domain. However, this mutation is unlikely to be pathogenic because the change in amino acid type is not significant. The mutation p.S280L is considered pathogenic because the change from serine to leucine is a significant change. Serine and threonine residues can be autophosphorylated, but not leucine. The mutation p.I320V occurred at highly conserved domain, predicted as possibly damaging by PolyPhen analysis and "tolerated" by SIFT analysis. However, this mutation is unlikely to be pathogenic because the change in amino acid type is not significant. The mutation p.G322D is predicted to be benign and tolerated on PolyPhen and SIFT analysis. But it is considered pathogenic, because of the significant change in amino acid, the conservation in 11 different species and the occurrence in the protein kinase domain. In patient with mutation p.P327L, the DNA of affected mother and affected uncle showed the same mutation. This mutation is predicted to be probably damaging and affects protein function by

both PolyPhen and SIFT analysis. The location of this mutation in protein kinase domain and the high conservation across 11 species, make this mutation strongly suggested as pathogenic. The mutation p.Y428D is predicted to be probably damaging and affects protein function by PolyPhen and SIFT analysis, respectively. With regard to a big change in amino acid, this mutation is considered pathogenic.

Mutation p.R487Q has been found previously to be pathogenic in other studies.[10,11,35] Akutsu et al (2007) found this mutation in patient with acute aortic dissection, mesenteric artery aneurysm and bilateral pneumothorax, without other features of MFS or LDS.[34] Matyas et al (2006) found this mutation in patient with thoracic aortic aneurysm and dissection, also without any features of MFS and LDS.[11] Loeys et al (2006) described this mutation in patients with thoracic aortic aneurysm and thoracic aortic dissection, without other features of MFS or LDS.[10] From those three previous publications, it seems that a mutation in this location causes aortic aneurysm and dissection, and is not likely to cause skeletal or eye abnormality.

Seven out of nine different mutations in this study occurred at highly conserved kinase domains, more specifically, the serine-threonine kinase domain. This domain is responsible for the formation of kinase, an enzyme that plays a role in cellular processes, including division, proliferation, apoptosis and differentiation.[35] Most of the mutations in *TGFBR1* that have been published are located in this domain.[10,11,13,20] Thus, it is strongly suggested that this domain has

a very crucial role in the formation of TGFβR1 and mutations in this domain are pathogenic.

A polymorphism in exon 1 (c.70_78delGCGGCGGCG), the 6Ala allele, was predicted to act as low penetrance allele of the clinical features in Marfan Syndrome.[37] This 6Ala allele has been previously associated with a higher risk of colorectal cancer, breast cancer and ovarian cancer. 6Ala/6Ala Homozygosity even leads to higher risk than 6Ala/9Ala heterozygosity.[38] Whether it acts in the same way in MFS and related disorders, however, needs a broader analysis which includes a large number of controls.

Among 194 patients with MFS and related disorders, we found 10 patients carrying *TGFBR1* mutations. Based on the data available, none of these patients was diagnosed clinically as MFS fulfilling the Ghent criteria, nor had features of LDS or other syndromes. These diseases have common nature, that the features might be shown to be age-dependent, thus a clinical follow up should be provided to confirm present diagnosis. In this study, all patients are more than 17 years old. Since they already exceed pubertal ages, the chance for developing new feature is not likely. But the existing feature still should be monitored for developing worse.

From 10 cases with *TGFBR1* mutations (7 pathogenics, 3 non pathogenics), seven out of ten patients with *TGFBR1* mutation have aortic aneurysm, with 3 of them also have minor features of MFS. Seven of them are familial cases. The exact phenotype due to *TGFBR1* mutations cannot be clearly concluded, since these patients have features ranging from isolated aortic aneurysm to skeletal abnormalities. Singh et al found *TGFBR1* mutations in

patients with typical Marfan Syndrome.[19] Loeys et al, Akutsu et al, Drera et al found mutations in patients with features of Loeys-Dietz syndrome.[21,35,39] Ades et al reported mutations in Furlong Syndrome.[12] Matyas et al found it in even larger variety of clinical features variation: in TAAD, LDS and typical MFS patients.[11] Thus, it is suggested that mutations in *TGFBR1* have variable clinical outcomes, indeed. The likeliness of patients with *TGFBR1* mutation having aortic aneurysms might be one sign that lead us to do *TGFBR1* mutation analysis in MFS and related disorders patient. It has been recognized also, that mutation in *TGFBR* genes rarely found in classic type of MFS. In this study we found none of patient with classic MFS has mutation in *TGFBR*.

Marfan Syndrome, Loeys-Dietz Syndrome, Familial Thoracic Aortic Aneurysms and Dissections and other Marfan-related disorders are inherited in autosomal dominant manner. An individual whose parent is carrying a heterozygous mutation of the gene causing this disorder will have 50% chance to develop the disorder. In this study, 6 familial cases with *TGFBR1* mutations have been found. In these families, genetic counseling should be provided to inform them about the risk and how to deal with the disorders in the future. The nature of MFS and aortic aneurysms is that the clinical presentations develops with age. Patients must be warned and counseled that once they are diagnosed, they need to be monitored for aortic widening, skeletal growth, etc. The medicinal treatment needs to be taken a life long, and surgical treatment may be needed.

Furthermore, mutations in *TGFBR* genes are related with more severe vascular manifestation with probably a shortened life expectancy.[7] In EDS

vascular type, there's a shortened survival, with the mean age is 48 years. The presence of tissue fragility make it difficult to do aorta repair, and increase the risk of visceral rupture.[40] Patients affected by Loeys-Dietz Syndrome have high risk of aortic aneurysm and dissection. But the repairing surgery is usually less complication and more successful in LDS compared to EDS vascular type.[10] In this case, a careful recognition on clinical presentation and molecular examination play an important role.

In the rest of the 184 patients, the causative gene responsible for their disorders has been left unknown. Further research on genes which play role in the TGF-β signaling pathway, such as *LTBP4*, or *TGFB* itself[9] will be needed. In patients having aortic aneurysms, other candidate genes which play role in maintaining the aortic structure, such as *COL3A1*, *ACTA2* and *MYH11*[41,42] may also be involved.

Chapter VI

CONCLUSION AND SUGGESTION

VI.1 CONCLUSION

1. Marfan Syndrome and related disorders patients who have no mutation in *FBN1* and *TGFBR2* genes, could be positive for *TGFBR1* gene mutation. Missense mutation is the commonest type of mutation which could be found in these individuals.

2. Among 9 different mutations found in this study, 7 mutations are considered pathogenic and 2 mutations are not pathogenic.

3. Clinical features of patients carrying the mutation are ranging from suspected Marfan Syndrome to isolated aortic aneurysm. None of the patients with classic MFS has mutation in *TGFBR1* gene.

4. Most of the patients carrying the *TGFBR1* gene mutation have aortic aneurysm as clinical feature.

IV.2 SUGGESTION

1. The mutation analysis in this study were done based on database references and software-based analysis. This still need a functional study to know the expression of TGF-β to confirm the pathogenicity of the mutations.

2. Genotype-phenotype correlation would be better seen in a larger number of patients. Thus, more samples are needed to see the *TGFBR1* mutation

frequency in higher population. The stricter inclusion criteria will be expected to have better conclusion in this matter, too.

3. Better study design in which we can follow the disease progress in MFS and related disorders patients should be used to get better understanding on the involvement and the impact of *TGFBR1* mutations in MFS and related disorders pathogenesis and severity.

4. This field of research has large potential to be explored in Indonesian population. The DNA samples from Indonesian patients is needed to perform the research. Centers which provide DNA sequencing services should be available.

5. When the cost of DNA sequencing becomes a major problem, mutation screening methods, such as MLPA (Multiplex Ligation Dependent Probe Amplification) may be considered. By mutation screening methods, the requirement for DNA sequencing can be significantly reduced.

6. Mutation analysis in other genes should be done on the rest of patients whose cause of the disorders are still unknown.

Chapter VII

SUMMARY

Marfan Syndrome (MFS), a common autosomal dominant inherited disorder of fibrous connective tissue, mostly affects three organ systems : skeletal, ocular and cardiovascular system. Cardiovascular involvements, the aortic aneurysms leading to aortic dissection or rupture, is the most life-threatening.

Diagnosis of MFS can be established by the Ghent criteria. However, the interpretation of these criteria is not always easy, due to the presence of many disorders which are clinically similar to MFS. Those disorders, termed as related disorders of MFS include Loeys-Dietz syndrome, Sphrintzen-Goldberg Syndrome, Familial Aortic Aneurysm, Bicuspid Aortic Valve with Aortic Dilatation, Familial Ectopia Lentis, MASS phenotype, Marfan Body Type, Mitral Valve Prolapse Syndrome, Congenital Contractural Arachnodactily (Beals syndrome), Stickler syndrome and Ehlers-Danlos syndrome.

Previously, the pathogenesis of MFS was explained based on the concept of fibrillin-rich micro fibrils as purely architectural elements in the extra cellular matrix. Mutations in the fibrillin-1 gene (*FBN1* gene), known to cause MFS, however, the mutations have not always been found in MFS patients.

Recent findings on the pathogenesis of MFS demonstrate changes in growth factor signaling and other changes in matrix-cell interactions. Mouse models of MFS with *FBN1* mutation which have lung emphysema as phenotypic manifestation, showed increased TGFβ signalling. The involvement of TGFβ-

receptor gene mutation in MFS has been shown in a Japanese patient with MFS who had a balanced chromosomal translocation involving chromosome 3p24. This locus had been found to show genetic linkage with MFS in a large French pedigree. The breakpoint in the Japanese patient disrupted the *TGFBR2* gene.

The proteins fibrillin-1, TGFBR1 and TGFBR2 take part in transforming growth factors β (TGFβ) signaling, thus mutations in one of these gene could cause similar phenotypes. Mutation analysis on *FBN1* and *TGFBR2* genes in MFS and related disorders have been well established, and is important to distinguish those Marfan spectrum disorders from one and another. However, since there are still many cases without any mutation in either *FBN1* or *TGFBR2,* mutation analysis on other candidate gene is needed to be performed. Mutation analysis on *TGFBR1* gene as one of candidate gene, which include the recognition of mutation and its kind, the prediction on pathogenicity, the distribution of phenotypes on genotypes and the recognition of clinical sign which may lead to this gene, are need to be done.

The *TGFBR1* gene is also known as activin A receptor like kinase, or serine / threonine-protein kinase receptor R4 gene. The DNA size is approximately 45kb long, the mRNA size is 2308bp, contains of 9 exons and is located on chromosome 9q22.33. The protein domains of TGFBR1 consist of extra cellular domain, transmembrane domain, cytoplasmic domain, glycine-serine rich domain, and serine-threonine kinase domain. These domains are highly conserved across species.

This research is in the field of medical genetics, held in the DNA Diagnostic Laboratory of Vrije Universiteit Medisch Centrum (VUmc), Amsterdam, The Netherlands. This is a descriptive study. The population of this research is the DNA samples of patients with Marfan Syndrome and related disorders which have been reffered to DNA Diagnostic Laboratory of VUmc Hospital Amsterdam, The Netherlands from the year 1998-2008. The DNA samples were donation with permission from Gerard Pals, PhD as the principal investigator of Connective Tissue Disorders research in the DNA Diagnostic Laboratory of VUmc Hospital Amsterdam, The Netherlands. All of the samples used in this research are part of Connective Tissue Disorders research project, and have been consent to be included in research.

One hundred and ninety four DNA of unrelated patients with MFS, suspected MFS, or related disorders, have been included. The inclusion criteria were having at least one major criterion of MFS and found to be negative for *FBN1* and *TGFBR2* mutations on previous examination. The samples were excluded if the amount of the DNA were not enough for further analysis. The phenotypic characteristics of the patients were then traced from their laboratory request form.

PCR was done to amplify the whole 9 exons of *TGFBR1*. The PCR products were then confirmed by gel electrophoresis, and underwent a pre-sequencing preparation before go to an automated sequencing machine. The results of the DNA sequencing were then analyzed for the presence of variants. When a variant has been found, the database of mutations and polymorphisms would be used to

confirm whether the variant is a mutation or polymorphism. When it was not in the database, then the following things would be considered to decide the pathogenicity : the amino acid changes, domain localization, conservation across species using multiple sequence alignment, and the prediction results from internet-based software : PolyPhen and SIFTblink.

The patients were grouped into several diagnoses based on clinical findings and matched with Ghent Criteria. A diagnosis of MFS was based on Ghent Criteria. Incomplete Ghent Criteria, or having at least one major criterion in an organ system with minor criterion of another organ, or more than one minor criterion, would be considered as Suspected MFS. The patients with only specific clinical features (such as only has aortic aneurysm, ectopia lentis, dural ectasia or joint hypermobility) would be grouped as the clinical findings, recognized as Marfan Syndrome, Suspected MFS, Aortic Aneurysms and/ Dissections, Familial Aortic Aneurysms and/ Dissections, Ectopia Lentis, Dural Ectasia, Joint Hypermobility. There are 10 MFS, 78 Suspected MFS, 60 Aortic aneurysms and/ Dissections, 42 Familial aortic aneurysms and dissections, 2 Ectopia Lentis, 1 Dural ectasia and 1 joint hypermobility patients.

On sequencing all 9 exons of *TGFBR1*, a total of 9 mutations, 7 different polymorphisms and 3 unclassified variants in *TGFBR1* were found. The mutations were found in 10 patients. The 9 mutations, occured in 7 different exons.

The first mutation c.113G>A; p.C38Y is located in exon 2 of *TGFBR1* gene, at the position 113 of cDNA, in which guanine is replaced by adenine, resulted in the change of amino acid 38 from cysteine (a polar-neutral amino acid) to tyrosine

(a polar-neutral), and is predicted to be pathogenic. This mutation is happened in patient with Familial aortic aneurysm and dissection.

The mutation c.451C>T; p.R151C is located in exon 3 of *TGFBR1* gene, at the position 451 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 151 from arginine (a polar-basic amino acid) to cysteine (a polar-neutral amino acid), and is predicted to be pathogenic. This mutation is present in patient with suspected MFS, with the clinical features aortic aneurysms and minor signs of MFS.

The mutation c.605C>T; p.A202V is located in exon 4 of *TGFBR1* gene, at the position 605 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 202 from alanine (a nonpolar-neutral amino acid) to valine (a nonpolar-neutral amino acid), and is predicted to be non pathogenic. This mutation occurs in patient with familial thoracic aortic aneurysms and dissection.

The mutation c.839C>T; p.S280L is located in exon 5 of *TGFBR1* gene, at the position 839 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 280 from serine (a polar-neutral amino acid) to leucine (a nonpolar-neutral amino acid), and is predicted to be pathogenic. This mutation happened in suspected MFS patient, with the clinical signs tall and long extremities, contractures of the hands, recurrent shoulder luxation and arachnodactyly.

The mutation c.958A>G; p.I320V is located in exon 5 of *TGFBR1* gene, at the position 958 of cDNA, in which adenine is replaced by guanine, resulted in the change of amino acid 320 from isoleucine (a nonpolar-neutral amino acid) to

valine (a nonpolar-neutral amino acid), and is predicted to be non pathogenic. This mutation occurred in patient with suspected MFS.

The mutation c.965G>A; p.G322D is located in exon 5 of *TGFBR1* gene, at the position 965 of cDNA, in which guanine is replaced by adenine, resulted in the change of amino acid 322 from glycine (a nonpolar-neutral amino acid) to aspartic acid (a polar-acidic amino acid), and is predicted to be pathogenic. This mutation occurred in patient with aortic aneurysms and dissections.

The mutation c.980C>T; p.P327L is located in exon 6 of *TGFBR1* gene, at the position 980 of cDNA, in which cytosine is replaced by timine, resulted in the change of amino acid 327 from proline (a nonpolar-neutral amino acid) to leucine (a nonpolar-neutral amino acid) and is predicted to be pathogenic. This mutation occurred in patient with suspected MFS, with aortic aneurysms and minor signs of MFS.

The mutation c.1282T>G; p.Y428D is located in exon 8 of *TGFBR1* gene, at the position 1282 of cDNA, in which timidine is replaced by guanine, resulted in the change of amino acid 428 from tyrosine (a polar-neutral amino acid) to aspartic acid (a polar-acidic amino acid), and is predicted to be pathogenic. This mutation occurred in patient with aortic aneurysms.

The mutation c.1460G>A; p.R487Q is located in exon 9 of *TGFBR1* gene, at the position 1460 of cDNA, in which guanine is replaced by adenine, resulted in the change of amino acid 487 from arginine (a polar-basic amino acid) to glutamine (a polar-neutral amino acid), and is a pathogenic mutation. This

mutation occurred in patient with aortic aneurysms and dissection with joint hypermobility.

REFERENCES

1. OMIM (Online Mendelian Inheritance in Man). Marfan Syndrome #154700. Available at URL :
   http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=154700

2. Colloud-Beroud G, Boileau C. Marfan syndrome in the third millenium. Eur J Hum Genet. 2002; 10 : 673-681

3. Millewicz DM, Dietz HC, Miller DC. Treatment of aortic disease in patients with Marfan Syndrome. Circulation. 2005 ; 111 : e150-e157

4. Robinson PN, Arteaga-Solis E, Baldock C, Collod-Beroud G, Booms P, De Paepe A, et al. The molecular genetics of marfan syndrome and related disorders. Journal of Medical Genetics 2006; 43 : 769-787

5. De Paepe A, Devereux RB, Dietz HC, Hennekam RCM, Pyeritz RE. Revised diagnostic criteria of marfan syndrome. American Journal of Medical Genetics. 1996 ; 62 : 417-476

6. Neptune ER, Frischmeyer PA, Arking DE, Myers L, Bunton T, Gayraud B, et al. Dysregulation of TGFβ activation contributes to pathogenesis in marfan syndrome. Nature Genetics. 2003 (33); 407-411

7. Mizuguchi T, Collod-Beroud G, Akiyama T, Abifadel M, Harada N, Morisaki T, et al. Heterozygous *TGFBR1* mutations in marfan syndrome. Nature Genetics Advance Online Publication. 2004. doi:10.1038/ng1392

8. Habashi JP, Judge DP, Holm TM, Cohn RD, Loeys BL, Cooper TK, et al. Losartan, an AT1 antagonist, prevents aortic aneurysm in a mouse model of marfan syndrome. Science. 2006; 312 : 117

9. Ten Dijke P, Arthur HM. Extracellular control of TGFβ signalling in vascular development and diseases. Nature Reviews on Molecular Cell Biology. 2007 ; 7 : 857-869

10. Loeys BL, Schwarze U, Holm T, Callewaert BL, Thomas GH, Pannu H, et al. Aneurysm syndromes caused by mutations in the TGF-beta receptor. N Engl J Med. 2006 Aug 24;355(8):788-98.

11. Matyas G, Arnold E, Carrel T, Baumgartner D, Boileau C, Berger W, Steinmann B. Identification and in silico analysis of novel TGFβR1 and TGFβR2 in MFS-related disorders. Hum Mutat. 2006 ; 27(8) : 760-769

12. Ades LC, Sullivan K, Biggin A, Haan EA, Brett M, Holman KJ, et al. *FBN1, TGFBR1*, and the Marfan-craniosynostosis/mental retardation disorders revisited. Am J Med Genet A. 2006 May 15;140(10):1047-58.

13. UMD (Universal Mutation Database) for *TGFBR1* mutations. Available at URL: http://www.umd.be

14. Dietz, HC. Gene Review : Marfan syndrome, 2005. Available at http://www.genetesthttp://www.genetests.org/servlet/access?db=geneclinics&site=gt&id=8888891&key=eK2Jaf8ktIuY&gry=&fcn=y&fw=fcPS&filename=/profiles/marfan/index.html

15. Dean, JCS. Marfan Syndrome : clinical diagnosis and management. Eur J Hum Genet. 2007; 15:724-733

16. GenAtlas : *TGFBR1*. Available at URL : http://genatlas.medecine.univ-paris5.fr/fiche.php?symbol=TGFBR1

17. OMIM (Online Mendelian Inheritance in Man). Transforming growth factor-beta receptor, type I ; TGFBR1*190181. Available at URL : http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=190181

18. Ace view : a comprehensive cDNA-supported genes and transcript annotation, Genome Biology 2006, 7(Suppl 1):S12

19. Singh KK, Rommel K, Mishra A, Karck M, Haverich A, Schmidtke J, Arslan-Kirchner M. *TGFBR1* and *TGFBR2* mutations in patients with features of Marfan syndrome and Loeys-Dietz syndrome. Hum Mutat. 2006 Aug;27(8):770-7.

20. Loeys Bl, Chen J, Neptune ER, Judge DP, Podowski M, Holm T, et al. A syndrome of altered cardiovascular, craniofacial, neurocognitive and skeletal development caused by mutations in *TGFBR1* or *TGFBR2* (Letters). Nature Genetic. 2005; 37(3), 275-281

21. Kosaki K, Takahashi D, Udaka T, Kosaki R, Matsumoto T, Ibe S, et al. Molecular pathology of Sphrintzen-Goldberg syndrome. Am J Med Genet A. 2006; 140:104-10

22. Bell J, Bodmer D, Sistermans E, Ramsden SC. Practice guidelines for the interpretation and reporting of unclassified variants (Uvs) in clinical molecular genetics. UK Clinical Molecular Genetics Society (CMGS) and The Dutch

Society of Clinical Genetic Laboratory Specialists (Vereniging Klinisch Genetische Laboratoriumspecialisten; VKGL). 2007.

23. Strachan T, Read AP. Human molecular genetics 3. Garland Publishing, London and New York. 2004; 317-321.

24. Pauline C Ng, Steven Henikoff. Predicting the effects of amino acid substitutions on protein function. Annu Rev. Genomics Hum.Genet.2006.7:61-80

25. PolyPhen : prediction of functional effect of human nsSNPs. Available at URL : http://coot.embl.de/PolyPhen/

26. Polyphen help. Available at URL : http://genetics.bwh.harvard.edu/pph/pph_help.html

27. Sorting Intolerant From Tolerant (SIFT) Available at URL : http://blocks.fhcrc.org/sift/SIFT_BLink_submit.html

28. Madiyono B. Perkiraan besar sampel. In Dasar-Dasar Metodologi Klinis. Editor: Sastroasmoro,et.al. Second Edition. Sagung Seto, Jakarta: p:270

29. Gene sequence information for ENSG00000106799. Available at URL : http://www.ensembl.org/Homo_sapiens/geneseqview?db=core;gene=ENSG00000106799

30. Ensembl transcript report. Available at URL : http://www.ensembl.org/Homo_sapiens/transview?db=core;transcript=ENST00000374994

31. UniprotKB/Swiss-Prot P36897 (TGF-beta R_Human). Available at URL : http://www.uniprot.org/uniprot/P36897#section_features

32. Human Splicing Finder. Available at URL : http://www.umd.be/HSF/

33. Akutsu K, Morisaki H, Takeshita S, Sakamoto S, Tamori Y, Yoshimuta T, et al. Phenotypic heterogeneity of Marfan-like connective tissue disorders associated with mutations in the transforming growth factor-beta receptor genes. Circ J. 2007 Aug;71(8):1305-9.

34. Sakai H, Visser R, Ikegawa S, Ito E, Numabe H, Watanabe Y, et al. Comprehensive genetic analysis of relevant four genes in 49 patients with Marfan syndrome or Marfan-related phenotypes. Am J Med Genet A. 2006 Aug 15;140(16):1719-25.

35. TGFBR1 HomoloGene. Available at URL :
http://www.ncbi.nlm.nih.gov/sites/entrez?cmd=Retrieve&db=homologene&dopt=MultipleAlignment&list_uids=3177

36. Mizuguchi T, Matsumoto N. Recent progress in genetics of Marfan Syndrome and Marfan-associated disorders. J Hum Genet (2007) 52:1-12

37. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man (Review). TRENDS in Biochemical Sciences (2002) 27(10) : 514-520

38. Lucarini L, et al. May TGFBR1 act also as low penetrance allele in Marfan Syndrome? Int J Cardiol (2007), doi:10.1016/j.ijcard.2007.07.048

39. De la Chapelle, A. Genetic predisposition to colorectal cancer. Nature Reviews : Cancer (2004) 4:769-780

40. Drera B, Tadini G, Barlati S, Colombi M. Identification of a novel *TGFBR1* mutation in a Loeys-Dietz syndrome type II patient with vascular Ehlers-Danlos syndrome phenotype. Clin Genet. 2007 Dec 6;. [Epub ahead of print]

41. Child AH, Neumann L, Robinson PN. Diagnosis and treatment of Marfan Syndrome—a summary, in : Marfan Syndrome : A Primer for Clinicians and Scientists. Kluwer Academics / Plenum Publisher. New York : 2004; 13-23

42. Pepin MMS, Schwarze U, Superti-Furga A, Byers PH. Clinical and genetic features of Ehlers-Danlos Type IV, the vascular type. N Engl J Med. 2000 ; 342(10):673-680

43. Patel HJ, Deeb GM. Ascending and arch aorta : pathology, natural history and treatment. Circulation. 2008;118:188-195

44. Millewicz DM, Guo DC, Tran-Fadulu V, Lafont AL, Papke CL, Inamoto S, Pannu H. Genetic basis of thoracic aortic aneurysms and dissections : focus on smooth muscle contractile dysfunction. Annu. Rev. Genomic Hum. Genet. 2008; 9:283-30

45. Krivokapich J, Child JS, Dadourian BJ, Perloff JK. Reassessment of echocardiographic criteria for diagnosis of mitral valve prolapse. Am J Cardiol. 1988; 61:131-135

46. Sponseller P, Shindle M. Orthopedic Problems in marfan Syndrome., in Marfan Syndrome : A Primer for Clinicians and Scientists. Kluwer Academics / Plenum Publisher. New York : 2004; 26-27

47. Von Kodolitsch Y, Rybczynski M. Cardiovascular aspect in marfan syndrome : a systematic review. In : Marfan Syndrome : A Primer for Clinicians and Scientists. Medical Intelligent Unit, Plenum Publisher. 2004;45-69

ATTACHMENT 1

Ghent Criteria of Marfan Syndrome[5]

Diagnostic requirements :

Index case:
Major criteria in 2 different organ systems
AND involvement of a third organ system.

Relative of index case:
1 major criterion in family history
AND 1 major criterion in an organ system
AND involvement in second organ system.

SKELETAL

Major (Presence of at least 4 of the following manifestations)

__ pectus carinatum
__ pectus excavatum requiring surgery
__ reduced upper to lower segment ratio (Note 1)
OR arm span to height ratio >1.05
Height ____ Arm span ____ Upper segment ____ Lower segment ____
__ wrist (Note 2) and thumb (Note 3) signs
__ scoliosis of >20° or spondylolisthesis
__ reduced extension at the elbows (<170°)
__ medial displacement of the medial malleolus causing pes planus
__ protrusio acetabulae of any degree (ascertained on radiographs)

Minor

__ pectus excavatum of moderate severity
__ joint hypermobility
__ high arched palate with crowding of teeth
__ facial appearance
__ dolichocephaly,
__ malar hypoplasia,
__ enophthalmos,
__ retrognathia,
__ down-slanting palpebral fissures

__ INVOLVEMENT: 2 major criteria or 1 major and 2 minor

OCULAR

Major
__ ectopia lentis

Minor
__ flat cornea
__ increased axial length of the globe
__ hypoplastic iris OR hypoplastic ciliary muscle causing decreased miosis

__ INVOLVEMENT: 2 minor criteria

CARDIOVASCULAR

Major
__ dilatation of the ascending aorta with or without aortic regurgitation
and involving at least the sinuses of Valsalva
__ dissection of the ascending aorta

Minor
__ mitral valve prolapse with or without mitral valve regurgitation
__ dilatation of the main pulmonary artery, in the absence of valvular or
peripheral pulmonic stenosis below the age of 40 years
__ calcification of the mitral annulus below the age of 40 years
__ dilatation or dissection of the descending thoracic or abdominal aorta
below age of 50 years

__ INVOLVEMENT: 1 minor criterion

PULMONARY

Minor (only)
__ spontaneous pneumothorax
__ apical blebs

__ INVOLVEMENT: 1 minor criterion

SKIN AND INTEGUMENT

Minor (only)
__ striae atrophicae
__ recurrent or incisional hernia

__ INVOLVEMENT: 1 minor criterion

DURA

Major
__ lumbosacral dural ectasia by CT or MRI

FAMILY/GENETIC HISTORY

Major
__ first degree relative who independantly meets the diagnostic criterian.
__ presence of mutation in FBN1 known to cause Marfan syndrome
__ presence of haplotype around FBN1 inherited by descent and unequivocally
associated with diagnosed Marfan syndrome in the family

ATTACHMENT 2

DIAGNOSTIC CRITERIA OF SOME CONDITIONS OVERLAPPING WITH MARFAN SYNDROME

1. **Loeys-Dietz Syndrome**

   General :

   - Widely-spaced eyes (hypertelorism),
   - Bifid uvula,
   - Generalized arterial tortuosity with widespread arterial aneurysms and dissection

   Loeys-Dietz Syndrome type 1 :

   - If craniofacial involvement consisting of cleft palate, craniosynostosis and hypertelorisms were observed

   Loeys-Dietz Syndrome type 2 :

   - No evidence of craniofacial involvement but only isolated bifid uvula

2. **Ehler-Danlos Syndrome**

   General :

   - Skin hyperextensibility,
   - Joint hypermobility,
   - Easy bruising,
   - Tissue fragility,
   - Mitral valve prolapse,
   - Aortic dilatation (uncommon)
   - Chronic joint and limb pain

   Classic type :

   - Inheritance : autosomal dominant
   - Major criteria : skin hyperextensibility, widened atrophic scars, joint hypermobility

- Minor criteria : smooth, velvety skin, molluscoid pseudotumors, muscle hypotonia, easy bruising, hiatal hernia, anal prolapse, positive family history

Hypermobility type :

- Inheritance : autosomal dominant
- Major criteria : hyperextensibility and or smooth velvety skin, generalized joint hypermobility
- Minor criteria : recurring joint dislocation, chronic joint/limb pain, positive family history

Vascular type :

- Inheritance : autosomal dominant
- Major criteria : thin, translucent skin, arterial/intestinal/uterine fragility or rupture, extensive bruising, characteristic facial appearance
- Minor criteria : acrogeria, hypermobility small joints, tendon and muscle rupture, clubfoot, early-onset varicose veins, arteriovenous or carotid-cavernous sinus fistula, pneumothorax, gingival recession, positive family history

3. **MASS phenotype :**
   - Mitral valve prolapse,
   - Aortic root diameter at the upper limit of normal,
   - Stretch mark (striae),
   - Skeletal features of Marfan (joint hypermobility, pectus excavatum/carinatum, scoliosis)

4. **Congenital Contractural Arachnodactily (Beals Syndrome)**
   - Inability to fully extend multiple joints such as fingers, elbows, knees, toes, and hips
   - Crumpled ear
   - Arachnodactily

- Scoliosis

- Kyphoscoliosis

- Osteopenia

- Dolichostenomelia

- Pectus excavatum or pectus carinatum

- Muscular hypoplasia

- Micrognathia

- High-arched palate

5. Mitral Valve Prolapse Syndrome[45]

Mitral valve prolapse with the signs :

Auscultation :

- Unequivocal mid- to late-systolic click, late systolic apical murmur, or both

Echocardiographic :

- Severe bowing of leaflets

- Coaptation of leaflets on the atrial side of the mitral annulus

- Moderate to severe Doppler mitral regurgitation with any leaflet bowing

- Mild Doppler mitral regurgitation with moderate bowing

6. Ectopia Lentis

The displacement of the lens, also named dislocation or subluxation due to an increasing elongation of the zonula fibres.

7. Dural Ectasia[46]

Widening of dural sac, with the criteria (developed by Ahn et al) The sagittal width of the dural sac at S1 or below is greater than the width of the dural sac above L4, or the presence of anterior meningocele (major criterion). Minor criteria : a nerve root sleeve at L5 > 6.5 mm in diameter or scalloping at S1 > 3.5 mm.

8. Sphrintzen-Goldberg Syndrome

- Omphalocele
- Scoliosis
- Laryngeal/pharyngeal hypoplasia
- Mild dysmorphic face
- Learning disabilities

ATTACHMENT 3
DIAGNOSTIC CRITERIA OF AORTIC ANEURYSMS

The classical approach to assess aortic root dimensions is to use M-mode echocardiography with measurements from the most anterior portion of the anterior aortic wall to the most anterior portion of the posterior aortic wall at end-diastole; in subjects ≥ 16 years of age dilatation of the aortic root is present with at least two of the following criteria:

1. width index of the aorta > 22 mm/m2,

2. aortic diameter > 37 mm

3. left atrial to aortic diameter ratio < 0.7.9

In addition, M-mode nomograms are available to compare aortic root dimensions at the sinuses of Valsalva with body surface area. More recently, two-dimensional echocardiography is used to assess aortic root dimensions at the level of the valve annulus, the aortic sinuses, the sinotubular junction and the proximal ascending aorta; such measurements are systematically larger (2 mm at the level of the aortic sinuses) than those made by M-mode echocardiography.

Currently, two-dimensional echocardiography is used to diagnose aortic root dilatation by means of nomograms relating aortic root size to body surface area; such nomograms are available for children < 18 years of age, for adults < 40 years of age and for adults ≥ 40 years of age;11 the use of these nomograms is recommended by the Ghent nosology and current European guidelines. In addition, adjusted nomograms are available for adults exceeding the 95th

percentile for body height ($\geq$ 189 cm in men; $\geq$ 175 cm in women) and for children with suspected MFS (who are shown to present with a body surface area above the 50th percentile despite exclusion of MFS).

Aortic ratios allow for comparison of individuals irrespective of age and body size. For calculation of an aortic ratio, the observed maximum diameter of the aortic root is divided by the predicted diameter based on age and body surface area (BSA) of normal individuals. The predicted sinus diameter (cm), for instance, can be calculated using the following regression formulas:

• in children (age < 18 years) = 1.02 + (0.98 x BSA (m2));

• in adults (age 18-40 years) = 0.97 + (1.12 x BSA (m2));

• in adults (age $\geq$ 40 years) = 1.92 + (0.74 x BSA (m2)).

Thus, an aortic sinus ratio of 1.3 indicates a 30 percent enlargement of the aortic sinus above the mean of normal individuals of the respective age and body surface area. Nomograms are less helpful in adults over 40 years of age, because obesity and aortic media degeneration account for a looser relationship between aortic size and body surface area; as a rule of thumb, in these individuals the aortic root is normal with diameters of < 37 mm, the ascending aorta is dilated with diameters $\geq$ 38 mm and < 50 mm, and aneurysm is present with diameters $\geq$ 50 mm.

ATTACHMENT 4

LABORATORY REQUEST FORM AND INFORMED CONSENT

VU medisch centrum

Afdeling Klinische Genetica eiwitdiagnostiek
Sectie Genoomdiagnostiek

Laboratorium voor DNA- en eiwitdiagnostiek
Afdeling Klinische Genetica - VUMC
Postbus 7057; intern BS7-J379
1007 MB AMSTERDAM
afleveradres voor koeriers:
v.d. Boechorststraat 7, 3de etage kamer J379
1081 BT AMSTERDAM
Klinisch moleculair genetici
Dr. E.A. Sistermans (hoofd)
Dr. J.J.P. Gille (subhoofd)
Dr. G. Pals (hoofd research)
Dr. G.S. Salomons
Secretariaat
Tel : 020-4448346; Fax: 020-4448293
E-mail: DNAdiagnostiek@vumc.nl
website: www.vumc.nl/genoomdiagnostiek

Secretariaat
Tel : 020-4448346; Fax: 020-4448293
E-mail: DNAdiagnostiek@vumc.nl
website: www.vumc.nl/genoomdiagnostiek

Aanvraag DNA- en

Naam
Geb. datum
Geslacht
Adres

Verzekering
Huisarts

per persoon een aanvraagformulier invullen

_____

Aanvrager
naam:	telefoonnummer:
zh/instelling:	afdeling:
adres:	uw referentie:
plaats:	c.c. uitslag:

_____

Materiaal
2 x 7 ml EDTA ontstold bloed (kleine kinderen 2 x 3 ml) voorzien van naam + geb. datum verzenden per post bij
kamertemperatuur. Monsters die niet zijn voorzien van een deugdelijke identificatie worden geweigerd. Voor sommige
indicaties is een huidbiopt of een fibroblastenkweek noodzakelijk (zie pag. 2). **Datum afname:**

_____

Indicatie
Aangeven in de tabel op pagina 2.
Relevante klinische gegevens:
Vraagstelling
  bevestigen/uitsluiten klinische diagnose   overig
  prenataal onderzoek (**vooraf aanmelden**)   opslag, nl. voor:
  screening op bekende mutatie in de familie, nl.:

_____

Is er al eens eerder materiaal van deze patiënt of van een familielid ingestuurd?
  Nee

Ja, nl. naam: geb. datum: ref. nr.

Stamboom (eventueel aparte stamboom meesturen):

Betrokkene geeft **geen** toestemming voor anoniem gebruik van lichaamsmateriaal voor research (zie 5.3 op pag. 3).

_____

In te vullen door het laboratorium
ZIS-nr.:
familienummer:
VD-nummer aanwezig materiaal:

ontvangen materiaal: paraaf staflid:

Indicaties voor DNA-onderzoek

Achondroplasie (FGFR3)
Alzheimer
PSEN1
PSEN2
APP
Apert syndroom
Azoöspermie/oligospermie (CFTR)
Azoöspermie/oligospermie (AZFa/b/c deleties)
Basaal Cel Nevus syndroom (PTCH)
Birt-Hogg-Dubé syndroom (FLCN)
Blackfan-Diamond anemie (RPS19)
Borst- en ovariumkanker
BRCA1
BRCA2
BPES (Blepharophimosis, ptosis, en epicanthuis inversus syndroom; FOXL2)
CBAVD (CFTR)
Chorea, erfelijke benigne (TITF1)
Craniosynostose (FGFR2, TWIST)
Crouzon syndroom
Cystic fibrosis (CFTR)
Darmkanker, Lynch syndroom
MLH1
MSH2
MSH6
Darmkanker, MUTYH geassocieerde adenomateuze polyposis
DiGeorge syndroom (22q11-deletie)
Ehlers-Danlos syndroom
COL3A1 (fibroblastenkweek of huidbiopt nodig)
COL5A1 (fibroblastenkweek of huidbiopt nodig)
Elastine (ELN)
Fanconi anemie (alleen na overleg)
Fragiele X syndroom (FRAXA)
Frontotemporale dementie
MAPT
PGRN
CHMP2B
Gorlin syndroom (PTCH)
Hyperferritinemie-cataract syndroom (FTH1)
Hypochondroplasie (FGFR3)
Langer mesomele dysplasie (SHOX)
Loeys-Dietz syndroom
TGFBR1

TGFBR2
Marfan syndroom
FBN1
TGFBR2
Maternale contaminatie
MLPA microdeletie syndromen (o.a. 22q11 en Williams syndr.)
MLPA subtelomeren
Obesitas (MC4R)
Osteogenesis imperfecta
COL1A1
COL1A2
Parkinson, ziekte van
Parkin (Park2)
DJ-1 (Park7)
Pink1 (Park6)
SNCA (Park4)
LRRK2 (Park8)
Pelizaeus-Merzbacher, ziekte van (PLP1)
Pelizaeus-Merzbacher-like disease, autosomaal recessief (GJA12)
Peutz-Jeghers syndroom (STK11)
Pfeiffer syndroom (FGFR2, FGFR3)
Porencephalie (COL4A1)
Prematuur ovarieel falen (FMR1 premutaties)
Pulmonale arteriële hypertensie, idiopatische (BMPR2)
Schmid dysplasie (COL10A1)
Saethre-Chotzen syndroom (FGFR3/TWIST)
Surfactant proteïne B deficiëntie (SFTPB)
Thanatofore dysplasie (FGFR3)
Uniparentale disomie (UPD)
Van de Woude syndroom (IRF6)
Andere indicatie (alleen na telefonisch overleg)

Overig DNA-onderzoek

Onderzoek dat **uitsluitend** kan worden aangevraagd na overleg
met prof. dr. M.S. van der Knaap, kinderneuroloog (ms.vanderknaap@vumc.nl)
Megalencephalic leukoencephalopathy with subcortical cysts (MLC1)
Leukoencephalopathie with vanishing white matter (VWM)
Leukoencephalopathy with brain stem and spinal cord involvement and lactate elevation (LBSL)

Onderzoek dat wordt verricht binnen het Metabool Laboratorium van het VUmc (Dr. G.S. Salomons)

Hiervoor een ander formulier gebruiken dat kan worden aangevraagd bij g.salomons@vumc.nl
• Alexander, ziekte van (GFAP)
• Canavan, ziekte van (ASPA)
• Cerebraal creatine deficiëntie syndroom (AGAT, GAMT,
SLC6A8)
• D-2-hydroxyglutaric dehydrogenase deficiëntie (D2HGDH)
• Epilepsie, pyridoxine afhankelijke (ALDH7A1
• GABA metabolisme (ALDH5A1, SSADH, GABA-T)
• Glutaryl-CoA dehydrogenase deficiëntie (GCDH)
• Homocysteine metabolisme (CBS, MTHFR,MMACHC)
• L-2-hydroglutaric dehydrogenase deficiëntie, (L2HGDH)
• Malonyl CoA decarboxylase deficiëntie (MLYCD)
• Ribose-5-phosphate isomerase deficiëntie (RPI)
• Tarui, ziekte van (PFKM)
• Transaldolase deficiëntie (TALDO)
• X-gebonden creatine transporter defect (SLC6A8)

Indicaties voor eiwitonderzoek
Fibroblastenkweek voor enzymonderzoek elders*
Osteogenesis imperfecta, type ___*
Ehlers-Danlos syndroom type ___*
Primaire Ciliaire Dyskinesie/Kartagener syndroom (respiratoir epitheelbiopt nodig)
* hiervoor is inzending van een fibroblastenkweek of een huidbiopt noodzakelijk

Huidbiopten
Afname:
• huidbiopten onder steriele condities afnemen, na desinfectie met 70% alcohol (**geen jodiumtinctuur**) bij
voorkeur aan de binnenkant van de onderarm of tijdens een operatie van de randen van de incisieplaats.
• Het biopt opvangen in steriel kweekmedium (op verzoek kan dit toegezonden worden). Alleen in noodgevallen een steriele
fysiologische zoutoplossing gebruiken.
• Indien buiten normale laboratoriumwerktijden een biopt moet worden afgenomen, het materiaal bewaren bij kamertemperatuur
(**niet op ijs**) en de volgende werkdag versturen.
Verzending
• het materiaal bij voorkeur op maandag, dinsdag of uiterlijk woensdag inzenden per TPG post. Op andere dagen alleen via een koerier.
• het materiaal goed inpakken ter bescherming tegen breuk en forse temperatuurdalingen.
• op het pakje vermelden "breekbaar" en "bewaren bij kamertemperatuur".

1. Aanvragen
1.1. Om fouten en vertragingen te vermijden behoren aanvragen op een duidelijke en ondubbelzinnige wijze te worden ingediend. Door gebruik te maken van dit aanvraagformulier komen alle gewenste gegevens aan de orde.
1.2. Met de acceptatie van een aanvraag verplicht de laboratorium zich tot het met zorg en vakmanschap uitvoeren van de gevraagde werkzaamheden volgens de voor de laboratorium geldende kwaliteitscriteria.
1.3. Aanvragen kunnen worden geweigerd indien deze onvoldoende gegevens bevatten om een resultaat te kunnen bereiken dat voldoet aan de geldende kwaliteitscriteria.
1.4. Het laboratorium moet in de gelegenheid gesteld te worden om met de aanvrager/behandelaar te kunnen overleggen over het gevraagde onderzoek.
1.5. De aanvrager wordt verzocht om alvorens patiëntenmateriaal in te sturen, na te gaan of de betreffende patiënt is verzekerd voor klinisch genetische zorg. Indien na uitvoering van een verrichting de patiënt niet verzekerd blijkt, wordt de rekening naar de patiënt gestuurd.
2. Monsters
2.1. De aanvrager levert de te onderzoeken monsters aan bij het laboratorium, voorzien van een deugdelijke identificatie (naam en geboortedatum) en een volledig ingevuld aanvraagformulier.
2.2. Per patiënt 2 x 7 ml EDTA bloed afnemen in onbreekbare buizen (geen glazen buizen), bij kleine kinderen 2 x 3 ml, en per post opsturen bij kamertemperatuur.
2.3. Indien niet wordt voldaan aan het gestelde in 2.1 en 2.2 is het laboratorium niet gehouden het ingestuurde monster in ontvangst te nemen.
2.4. Voor zover bij de indiening van de aanvraag daarover niets is overeengekomen, zal het laboratorium de monsters, c.q. de restanten daarvan na onderzoek, overeenkomstig de eigen voorschriften voor onbepaalde tijd bewaren.
2.5. Alle handelingen en opslag voorafgaand aan de in ontvangstname van een monster vallen buiten de verantwoordelijkheid van het laboratorium.

3. Resultaten

3.1. Resultaten in de vorm van onderzoeksuitslagen, adviezen, informatie of welke andere vorm dan ook, worden door het laboratorium in schriftelijke vorm aangeleverd.

3.2. Resultaten komen doorgaans beschikbaar binnen:

• Prenataal onderzoek: 2-3 weken

• Presymptomatisch / dragerschapbepaling / bevestiging diagnose (bekende mutatie): 6-8 weken

• Mutatie scanning (opsporen van nog onbekende mutatie): 3-6 maanden. In geval van spoed kunnen in overleg andere uitslagtermijnen worden afgesproken.

4. Geheimhouding

4.1. Geheimhouding van gegevens is gewaarborgd en vastgelegd in de ziekenhuisvoorschriften van het VU medisch centrum (zwijgplicht over patiëntengegevens).

5. Gebruik patiëntenmateriaal

5.1. Het laboratorium bewaart het verkregen DNA monster van de patiënt voor onbepaalde tijd tenzij een schriftelijk verzoek om het monster te vernietigen is ontvangen van de patiënt of diens wettelijke vertegenwoordigers.

5.2. Het laboratorium gebruikt herleidbaar geanonimiseerd patiënten materiaal voor verder onderzoek (research) in lijn met de oorspronkelijke diagnostische vraagstelling. In geval dit resulteert in voor de patiënt relevante bevindingen zal deze via de oorspronkelijke aanvrager worden geïnformeerd.

5.3. Voor het ontwikkelen van nieuwe en het verbeteren van bestaande technieken gebruikt het laboratorium herleidbaar geanonimiseerd patiëntenmateriaal, o.a. voor controles en validatie. Het laboratorium verzoekt de aanvrager de patiënt hierover te informeren. Mocht deze bezwaar maken tegen het anoniem gebruik van lichaamsmateriaal, dan kan dit op pagina 1 van het aanvraagformulier worden aangegeven.

ATTACHMENT 5

POLYPHEN USER'S GUIDE[26]

POLYPHEN INPUT

PolyPhen works with human proteins and identifies them either by **ID** or **accession number** from *hs_swall* database or by the amino acid sequence itself. In the latter case, PolyPhen tries to find exact match of the sequence in *hs_swall*. If a sequence is identified as a database entry, all entry information (complete sequence, FT, etc.) is used. Amino acid replacement is characterised by **position number** and **substitution**, consisting of two amino acid variants, AA_1 and AA_2.

1. QUERY DATA

The input form contains the following fields:

Protein identifier (ACC or ID) from the SWALL database which is case-insensitive, e.g., pexa_human, XYZ_HUMAN, P12345, p12345, aah01234, etc. PolyPhen maps this value to primary accession number and works with it.

Amino acid sequence in FASTA format which should obey the "classical" FASTA format, e.g., provide sequence identifier

User is supposed to complete only one of the fields above.

Position is checked not to exceed the protein length

Substitution is given by two amino acid variants; the first one is checked to correspond to the actual protein sequence, whereas the second is checked to differ from the first one.

Description is an optional short string (up to 60 characters) providing descriptive name and/or comment for your query. It will be displayed in the query management page to facilitate identifying particular query instances which may be useful when you submit a large number of them.

2. OPTIONS

Structural database (PDB/PQS)

PolyPhen can use two protein structure databases, PDB and PQS. In general, queries against PDB can be faster than those against PQS. However, use of PQS (default) is strongly recommended if a user is concerned with residue contacts, especially inter-subunit.

Sort hits by (Identity/E-value)

Hits are sorted according to the sequence identity or E- value (default) of the sequence alignment with the input protein.

Map to mismatch (No/Yes)

By default, a hit is rejected if its amino acid at the corresponding position differs from the amino acid in the input sequence. Mapping to mismatching amino acid residue should be used with caution only when a protein with known structure and matching amino acid can not be found.

Calculate structural parameters (For first hit only/For all hits)

In some cases a user may want to check the conservation of structural parameters of a residue in all hits. By default, parameters are calculated for the first hit only, since they are expected to be very close in all homologous structures.

Calculate contacts (For first hit only/For all hits)

Contrary to the structural parameters, contacts are by default calculated for all found hits with known structure. This is essential for the cases when several PDB(PQS) entries correspond to one protein, but carry different information about complexes with other macromolecules and ligands (for example, see Fig.2 in [Sunyaev *et al* 2001])

Minimal alignment length (integer number, default: 100)

PolyPhen will filter out hits with structure whose alignment length with the query sequence is smaller than the given value.

Minimal identity in alignment (floating point value, not exceeding 1, default: 0.5)

Hits with structure whose sequence identity to the query sequence is smaller than the given threshold are filtered out

Maximal gap length in alignment (integer number, default: 20)

PolyPhen will filter out hits with structure whose alignment with the query sequence contains gaps with total length greater than this value

Threshold for contacts (floating point value, default: 6.0Å)

PolyPhen will report residue contacts below this threshold

POLYPHEN OUTPUT

PolyPhen output is divided into three main sections and consists mainly of the tables whose contents are discussed below.

1.QUERY

This section contains query data, mostly resembling the input:

Acc number	For entries from *hs_swall* this column contains link to the SRS system.
Position	Substitution position.
AA$_1$	First amino acid variant.
AA$_2$	Second amino acid variant.
Description	For entries from *hs_swall* this column contains protein description from the corresponding database field.

2.PREDICTION

This section contains prediction itself, e.g., "**This variant is predicted to be probably damaging"**, and the supporting information:

Available data	**FT**, **alignment**, **structure** Data available for prediction as <u>described above</u>
Prediction	**benign**, **possibly damaging**, **probably damaging**, **uknonwn**: one of four predictions, <u>also see above</u>
Prediction basis	**sequence annotation**, **sequence prediction**, **multiple alignment**, **structure**: <u>also see above</u>

Substitution effect	For some rules predicting damaging effect, a brief description of expected effect is given. Hierarchy of possible damaging effects is <u>given above</u> . In this column PolyPhen also shows more "friendly" description of effect, e.g., **Hydrophobicity change at buried site** that corresponds to
	 `1.1.1. structural, buried site,hydrophobicity`
	 `disruption`

Prediction data	In case of a damaging substitution, this column summarises (mostly quantitative) data used to make a prediction, e.g., **Normed accessibility: 0.07, Hydrophobicity change: 1.3**
Remarks	Amino acid replacement features that were not used when making prediction, but may nevertheless be interesting, e.g., interchain contacts of a residue.

PREDICTION

The table below contains rules used by PolyPhen to predict effect of nsSNPs on protein function and structure. One row corresponds to one rule which may consist of several parts connected by logical "and". For a given substitution, all

rules are tried one by one, resulting in prediction of functional effect. If no evidence for damaging effect is seen, substitution is considered benign. Prediction basis and Substitution effect are described below.

	RULES (connected with logical AND)			PREDICTION	BASIS	EFFECT
	PSIC score difference:	Substitution site properties:	Substitution type properties:			
1	arbitrary	annotated as a functional site[+]	arbitrary	**probably damaging**	sequence annotation	functional, functional site (2.2)
2	arbitrary	annotated as a bond formation site[++]	arbitrary	**probably damaging**	sequence annotation	structural, bond formation (1.2)
3	arbitrary	in a region annotated as transmembrane	PHAT matrix difference resulting from substitution is negative	**possibly damaging**	sequence annotation	functional, functional site, transmembrane (2.2.2)
4	arbitrary	in a region predicted as transmembrane		**possibly damaging**	sequence prediction	
5	<=0.5	arbitrary	arbitrary	**benign**	multiple alignment	
6	>1.0	atoms are closer than 3Å to atoms of a ligand	arbitrary	**probably damaging**	structure	functional, functional site, ligand binding (2.2.3)
7		atoms are closer than 3Å to atoms of a residue annotated as BINDING,	arbitrary	**probably damaging**	structure	functional, functional site, indirect (2.1)

		ACT_SITE, or SITE				
8		with normed accessibility <=15%	change of accessible surface propensity is >=0.75	**possibly damaging**	structure	structural, buried site, hydrophobicity disruption (1.1.1)
9			change of side chain volume is >=60	**possibly damaging**	structure	structural, buried site, overpacking (1.1.2)
10	in the interval (0.5..1.5]		change of side chain volume is <=-60	**possibly damaging**	structure	structural, buried site, cavity creation (1.1.3)
11		with normed accessibility <=5%	change of accessible surface propensity is >=1.0	**probably damaging**	structure	structural, buried site, hydrophobicity disruption (1.1.1)
12			change of side chain volume is >=80	**probably damaging**	structure	structural, buried site, overpacking (1.1.2)
13			change of side chain volume is <=-80	**probably damaging**	structure	structural, buried site, cavity creation (1.1.3)
14	in the interval (1.5..2.0]		change of accessible surface propensity is >=1.0	**probably damaging**	structure	structural, buried site, hydrophobicity disruption (1.1.1)

15			change of side chain volume is >=80	**probably damaging**	structure	structural, buried site, overpacking (1.1.2)
16			change of side chain volume is <=-80	**probably damaging**	structure	structural, buried site, cavity creation (1.1.3)
17		arbitrary	arbitrary	**possibly damaging**	structure	structural, buried site, cavity creation (1.1.3)
18	>2.0	arbitrary	arbitrary	**probably damaging**	multiple alignment	

[+]BINDING, ACT_SITE, SITE, MOD_RES, LIPID, METAL, SE_CYS

[++]DISULFID, THIOLEST, THIOETH

2.AVAILABLE DATA

PolyPhen makes its predictions using three main source of data:

(1) FT, sequence annotation (or prediction) being a fragment of SWALL feature table (FT) describing the substitution position,

(2) alignment, PSIC profile scores derived from multiple alignment,

(3) structure, structural information, obtained if a search against structural database was successful.

The presence of all three data sources indicates the highest reliability of a prediction. However, as a rough estimate one can expect that approximately only ~10% of all sequences have homologous proteins with known structure.

2.PREDICTION BASIS

As can be seen from the table above, a prediction is based on one of the following:

- **sequence annotation**

- **sequence prediction**
- **multiple alignment**
- **structure**

depending on the rule used to make it.

ATTACHMENT 6

SIFT USER GUIDE[27]

SIFT takes a query sequence and uses multiple alignment information to predict tolerated and deleterious substitutions for every position of the query sequence. SIFT is a multistep procedure that (1) searches for similar sequences, (2) chooses closely related sequences that may share similar function to the query sequence , (3) obtains the alignment of these chosen sequences, and (4) calculates normalized probabilities for all possible substitutions from the alignment. Positions with normalized probabilities less than 0.05 are predicted to be deleterious, those greater than or equal to 0.05 are predicted to be tolerated.

Procedure (the details):

1. **Get related sequences.** A PSI-BLAST search against a database is executed on the query sequence.
   Parameters: 4 iterations, expectation value .0001, e-value threshold for inclusion in multipass model 0.002
   Update 05/15/01 Number of PSI-BLAST iterations reduced to 2 to save time and prevent the search from diverging.

2. **Choose closely related sequences.**

   As described in *Genome Research* 11:963-87: We desire to have sequences that are similar in function as well as structure to the query sequence. To do so, we select only a subset of sequences from the PSI-BLAST results.

   a. Group sequences found from the PSI-BLAST search that are more than 90% identical together and make a consensus sequence for each group by choosing the amino acid that occurs most frequently at each position.
   b. MOTIF finds conserved regions among the query sequence and the consensus sequences from (a) that were derived from at least two sequences.
   c. After the conserved regions in the query sequence have been identified by MOTIF, these regions are extracted from the sequences aligned by PSI-BLAST.
   d. The conserved regions of the query sequence and those consensus sequences more than 90% identical are converted to a PSI-BLAST checkpoint file.
   e. The checkpoint file is given to PSI-BLAST to search among the remaining conserved regions of the consensus sequences not included in the seed checkpoint file. The top hit is added to the alignment corresponding tothe seed checkpoint file and the conservation over the entire alignment of conserved regions is calculated. If conservation does not decrease, the

consensus sequence is added to the alignment and the checkpoint file rebuilt. (e) iterates until conservation decreases.

OR

SIFT by conservation: In the original version of SIFT, an arbitrary number of sequences is added. In this version, sequences are continually added until they reach a sequence conservation cutoff, set by the user. If the sequences for which prediction is based on are very diverse (low conservation cutoff), only substitutions at the strongly conserved positions will be predicted as deleterious. If the sequences chosen for prediction are very similar to each other (high conservation cutoff), then most substitutions will be predicted as deleterious.
Users can choose the degree of sequence conservation: they can opt for detecting most of the deleterious substitutions (use a high sequence conservation) , or predict fewer deleterious substitutions but with a high level of certainty (use a low sequence conservation).

f. Group sequences found from the PSI-BLAST search (step 1) that are more than 90% identical together and make a consensus sequence for each group by choosing the amino acid that occurs most frequently at each position.
g. The query sequence and its checkpoint file is given to PSI-BLAST to search among the consensus sequences. The top hit is added and aligned to the query sequence. Information is calculated for each position in the alignment, and the median of these values is obtained. If the median conservation over all positions does not fall below a given cuttoff, the hit is retained in the alignment and the checkpoint file rebuilt. The process repeats until the median conservations as long as the median information does not fall below the cutoff.

The sequences picked from this iterative procedure are chosen as closely related sequences. You can also submit your own sequences.

3. **Obtain alignment.** Since PSI-BLAST alignments are fairly accurate and long (Sauder & Dunbrack, 2000), we obtain the alignment of the sequences chosen in (2) from the initial PSI-BLAST search results (1).
You can also submit your own alignment of your query sequence with other sequences.

4. **Calculate probabilities.** At each position of the alignment, each amino acid i appears at a frequency n_i. Using the n_i's, the probabilities of amino acids are estimated according to Dirichlet mixtures (d_i's. The final probability of an amino acid appearing at a position, p_i, is a weighted average of the observed

frequencies and the Dirichlet estimation. The weight of the observed frequencies is the number of sequences used to construct the alignment. The weight of the Dirichlet estimated probabilities is an exponential function of a diversity measure (**Div**) calculated by

$$\textbf{Div} = \text{SUM} (\text{rank}_i * n_i)$$

where rank_i is the rank amino acid i has in reference to the original amino acid when BLOSUM62 substitution scores for the original amino acid are ranked from highest to lowest.
Probabilities are normalized by dividing by $\max\{\Pr(\text{amino acid})\}$.

Update: 08/08/01: Prior to calculating the probabilities, sequences > 90% identical to the query sequence are removed. This eliminates the possibility that the sequence containing your substitution of interest is already represented in the database therefore and will trivially be predicted as tolerated.

- We have found by comparison to experimental data that substitutions with less than 0.05 are deleterious. We use this as a cutoff for prediction. We strongly suggest users examine the normalized probabilities manually. If your substitutions are slightly above the 0.05 cutoff, you might want to consider this as a deleterious substitution.