

https://dspace.org/

Investigating the genetic aetiology of heritable connective tissue disorders in South Africa using next generation sequencing

Mhlongo, Felicity Lindokuhle

2021
https://hdl.handle.net/10539/33203

Downloaded from a DSpace repository. https://dspace.org/

INVESTIGATING THE GENETIC AETIOLOGY OF HERITABLE CONNECTIVE TISSUE DISORDERS IN SOUTH AFRICA USING NEXT GENERATION SEQUENCING

FELICITY LINDOKUHLE MHLONGO

A Dissertation Submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine.

Johannesburg, 2021

WITS UNIVERSITY

NATIONAL HEALTH LABORATORY SERVICE

DECLARATION

I Felicity Lindokuhle Mhlongo declare that this Dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science in Medicine in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.



…….. ……13/08/2021……

Felicity Lindokuhle Mhlongo Date

DEDICATION

My hardworking self.

My beautiful, supportive and God-fearing family.

PRESENTATIONS ARISING FROM THIS RESEARCH PROJECT.

“Investigating the genetic aetiology of heritable connective tissue disorders in South Africa using Next Generation Sequencing”, oral presentation at the University of the Witwatersrand, Johannesburg, Faculty of Health Sciences Research day, 15 October 2020.

ABSTRACT

Heritable connective tissue disorders (HCTDs) are a heterogeneous group of disorders affecting the connective tissue. They are caused by pathogenic variants in genes that encode proteins in the extracellular matrix or the post-translational modifications thereof. There is a wide phenotypic spectrum and extensive overlap in the clinical features of the different HCTDs. For these reasons, making a definitive clinical diagnosis is difficult.

Currently in South Africa, the diagnosis of HCTDs is achieved using clinical assessments and available diagnostic criteria. However, these may not always assist because many physical manifestations of HCTDs are age-dependent in their penetrance making diagnosis challenging in young individuals. This study aimed to determine the value of a targeted Next-Generation Sequencing (NGS) panel for known HCTDs genes in the molecular diagnosis of these disorders.

Fifteen (15) patients with a clinical diagnosis of one of seven selected HTCDs were invited to participate in the study. Automated DNA Library, template preparation and sequencing were executed using the Ion Torrent™ system. The Ion Torrent Suite™ and Ion Reporter™ software, as well as the Ensembl Variant Effect Predictor tool were used for variant calling and annotation, respectively.

Six different putative disease-causing variants were identified in 12 family entities, yielding a mutation detection rate of 50%. Two of these variants are novel, contributing to the mutation spectrum of HCTDs. With the use of NGS, we confirmed and established the diagnosis in our cohort, and we recommend its use in a diagnostic setting.

ACKNOWLEDGEMENTS

I would like to acknowledge the following people:

- Participants for giving consent for use of their samples to carry out this research. This research project would not have been possible without their participation.
- Supervisors: Dr Nadia Carstens, Prof Amanda Krause, Dr Candice Feben-Small. Thank you for affording me the opportunity to learn and grow in my academic journey. I am grateful for having worked with all of you for the past two years and have learnt a lot. Dr Carstens was kind enough to read my work even during her maternity leave. Thank you Nadia.
- Funders (National Research Foundation, South Africa).
- My family for their support and prayers when it was not easy.
- My friends for their support always.
- Postgraduate students for their encouragement and support, especially Erin Manolas and Clarice Smal.

Contents

LIST OF ABBREVIATIONS viii
LIST OF FIGURES x
LIST OF TABLES xii
1. INTRODUCTION 1
1.1 Background 1
1.2 Genetic Aetiology of HCTDs 3
1.2.1 *FBN1* gene 3
1.2.3 *FBN2* gene 10
1.2.4 *COL3A1* gene 11
1.2.5 *TFGβR1&2*, *TGFβ2&3* and *SMAD2&3* genes 12
1.2.6 *CBS* gene 16
1.3 Diagnosis of HCTDs 16
1.3.1 Clinical Diagnosis of HCTDs. 16
1.3.2 Molecular Diagnosis of HCTDs 18
1.4 Management of HCTDs 20
1.5. Purpose of the Study 21
1.5.1 Research Question 21
1.5.2 Aim 21
1.5.3 Objectives 21
2. PARTICIPANTS AND METHODS 22
2.1 Ethical Clearance 22
2.2 Subjects 22
2.2.1 Patient Recruitment 22
2.2.2 Inclusion and Exclusion Criteria 22
2.3 Sample Preparation 24
2.3.1 Genomic DNA Extraction 25
2.3.2 NGS-Custom Panel Design 25
2.3.3 DNA Library and Template preparation 27
2.3.4. Sequencing 28
2.4 Data Analysis 29
2.4.1 Post-sequencing Metrics 29
2.4.2 Read Alignment and Variant calling 30
2.4.3. Variant Interpretation 30
2.4.4 Post-alignment Quality Control 36
2.5. Validation of Variants by Sanger Sequencing 36

3. RESULTS 39
3.1 Patient Characterization 39
3.2 NGS Data Quality Control 41
3.3 NGS Variants Identified 43
3.3.1 Putative Disease-Causing Variants 43
3.4 Diagnostic Yield 75
4. DISCUSSION 77
4.1 Project Rationale 77
4.2 Diagnostic Yield 78
4.2.1 HCTDs Diagnostic Yield 78
4.2.2 *FBN1* Diagnostic Yield 78
4.2.3 *FBN1* Mutation-Detection Rate in Present Cohort 80
4.3 Identified Putative Disease-Causing Variants 81
4.3.1 *FBN1* Missense Variants 82
4.3.2 *FBN1* Frameshift Variant 86
4.3.3 *FBN1* In-frame Deletion Variant 87
4.4 Mutation-Negative Cases 88
4.5 Implications of *FBN1* Testing 90
4.5.1 Diagnostic and Predictive Testing 90
4.5.2 *FBN1* Genotype-Phenotype Correlations 91
4.5.3 Genetic Counselling 92
4.5.4 Management 94
4.6 Limitations of the Study 94
4.7 Future Studies 96
4.8 Conclusion 96
REFERENCES 98
APPENDICES 121
Appendix A: Ethics Clearance Certificate 121
Appendix B: Connective Tissue Disorder Clinical Tick sheet 122
Appendix C: Information Sheet and Consent Forms. 124
Appendix D: Informed consent for patients' medical photographs adapted from ACMG 135
Appendix E: ACMG-AMP Guidelines of Variant Interpretation. 136
Appendix F: Sanger Sequencing Protocol 140
Appendix G: Turn-it-in Report 141

LIST OF ABBREVIATIONS

ACMG-AMP: American College of Medical Genomics and Genetics and the Association of Molecular Pathology.

A: Adenine

BAM: Binary Alignment Map

CADD: Combined Annotation Dependent Depletion

CCA: Congenital contractural arachnodactyly

CT: Connective Tissue

CTD: Connective Tissue Disorder

C-Cytosine

DANN: Deleterious Annotation of genetic variants using Neural Networks

dbscSNV: Database for functional prediction of splicing single nucleotide variants

dbSNP: Single Nucleotide Polymorphism Database

ddNTP: Di-deoxyribonucleotide triphosphate

DNA: Deoxyribonucleic acid

dNTP: Deoxyribonucleotide triphosphate

ECM: Extracellular matrix

EDS: Ehlers Danlos syndrome

ELS: Ectopia lentis syndrome

ExAC: Exome Aggregation Consortium

FATHMM: Functional Analysis Through Hidden Markov Models

GERP: Genomic Evolutionary Rate Profiling

gnomAD: The Genome Aggregation Database

GTEx: Genotype-Tissue Expression database

G-Guanine

HCTD: Heritable Connective Tissue Disorder

HGMD: Human Gene Mutation Database

IGV: Integrative Genomics Viewer

ISP: Ion Sphere Particle

LDS: Loeys-Dietz syndrome

LLC: Large latent complex

MAF: Minor Allele Frequency

MFS: Marfan syndrome

NCBI: National Center for Biotechnology Information

NGS: Next-Generation Sequencing

PCR: Polymerase Chain Reaction

PolyPhen2: Polymorphism phenotyping version 2

REVEL: Rare Exome Variant Ensembl Learner

RF: Random forest

SGS: Shprintzen-Goldberg syndrome

SIFT: Sorting Tolerance from Intolerance

SNVs: Single nucleotide variants.

T: Thymine

UMD: Universal Mutation Database

UniProtKB: Universal Protein Resource Knowledgebase

VAF: Variant Allele Fraction

VCF: Variant Call Format

vEDS: vascular Ehlers Danlos syndrome

VEP: Variant Effect Predictor

VUS: Variant of Uncertain Significance

LIST OF FIGURES

CHAPTER ONE

Figure 1.1: The types of connective tissues and their functions……1

Figure 1.2: The *FBN1* gene, located on chromosome 15q21.1 implicated in the causation of Marfan syndrome and ectopia lentis syndrome……4

Figure 1.3: Images showing an individual affected with Marfan syndrome……5

Figure 1.4: The regulation of the transforming growth factor beta pathway by the fibrillin1 protein……14

Figure 1.5: The diagram indicates how the genes implicated in the causation of HCTDs are related to each other……17

CHAPTER TWO

Figure 2.1: A diagram illustrating a summary of the methods……24

Figure 2.2: A diagram demonstrating an overview of the data interpretation workflow to identify putative disease-causing variants……30

Figure 2.3: Flow diagram indicating the filtering pipeline for the identification of variants implicated in the causation of HCTDs……35

CHAPTER THREE

Figure 3.1: Twelve patients' samples were sequenced in the first NGS-sequencing……41

Figure 3.2: Images from the second NGS-sequencing run report……42

Figure 3.3: Pedigree of Family HCTDs_2456 ……45

Figure 3.4: IGV images for variant c.5726T>C in exon 47 of the *FBN1* gene, identified in patients HCTDs_24 (A), HCTDs_25 (B) and HCTDs_26 (C)……49

Figure 3.5: The PCR products run in 2% agarose gel for patients HCTDs_24, HCTDs_ 25, HCTDs_26 and HCTDs_23……50

Figure 3.6: The sequence electropherograms for the *FBN1* heterozygous missense variant c.5726T>C in patients HCTDs_24, HCTDs_25 and HCTDs_26……51

Figure 3.7: IGV image indicating a change in nucleotide; c.3794G>A, in *FBN1*……54

Figure 3.8: This electropherogram depicts the c.3794G>A heterozygous missense variant……54

Figure 3.9: IGV image showing the c.640G>A variant in exon 7 of the *FBN1* gene……57

Figure 3.10: The PCR products ran in 2% agarose gel electrophoresis……58

Figure 3.11: The heterozygous missense variant c.640G>A in the *FBN1* gene……58

Figure 3.12: An image from Mutalyzer 2.0.32 depicting a shortened protein……60

Figure 3.13: An IGV image depicting an insertion the A base in exon 55 of the *FBN1*……61

Figure 3.14: An electropherogram depicting the insertion of the base T……62

Figure 3.15: This IGV image is indicative of the nucleotide change C>T……65

Figure 3.16: An electropherogram showing the heterozygous *FBN1* c.3037G>A missense variant……65

Figure 3.17: The electropherogram depicts the c.3037G>A heterozygous missense *FBN1* variant in the reverse strand……66

Figure 3.18: The pedigree for HCTDs_13 and HCTDs_14 with a suspected diagnosis of MFS with an identified likely pathogenic variant……67

Figure 3.19: A screenshot from Mutalyzer 2.0.32 showing the reference protein……69

Figure 3.20: These IGV images depict the coverage of the *FBN1* heterozygous in-frame deletion; c.23_34del……70

CHAPTER FOUR

Figure 4.1: A diagram showing the distribution of the *FBN1* likely pathogenic or pathogenic variants……81

LIST OF TABLES

CHAPTER ONE

Table 1.1: Inherited disorders of the connective tissues and some of their discriminating features……3

Table 1.2: Revised Ghent criteria for Marfan syndrome……7

Table 1.3: Major and minor diagnostic criteria for vascular Ehlers Danlos syndrome……11

Table 1.4: Clinical features of Loeys-Dietz syndrome in different systems……13

CHAPTER TWO

Table 2.1: The sizes and biologically relevant transcripts of the genes described as putative disease-causing in HCTDs……27

Table 2.2: The transcripts used in variant prioritization to identify putative pathogenic variants……33

Table 2.3: The pathogenicity predicting *in silico* tools used in the study……34

Table 2.4: Sanger Sequencing Primers for validation of putative disease-causing variants……37

Table 2.5: Optimized primer-specific PCR conditions for each variant……38

Table 2.6: Cycle sequencing thermal cycler conditions……38

CHAPTER THREE

Table 3.1: Patients' clinical features in different systems suggesting inherited disorders of connective tissue……40

Table 3.2: The mean depth of coverage observed for all patients in both NGS sequencing runs and the number of variants identified in each patient……43

Table 3.3: The coverage and allele fraction of the (likely) pathogenic variants identified in this study……44

Table 3.4: The three shared variants for further investigation in family HCTDs_2456……47

Table 3.5: Table showing the shortlisted variant in patient HCTDs_23……52

Table 3.6: The shortlisted variants for patient HCTDs_22……56

Table 3.7: The shortlisted variant in the *FBN1* gene identified patient HCTDs_21……59

Table 3.8: The three variants shortlisted in patient HCTDs_29……64

Table 3.9: The identified variant following the family- segregation variant filtering strategy……68

Table 3.10: Shortlisted variants identified in HCTDs_32 post variant filtering……71

Table 3.11: The candidate variant identified in the *COL3A1* gene of patient HCTDs_31……72

Table 3.12: Table showing the identified candidate variant in patient HCTDs_28……73

Table 3.13: The one variant identified in HCTDs_27 ……73

Table 3.14: The identified variant in patient HCTDs_30 detected in the *COL1A1* gene……74

Table 3.15: Table with the (likely) pathogenic variants in the *FBN1* gene identified in our study cohort……75

CHAPTER FOUR

Table 4.1: A table showing mutation screening *FBN1*- yields in different population groups ……80

1. INTRODUCTION

1.1 Background

The human body is composed of four types of biological tissues: nervous, epithelial, muscular and connective tissues (CTs). CTs have varied cell types and tissue architecture. They are the most abundant and widely distributed tissues. The characteristic components of CTs are cells, fibrous proteins (collagen and elastin) and a ground substance. Collagen is a long and straight fibrous protein with great tensile strength which resists stretching. Elastin is a protein that allows stretching and coiling. These fibrous proteins together, with the ground substance form the extracellular matrix (ECM) (Kusindarta and Wihadmadyatami, 2018).

CTs are classified into three types, based on the characteristic of the ground substance and the fibre within the ECM including CT proper, supporting CT and fluid CT. The examples of the types of connective tissues are adipose tissue, tendon, bone, cartilage, blood and lymph. For instance, blood has a liquid matrix. The different CTs have various functions. Their functions (as indicated in Figure1.1) are to join organs and tissues together to provide structure, framework, protection, transport substances in the body and also store surplus energy as fat (McKee *et al.,* 2019).

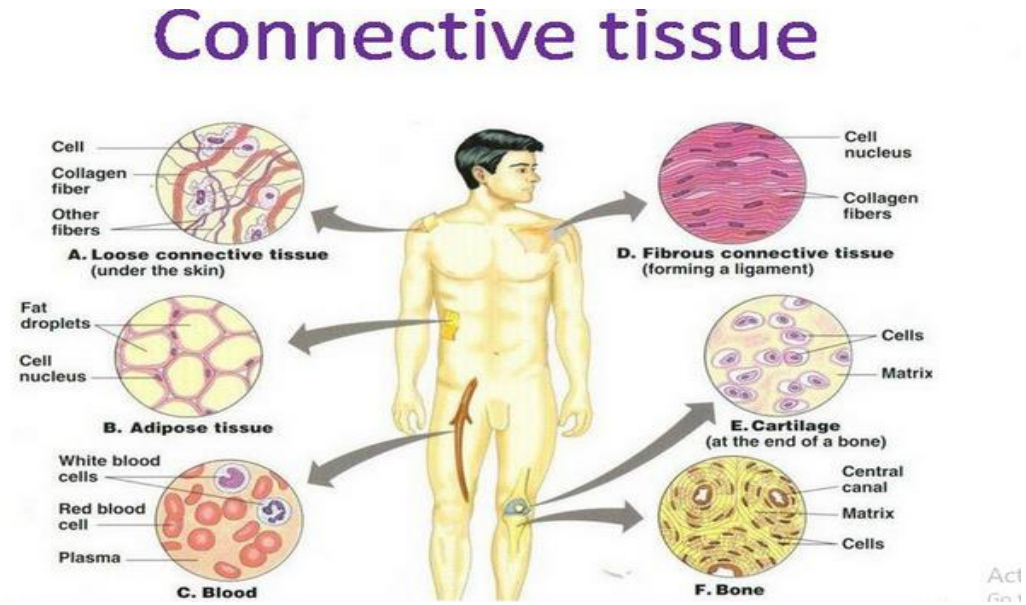


Figure 1.1. Types of connective tissues (CT) and their functions. **A.** The loose CT under the skin holds organs in place **B.** Adipose tissue stores energy as fat. **C.** Blood transports substances in the body **D.** Ligaments attach bones to bones **E.** Cartilage are joints between bones **F.** Bones provide structure (Karki, 2018).

A defect in any of the components of CT could result in a connective tissue disorder (CTD). CTDs are disorders characterised by an abnormal element of CT and can either be acquired or inherited. Heritable connective tissue disorders (HCTDs) are a group of heterogeneous, phenotypically related disorders with the CT as the site of pathology that are caused by single gene pathogenic variants in genes encoding the fibrous proteins (Grahame and KazKaz, 2014) or the enzymes involved in the modification of these proteins after translation (De Paepe and Malfait, 2012). Tissue laxity and fragility are characteristic features of HCTDs and impair the tissue's tensile strength (Grahame and KazKaz, 2014). Inherited disorders of connective tissue may be associated with developmental delay. Developmental delay is a condition that affects major domains such as motor, speech, cognitive and social skills (Olusanya *et al.,* 2018).

There are various inherited CTDs which affect a wide diversity of organs; some of the disorders include Marfan syndrome (MFS), ectopia lentis syndrome (ELS), Shprintzen-Goldberg syndrome (SGS), vascular Ehlers-Danlos syndrome (vEDS), Loeys-Dietz syndrome (LDS), congenital contractural arachnodactyly (CCA), also known as Beals syndrome and homocystinuria.

HCTDs are one of the most difficult groups of disorders to diagnose (Baranov *et al.,* 2018). This is due to the wide dispersion of CTs in the human body and their presence in almost all body organs, leading to the wide clinical phenotype (Baranov *et al.,* 2018) and the extensive overlap in features between the different disorders (Grahame and KazKaz, 2014). Patients often present with clinical features suggestive of a possible HCTD, however physicians may be unable to classify it as a specific disorder. About 50% of HCTDs at disease onset are not classified as a specific HCTD (Mosca *et al.,* 2014). There are however some distinguishing features, shown in Table 1.1, which, if present, can be helpful.

Table 1.1. Inherited disorders of the connective tissues and some of their discriminating features (modified from Loeys *et al.,* 2010).

HCTD	**Gene(s) implicated**	**Discriminating feature(s)**
Marfan syndrome	*FBN1*	Tall stature, long and thin bones.
Ectopia Lentis syndrome	*FBN1* and *ADAMTSL4*	Dislocated lens with no aortic dilation
Shprintzen-Goldberg syndrome	*SKI*	Craniosynostosis, hypotonia
Beals syndrome	*FBN2*	Crumpled ears, arachnodactyly
Vascular Ehlers-Danlos syndrome	*COL3A1* and *COL1A1*	Uterine rupture, aortic aneurysms
Loeys-Dietz syndrome	*TGFβ2&3, TGFBR1&2* and *SMAD2&3*	Hypertelorism, osteoarthritis, cleft palate
Homocystinuria	*CBS*	Thromboembolism

ADAMTSL4- A disintegrin and metalloproteinase with thrombospondin motif, *CBS*- cystathionine- beta synthase, *COL1A1*-type I Collagen, *COL3A1* type III Collagen, *FBN1*-Fibrillin1, *FBN2*- Fibrilin-2, HCTDs-heritable connective tissue disorders, *SMAD2*-Mothers against decapentaplegic homolog2, *SMAD3*- Mothers against decapentaplegic homolog3, *SKI*- Sloan-Kettering Institute, *TGFβ2*- Transforming-growth factor beta2, *TGFβ3*-Transforming-growth factor beta3, *TGFβR1*- Transforming-growth factor beta receptor1, *TGFβR2*- Transforming-growth factor beta receptor 2.

1.2 Genetic Aetiology of HCTDs

There are several genes reported to be responsible for causing HCTDs when they are mutated (Loeys *et al.*, 2010). These include the fibrilin1 *(FBN1)*, sloan-kettering institute *(SKI)*, fibrilin-2 *(FBN2)*, type III collagen *(COL3A1)*, type I collagen *(COL1A1)*, transforming-growth factor beta receptor *1* and 2 *(TGFβR1* and *TGFBβR2)*, transforming-growth factor beta 2 and 3 *(TGFβ2* and *TGFβ3)*, mothers against decapentaplegic homolog 2 and 3 *(SMAD2* and *SMAD3)*, cystathionine- beta synthase *(CBS)*, and the a disintegrin and metalloproteinase with thrombospondin motif *(ADAMTSL4)* genes, as shown in Table 1.1.

1.2.1 *FBN1* gene

1.2.1.1 Gene Structure

The *FBN1* gene contains 65 exons and is located on chromosome 15q21.1 (Figure 1.2) (Chandra and Charteris, 2014). It encodes fibrillin-1, a cysteine-rich extracellular matrix protein which contributes to the formation of microfibrils, which are constituents of elastic fibres (Dietz, 2017). The fibrilin-1 protein is composed of 47 epidermal growth factor (EGF) and 7 TGFβ binding protein-like domains. Of the EGF domains, 43 are calcium-binding (cb-EGF) and play a role in the assembly and stability of microfibrils (Takeda *et al.,* 2018). There are also 2 hybrid domains, and these are composed of

both cb-EGF and the TGFβ like-domain properties (Baudhuin *et al.,* 2019). The fibrillin1 protein maintains elastic mechanical stability, controls elasticity (Dietz, 2017) and provides structural force-bearing support to many tissues and organs (Tunçbilek and Alanay, 2006).

Pathogenic variants in the *FBN1* gene result in either a severe reduction in microfibril number (Dietz, 2017) or a disruption of the microfibrillar architecture (Chandra and Charteris, 2014). More than 3000 different mutations have been identified in the *FBN1* gene (reviewed by Takeda *et al.,* 2018), with no clear genotype-phenotype correlation. The mutations reported are nonsense, missense, splice-site variants, indels and copy number variants (CNVs) (Stengl *et al.,* 2020). Mutations in the *FBN1* gene cause MFS, ELS, mitral valve prolapse (MVP) and MASS phenotype (myopia, mitral valve prolapse, mild aortic dilatation, striae and skeletal involvement) (Rippe *et al.,* 2016).

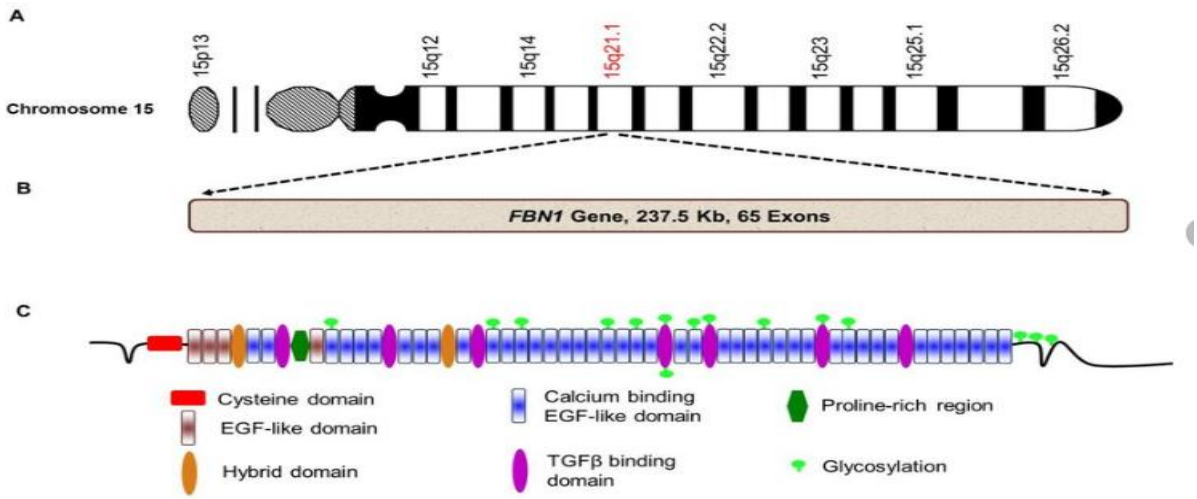


Figure 1.2. The *FBN1* gene, located on chromosome 15q21.1 implicated in the causation of Marfan syndrome and ectopia lentis syndrome (Matkar *et al*. 2015).

1.2.1.2 Marfan Syndrome

Marfan syndrome (MFS) is a systemic disorder that affects the CT with a prevalence of 1 in 5000- 1 in 10000 (Dietz, 2017). In South Africa, this figure is extrapolated to estimate that there will be 10000 individuals affected with MFS in the country (Child *et al.,* 2007). MFS occurs in equal frequencies in both sexes, in all countries and races (Von Kodolitsch, 2015) and it is inherited in an autosomal dominant manner.

MFS is associated with a wide continuum of clinical features. Clinical variability is observed even in the same family (Hoffjan, 2012). The affected systems include the cardiac, ocular, and skeletal systems. The most common clinical presentations include aortic dilatation, mitral or tricuspid valve prolapse, ectopia lentis, scoliosis and pectus deformity (Dietz, 2017). Defining clinical features of MFS include disproportionate growth of the long bones, as demonstrated in Figure 1.3 (Judge and Dietz, 2005). Morbidity and mortality are primarily related to abnormalities in the cardiovascular system (Verstraeten *et al.,* 2016).



Figure 1.3. Images adapted from Zhurayev *et al*. (2016) showing: **A**. An individual affected with Marfan syndrome with disproportionate overgrowth of the long bones. **B.** Arachnodactyly, a condition where the toes and fingers are slender and long, which is characteristic of Marfan syndrome.

Neonatal MFS is a severe form of MFS characterized by earlier onset of symptoms and rapid disease progression (Dietz, 2017). Patients with neonatal MFS are diagnosed early and have aortic complications within their first year of life (Liu *et al.,* 2018). Reported cases of neonatal MFS occur in 1 to 6 of 10000 individuals (Kitahara *et al*, 2016); hence, neonatal MFS is a rare disorder according to the European definition of rare disorders.

Phenotypic overlap exists between MFS and other HCTDs, making diagnosis difficult. In a proband or an individual with a family history of MFS, definitive diagnosis can be made based on the revised Ghent diagnostic criteria as shown in Table 1.2 (Loeys *et al.,* 2010). However, because many physical manifestations of MFS are age-dependent, the diagnostic criteria may not always assist because they can clearly establish a diagnosis in individuals 20 years or older and are not specific to younger individuals. It is also unclear whether the Ghent criteria are applicable to the South African population (Child *et al.,* 2007). According to Dietz (2017), an individual <20 years with aortic dilatation (Z- score ≤3), systemic score (≤7 points), without a mutation in the *FBN1* gene can be diagnosed with a “non-specific CTD”. An individual <20 years can be classified as having “potential MFS” if they have a mutation in the *FBN1* gene that has not been reported to be associated with aortic dilatation, and the aortic root Z-score is <3 (Dietz, 2017).

Cysteine residues play a crucial role in the structure of the fibrilin1 protein. Pathogenic variants in the *FBN1* gene that alter cysteine residues have been identified and described as the most common cause of MFS (Baudhuin *et al.,* 2019). Variants occurring in functional domains and highly conserved regions of the *FBN1* gene are most likely to be classified as pathogenic. This was best shown by Baudhuin *et al*. (2019) who evaluated the classification of *FBN1* variants reported in ClinVar (Landrum *et al.,* 2014) and deduced that variants in these regions were most likely to be classified as disease-causing.

Table 1.2. Revised Ghent criteria for Marfan syndrome (Loeys *et al.,* 2010).

Absence of family history	**Presence of family history**
Aortic dilation (≥2, Z-score) with a pathogenic variant in the *FBN1* gene	Ectopia lentis
Aortic dilation (≥2, Z-score) with ectopia lentis	Aortic dilation (≥2 and ≥3 Z-score, in individuals ≥20 years and ≤20 years, respectively)
Aortic dilation (≥2, Z-score) with a systemic score (≥7, points)	Systemic score (≥7 points)
Ectopia lentis with a pathogenic variant in the *FBN1* gene associated with aortic dilatation	

*The ≥2 and ≥3 z-scores indicate 2 and 3 standard deviations greater than the mean (when standardized to age and body size), respectively (Loeys *et al.,* 2010). The scoring of systemic features includes giving patient points for the manifestation of certain features in different organ systems, where a score of ≥7 points is indicative of positive systemic involvement (Loeys *et al.,* 2010).

1.2.1.3 Ectopia Lentis

Autosomal dominant ELS is a HCTD with the ocular system as the site of pathology and is caused by mutations in the *FBN1* gene. The prevalence of congenital ELS is 6 in 100 000 (Chandra and Charteris, 2014). Ectopia lentis may indicate systemic involvement of MFS, or it can be isolated (Neuhann *et al.,* 2015). A characteristic feature of ELS is the displacement of lens from their normal position in the globe and the defining feature is the absence of aortic dilatation (Loeys *et al.,* 2010).

ELS mutations in the *FBN1* gene lead to an abnormal distribution and configuration of the microfibrils at the site of the cilliary zonules, which are crucial structures of the lens (Chandra and Charteris, 2014). Fibrillin-1 is an important macromolecule of the cilliary zonules; this explains the presence of ectopia lentis in 60% of individuals with MFS (Chandra and Charteris, 2014). As many features of MFS show age-dependent penetrance, it may be difficult to exclude MFS in an individual with ectopia lentis in the absence of genetic testing. According to Loeys *et al*. (2010), diagnosis of autosomal dominant ELS is achieved when an individual presents with isolated ectopia lentis, with a pathogenic variant in the *FBN1* gene that has not been previously associated with aortic root dilatation. The most common pathogenic variants in the *FBN1* gene that are causative in ELS are missense mutations that alter cysteine residues, or mutations in the first 15 exons at the 5’ end of the gene (Chandra and Charteris, 2014).

ELS is not always caused by pathogenic variants in the *FBN1* gene. Autosomal recessive ELS is caused by biallelic mutations in the *ADAMTSL4* gene, located on chromosome 1q21.2 (Chandra and Charteris, 2014). The product of the *ADAMTSL4* gene is found in the eye (Gabriel *et al.,* 2012), which indicates a role in ocular development. Mutations in the *ADAMTSL4* gene are the most common cause of ELS (Neuhann *et al.,* 2015) and are associated with a severe phenotype, characterised by an earlier onset of ectopia lentis. This means that individuals with *ADAMTSL4*-related ELS are usually identified during early childhood.

1.2.1.4 Genetic *FBN1* Testing

FBN1-related diagnostic testing is challenging, specifically with regard to the interpretation of the generated sequencing variants (Groth *et al.,* 2016). The *FBN1* gene is among the largest in the human genome, constituting over 200 000 bases (Summers *et al.,* 2012). Pathogenic variants have been described along the length of the entire gene (Sakai *et al.,* 2016) and only approximately 10% of these are recurrent in MFS affected families (Gong *et al.,* 2019). Most *FBN1* gene mutations causing MFS are unique to a family. In addition, the expressivity of MFS even in a family with the same mutation is variable (Martinez-Quintana *et al.,* 2014). This may be because of the pleiotropic nature of the *FBN1* gene mutations or because the variant may not be the only factor determining the severity of the phenotype (Takeda *et al.,* 2018).

For these reasons, *FBN1* genetic testing is difficult, time-consuming and expensive (Summers *et al.,* 2006). Sanger sequencing alone is inappropriate due to the extensive coding regions of the gene (Aleem, 2019). On the contrary, high-throughput sequencing technologies such as the Next-generation Sequencing (NGS) approach allow for more extensive genetic testing than traditional Sanger sequencing (Weerakkody *et al.,* 2016). Developed countries such as Australia, China and Switzerland perform *FBN1* genetic testing by using a targeted NGS approach in the diagnosis of MFS and other HCTDs (Summer *et al.,* 2012, Gong *et al.,* 2019, Bombardiers *et al.,* 2020). In this regard, *FBN1* genetic testing using NGS is the best possible technique to cover the entire length of the gene with low costs and short run-time (Ma, Wang and Wei, 2016).

1.2.2 *SKI* gene

Shprintzen Goldberg syndrome (SGS) is a rare autosomal dominant disorder of CT. It is characterised by craniosynostosis and features in the cardiovascular, skeletal, craniofacial and central nervous systems (Greally, 2013). SGS shares several features with MFS and LDS (Schepers *et al.,* 2015). A clinical diagnosis of SGS is suspected when an individual has a combination of craniosynostosis, hernias, scoliosis, and aortic dilation (Greally, 2013). Due to the extensive clinical overlap between SGS and MFS, these HCTDs were initially thought to be caused by different mutations in the same gene.

Although mutations in the *FBN1* gene had initially been implicated in SGS, Greally (2013) reported that the *SKI* gene is the only gene with pathogenic variants causative of SGS. The *SKI* gene is localized on chromosome 1p36.33-p36.32 and it encodes a protein that represses the SMAD-dependent TGF-β signalling pathway by preventing the phosphorylation of SMAD2 and SMAD3, thereby preventing its translocation to the nucleus (Greally, 2013).

Mutations in the *SKI* gene lead to elevated levels of SMAD2, SMAD3 and TGF-β due to loss of suppression of the TGFβ signalling. The increase in TGFβ signalling is implicated in disorders with clinical presentations of syndromic aortic aneurysms such as SGS, MFS and LDS (Doyle *et al.,* 2012), however aortic dilatation is less frequent in SGS patients. With SGS having defining features such as intellectual disability and hypotonia, the SKI protein has also been assumed to play a crucial role in neuronal and muscle cell development (Schepers *et al.,* 2015). Transgenic mice with homozygous mutations in the *SKI* gene have severe hypotonia and defects in the closure of the cranial neural tube (Schepers *et al.,* 2015). All pathogenic variants that have been identified in the *SKI* gene thus far are in the first exon, in the SMAD-binding domain and Dachshund-homology domain (DHD) (Takeda *et al.,* 2018) and they are either missense or deletion mutations (Zhang *et al.,* 2019).

1.2.3 *FBN2* gene

Beals syndrome, also known as CCA is an autosomal dominant HCTD characterized by multiple contractures, underdevelopment of the muscles and “marfanoid” features such as scoliosis and arachnodactyly (Tunçbilek and Alanay, 2006). It is caused by mutations in the *FBN2* gene, situated on chromosome 5q23. The gene encodes a protein that promotes the assembly of elastic fibers in early embryogenesis (Guo *et al.,* 2016). Pathogenic variants in the *FBN2* gene alter the structure of the protein or decrease the protein production (Guo *et al.,* 2016) and these include nonsense, missense, indels (either point or gross) and splice site mutations (Deng *et al.,* 2016). There are 162 different *FBN2* variants reported in the Human Gene Mutation Database (HGMD) (Stenson *et al.,* 2003). Most pathogenic variants are missense mutations that substitute cysteine with another amino acid (Guo *et al.,* 2016). This suggests that alterations in the protein structure are the underlying pathogenic mechanism of CCA.

CCA shares several features with MFS. This makes it difficult to estimate the prevalence of CCA. The defining features of CCA are contractures, crumpled ears and the absence of aortic dilatation (Tunçbilek and Alanay, 2006). There is phenotypic overlap between neonatal MFS and CCA, and in such cases early onset of aortic dilatation or the presence of kyphoscoliosis can be indicative of neonatal MFS or CCA, respectively (Tunçbilek and Alanay, 2006). The pronounced overlap between patients with CCA and neonatal MFS is because mutations in the *FBN2* gene that cause CCA cluster in exons 24-32, known as the “neonatal region”, a region similar to where mutations in the *FBN1* gene cluster to cause neonatal MFS.

In addition, both causal genes of CCA and neonatal MFS are highly homologous at the nucleotide and protein level (Frédéric *et al.,* 2009). They are 84% homologous upon comparison of 58% of their sequence (Frédéric *et al.,* 2009), and the amino-acid homology is 61–69% (Aubart *et al.,* 2016). The domain organization and the sequence motifs of the fibrillin1 and fibrillin2 proteins are similar (Robinson and Godfrey, 2000). They are however expressed differently. The *FBN1* gene is expressed from development until adulthood. On the contrary, the *FBN2* gene is expressed primarily during early development (Aubart *et al.,* 2016).

1.2.4 *COL3A1* gene

The Ehlers-Danlos syndromes (EDS) are a heterogeneous group of monogenic disorders affecting multiple systems (De Paepe and Malfait, 2012) with complications involving the skin, musculoskeletal, and cardiovascular systems. They are classified into six groups with variable clinical phenotypes. The vascular form of EDS (vEDS) is one of the major subtypes of EDS. vEDS is an autosomal dominant condition characterized principally by arterial dilatation, aneurysm formation and vessel rupture (Byers *et al.,* 2017). It affects the cardiovascular, gastrointestinal, pulmonary, ocular and cutaneous systems. Clinical features suggestive of vEDS include easy bruising, thin, translucent skin and joint hypermobility (Malfait *et al.,* 2017). The 2017 International Clinical Classification of Ehlers-Danlos syndrome identified both minor and major diagnostic criteria, as shown in Table 1.3. A clinical diagnosis of vEDS can be established if a person has a family history of vEDS, an arterial rupture at <40 years, sigmoid colon perforation and spontaneous pneumothorax (Malfait *et al.,* 2017).

Table 1.3. Major and minor diagnostic criteria for vascular Ehlers-Danlos syndrome (Malfait *et al.,* 2017).

Major criteria	**Minor criteria**
1. Family history of vEDS with a described causative variant in the *COL3A1* gene.	1. Easy bruising
2. Arterial rupture	2. Thin, translucent skin with increased venous visibility
3. Spontaneous sigmoid colon perforation in the absence of known diverticular disease or other bowel pathology	3. Spontaneous pneumothorax
4. Uterine rupture during the 3rd trimester of pregnancy	4. Acrogeria, talipes equinovarus, congenital hip dislocation
5. Carotid-cavernous sinus fistula (CCSF) formation in the absence of trauma	5. Hypermobility of small joints, tendon and muscle rupture
	6. Keratoconus, gingival recession, early onset varicose veins

The prevalence of vEDS is 1 in 50 000 to 1 in 200 000. It is caused by mutations in genes encoding fibrillar collagen or genes encoding enzymes involved in the post-translation modification of the proteins (De Paepe and Malfait, 2012). The implicated genes involved in the causation of vEDS are *COL3A1* and *COL1A1*, localized on chromosome 2q32.3 and 17q21.33, respectively (Malfait *et al.,* 2017). The encoded collagen proteins are mainly found in blood vessels, the skin and hollow organs and

are responsible for forming a microfibrillar network for strength and structure of the ECM (De Paepe and Malfait, 2012). The characteristic structure of the proteins is a trimeric molecule with three polypeptide chains, with each of the polypeptides having the glycine-proline-hydroxyproline (Gly-Xaa-Yaa) subunit, with the Glycine required for correct formation of the helix (Vanakker *et al.,* 2015).

The nature of the mutation has an influence on the severity of vEDS. For example, most mutations in the *COL3A1* gene resulting in the substitution of glycine in the helical region of the collagen molecule (De Paepe and Malfait, 2012) are associated with a less severe phenotype (Byer *et al.,* 2017), while pathogenic variants that alter sequences in the carboxyl-terminal propeptide of the chains and splice site mutations lead to skipping of exons and are both associated with a full range phenotype or severe clinical features (Byers *et al.,* 2017). Pathogenic variants in the *COL1A1* gene are implicated in vEDS (Malfait *et al*. (2017), but rarely. Most *COL1A1* pathogenic variants cause osteogenesis imperfecta (Steiner and Basel, 2019). However, Malfait *et al*. (2007) identified rare pathogenic variants in the *COL1A1* gene that cause vEDS. These were all missense mutations substituting cytosine with thymine at different codons leading to a change in the amino acids from arginine to cysteine.

1.2.5 *TFGβR1&2*, *TGFβ2&3* and *SMAD2&3* genes

LDS is an autosomal dominant CT disorder that affects the vascular, skeletal, craniofacial and cutaneous systems (Loeys and Dietz, 2018). Its prevalence is unknown and manifestations include arterial aneurysms, uterine rupture, osteoarthritis, clubfoot, widely spaced eyes and thin, translucent skin (Loeys and Dietz, 2018). The clinical features are shown in Table 1.4. LDS has a number of overlapping features with MFS (Loeys *et al.,* 2010). This could be explained by the hypothesis that the fibrillin1 protein regulates the bioavailability of TGFβ (Hoffjan, 2012), therefore, mutations in both the *FBN1* gene and those in the TGFβ pathway can produce similar clinical features, and this is indicated in Figure 1.4. Despite the clinical overlap between LDS and MFS, LDS is associated with early onset of aortic dilatation (Hoffjan, 2012). There are also pregnancy-related complications in individuals affected with LDS, including uterine rupture. An individual can be diagnosed with LDS if they have a pathogenic variant in any of the genes involved in

the Transforming-growth factor β (TGF-β) pathway and aortic dilation (with a Z-score >2) or a combination of features in the systems LDS affects.

Table 1.4. Clinical features of Loeys-Dietz syndrome in different systems (Loeys and Dietz, 2018).

System	**Clinical presentation**
Cardiovascular	Arterial aneurysms, dissections and rupture, tortuosity.
Skeletal	Arachnodactyly, osteoarthritis, skeletal overgrowth, pectus deformity, congenital clubfoot, spine anomalies, joint hypermobility.
Craniofacial	Widely spaced eyes, craniosynostosis, cleft palate, malar flattening.
Cutaneous	Thin, translucent, velvety skin. Easy bruising, dystrophic scarring.
Ocular	Blue sclerae, retinal detachment.
Other	Asthma, food allergy, eosinophilic gastrointestinal disease. Spleen, bowel and uterine rupture.

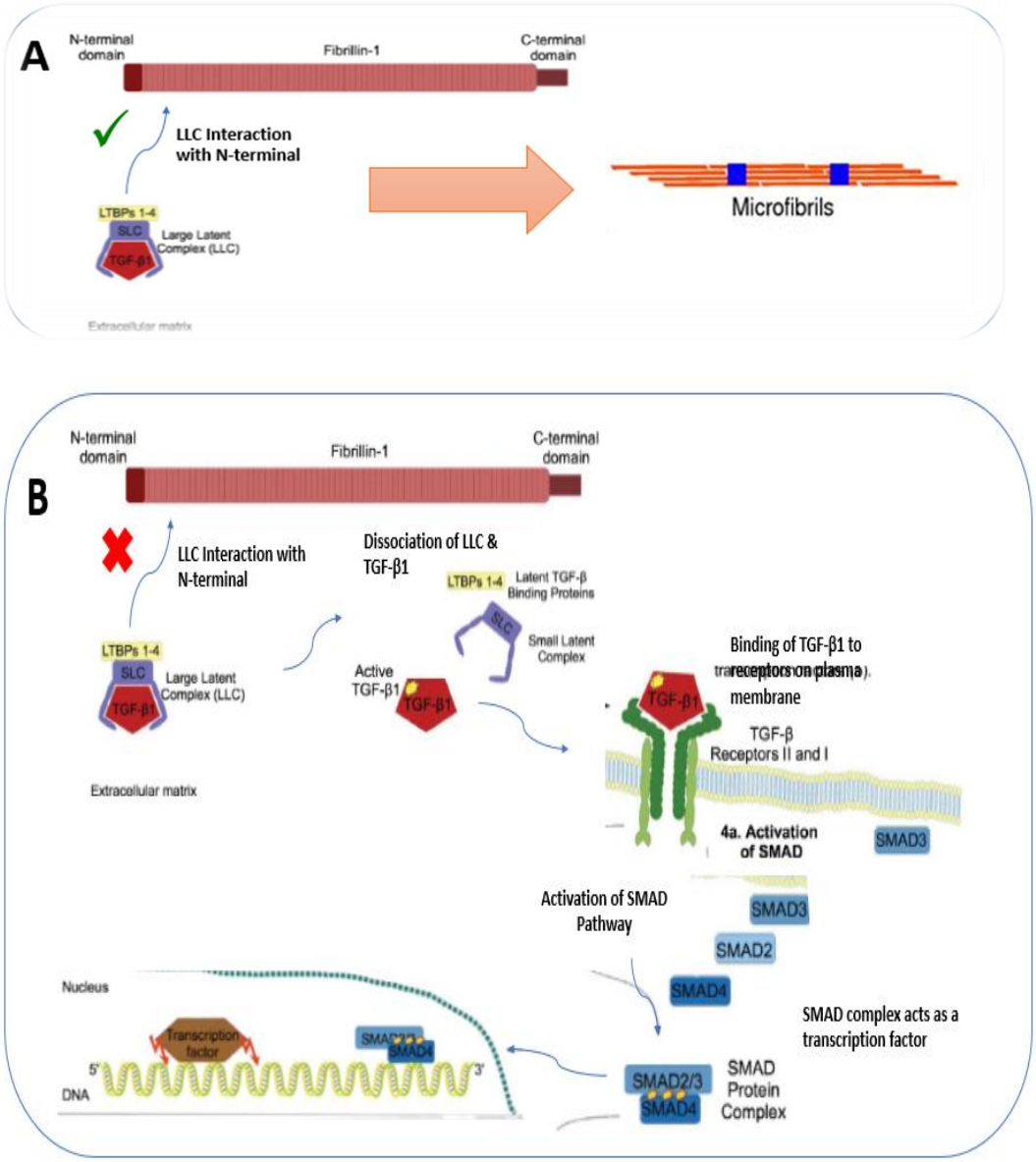


Figure 1.4. The regulation of the transforming growth factor beta pathway by the fibrillin1 protein. **Part A.** Inactive TGFβ is secreted by the cells as part of the large latent complex (LLC). The LLC binds to the N-terminal region of the fibrillin1 protein. This then forms microfibrils. **Part B**. Mutation of the *FBN1* gene causes the encoded protein structure to be altered. Binding of the LLC to fibrillin1 causes dissociation of the complex, thus the release and activation of TGFβ. TGFβ will then bind to *TGFβRII*, which leads to phosphorylation of *TGFβR*I. Phosphorylation of *TGFβRI* recruits the SMAD complex which then moves to the nucleus where it controls gene expression (Adapted from Sepetiene and Simonyte, 2016).

Pathogenic variants in any of the genes in the TGF-β pathway cause LDS (Loeys and Dietz, 2018). The *TGFβ2* and *TGFβ3* genes (on chromosomes 1q41 and 14q24.3, respectively) encode cytokines that regulate the behaviour of cells; including cell migration and proliferation, by binding TGFβRs. The *TGFβR1* and *TGFβR2* genes (on chromosomes 9q22.33 and 3p24.q, respectively) encode serine/threonine receptors that bind TGFβ cytokines. The *SMAD2* and *SMAD3* genes (on chromosomes 18q21.1 and 15q22.33, respectively) encode proteins that serve as signal transducers and transcriptional modulators and they are activated by TGF-β. The pathogenic variants identified in LDS- related genes include insertions, deletions, missense, nonsense and splice site variants (Schepers *et al.,* 2018). Most of the variants implicated in the causation of LDS are missense changes that affect the serine-threonine kinase domains of the TGFβ receptors (Hoffjan, 2012).

Due to the overlap in clinical features, it is often a challenge to clinically identify which of the LDS-related genes is likely to be involved in a particular affected individual and there is very little genotype-phenotype correlation (Loeys and Dietz, 2018). Clinical variability between and within family members with LDS is observed (Loeys and Dietz, 2018).

In a study by Campbell *et al*. (2011), a female patient had a deletion in the *TGFβR2* gene and presented with microcephaly and developmental delay, while another individual with a duplication mutation in the *TGFβR1* presented with features suggestive of LDS (Hoffjan, 2012). These findings are consistent with the assumption that elevated TGFβ signalling, not haploinsufficiency, is the main pathologic mechanism of LDS. This suggests that pathogenic variants in the genes that are in the TGFβ pathway and involved in the causation of LDS are gain-of-function (Gordon and Blobe, 2008, Takeda *et al.,* 2018). The early onset of aortic dilatation and aneurysm is thus as a result of increased signalling in the aortic wall (Gordon and Blobe, 2008, Hara *et al.,* 2019).

1.2.6 *CBS* gene

The *CBS* gene (located on chromosome 21q22.3) encodes an enzyme that converts homocysteine into cysteine (Sacharow *et al.,* 2017). Pathogenic variants in the *CBS* gene cause homocystinuria, a metabolic disorder characterized by a congenital error of metabolism where there is accumulation of toxic homocysteine. Homocystinuria is an autosomal recessive condition that affects multiple systems such as the ocular, skeletal, vascular and central nervous systems (Sellos-Moura *et al.,* 2020). The clinical manifestations include ectopia lentis, scoliosis, pectus deformity, seizures, psychiatric problems, thromboembolism and intellectual disability, with these features expressed variably in each patient (Sacharow *et al.,* 2017). The causes of morbidity and mortality in homocystinuria relate to thromboembolism. The most common pathogenic variants of the *CBS* gene causing homocystinuria involve the change in thymine to cytosine, guanine to adenine and cytosine to thymine at different nucleotide positions (Sacharow *et al.,* 2017). The prevalence of homocystinuria is 1 in 200 000 to 1 in 335 000 (Sellos-Moura *et al.,* 2020).

1.3 Diagnosis of HCTDs

1.3.1 Clinical Diagnosis of HCTDs.

HCTDs are difficult to diagnose (as illustrated in Figure 1.5.) due to the large number of differential diagnoses and the pronounced overlap in clinical features between the disorders. Additional factors include the heterogeneity of mutations, absence of genotype-phenotype correlations and the age-dependent physical manifestations (Loeys *et al.,* 2010). These may be attributed to the related nature of the genes implicated in HCTDs by either having a role in the same signalling pathway or encoding proteins with similar functions.

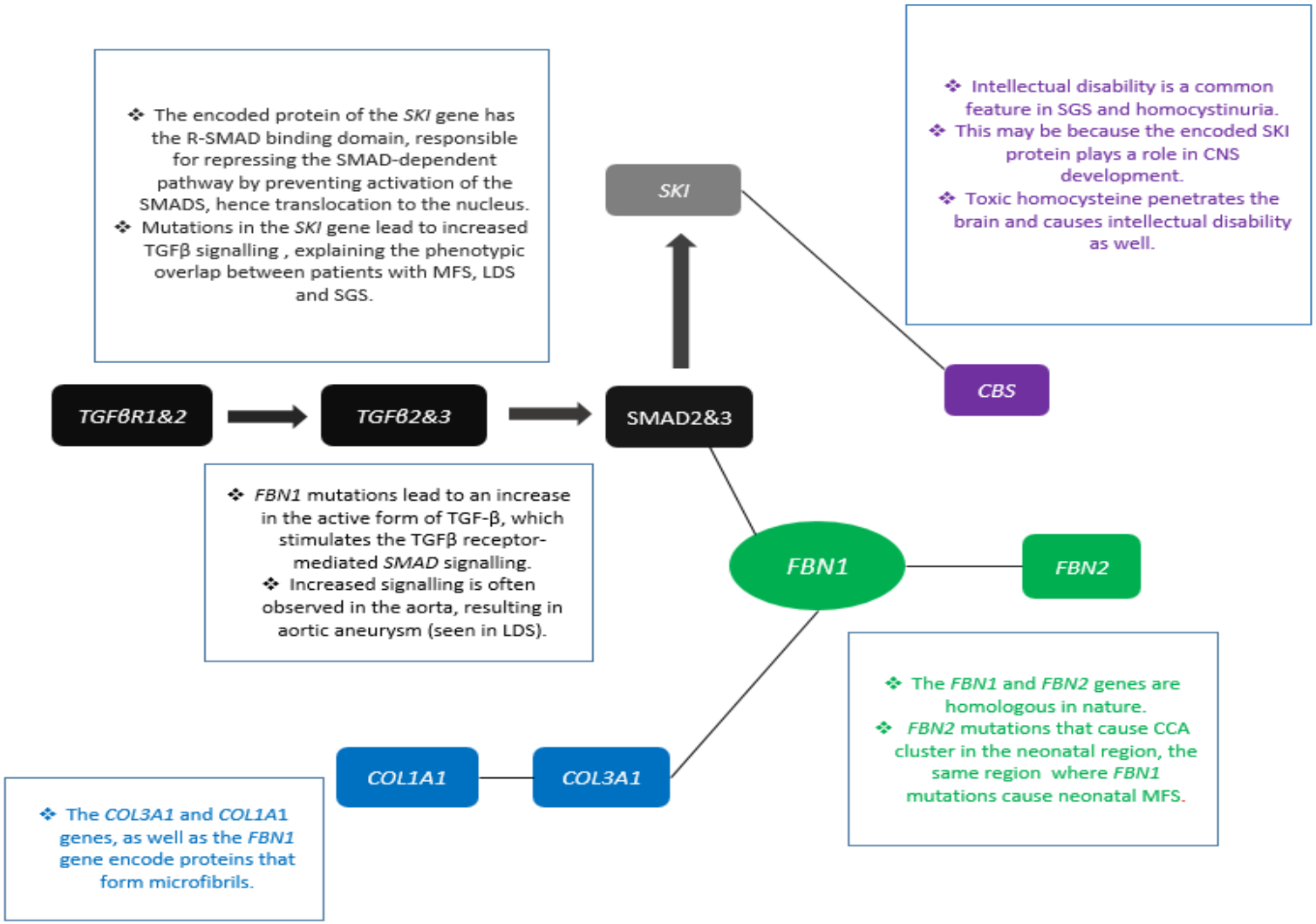


Figure 1.5. The diagram above indicates how the genes implicated in the causation of heritable connective tissue disorders are related to each other to cause an overlap in clinical features. The arrows indicate genes is the TGFβ pathway (*TGFβR1&2, TGFβ1&2* and *SMAD2&3)* and their co-repressor (*SKI*), while the lines illustrate the related nature of the genes with regards to their encoded protein function.

In South Africa currently, diagnosis of a HCTD is based on the clinical phenotype of the patient. In some cases, application of clinical scoring systems or diagnostic criteria aid in making a diagnosis. However, there is no certainty that the widely used diagnostic criteria for HCTDs are applicable to the South African population (Child *et al.,* 2007). Thus, accurate diagnosis of any HCTD should include the knowledge of the wide spectrum of differential diagnoses and their genetic background (Hoffjan, 2012).

1.3.2 Molecular Diagnosis of HCTDs

Genetic testing to confirm the clinical diagnosis is not available to the majority of South African patients at this stage, and testing for HCTDs is not offered to State patients. This may be due to economic restrictions (Child *et al.,* 2007) given that only 16% of the South African population belong to medical aid schemes while the remaining 84% rely on the State Healthcare system (Naidoo *et al.,* 2012). Additionally, there is unequal distribution of resources particularly because the 16% of South Africans with medical insurance consume over 50% of the whole health expenditure while the remaining 84% rely on the overburdened public sector (National Department of Health, 2014). Private patients have access to comprehensive genetic screening through local and international labs, while state patients have access to the limited genetics tests offered only by the National Health Laboratory Service (NHLS). These genetic services for state patients are provided through academic hospitals and these are not available in rural areas (Kromberg *et al*., 2013). In this regard, it makes it difficult for patients to access genetic services due to reasons that may include poor public transport and travelling long distances to cities where the genetic facilities are situated and the genetic services are covered by the state (Kromberg *et al*., 2013).

The identification of pathogenic variants aids in confirmation of the clinical diagnosis, streamlines clinical care, and provides information on prognosis and possible complications (Child *et al.,* 2007). Additionally, affected individuals and their family members may benefit from pre-symptomatic testing and pre-natal diagnosis. Genetic counselling as well as developing optimal personalised medicine can also be achieved with the identification of pathogenic variants. Precision medicine is the integration of primarily genetic variation, the environment and lifestyle with medicine to implement person-specific treatment (Parish and Ashley, 2017). As some genetic variants are

directly associated with the severity of HCTDs, knowledge of these can facilitate care choices. These factors suggest that molecular diagnostic testing of HCTDs could be extremely useful in South Africa.

NGS is a sequencing technique that allows simultaneous sequencing of small DNA fragments. It is able to sequence multiple different genes or gene regions and heterogeneous DNA samples in a single test (Yohe and Thyagarajan, 2017). NGS is a useful tool when studying heterogeneous diseases like HCTDs, where multiple genes are involved in the causation and a large phenotypic spectrum exists (Di Resta *et al.,* 2018). This makes it possible to investigate many of the putative disease-causing genes at the same time using NGS. As management and risk profiles of HCTDs are different (Loeys *et al.,* 2010), a molecular diagnosis is of great significance. This can be achieved either by the use of single-gene testing, a multi-gene panel including the most common genes reported to be causative of HCTDs, by whole exome sequencing (WES) or whole genome sequencing (WGS) (Dietz, 2017).

Due to the genetic heterogeneity of HCTDs, novel genes may be implicated in HCTDs and are yet to be identified. Thus, WES and WGS NGS are optimal sequencing approaches in variant detection. They are comprehensive sequencing methods, however, they are frequently used as secondary sequencing techniques in mutation negative cases (Hartman *et al.,* 2019, Yohe and Thyagarajan, 2017). This may be because despite the comprehensive nature of WES and WGS, these approaches are expensive and are associated with long turnaround times, which may be partially related to the interpretation of a large number of generated sequencing variants (Saudi Mendeliome Group, 2015, Miller *et al.,* 2017). A NGS targeted sequencing approach using a multi-gene panel is thus an efficient alternative in a South African setting.

A multi-gene panel is a genetic test where simultaneous targeted mutation analysis of several genes associated with similar clinical phenotypes is possible. Commercial laboratories including the University of Nebraska Medical Center (Human Genetics Laboratory, Omaha, Unites States of America), test for HCTDs using an NGS sequencing gene panel comprising over 67 genes (Vanderjagt and Butler, 2019). Amsterdam, in the Netherlands routinely uses an NGS sequencing gene panel incorporating 21 genes to test for HCTDs and uses a secondary technique such as

multiplex ligation-dependent probe amplification (MLPA) to detect copy number variations (CNVs) (Overwater *et al.,* 2018).

Technologies such as NGS can aid in understanding the molecular basis of HCTDs and countries such as Russia already use a NGS panel, covering 50 genes for the molecular diagnosis of HCTDs (Baranov *et al.,* 2018). While, as from 2013, the University of Oslo, Norway has used a HCTD-NGS panel for the diagnosis of HCTDs (Pope *et al.,* 2019). This indicates that other countries already used targeted NGS as a clinical service to provide molecular diagnosis of HCTDs. NGS also allows differentiation of disorders with clinical overlap, while it is rapid and cost effective (Wooderchak-Donahue *et al.,* 2015), thus making the technology best suited for the molecular diagnosis of HCTDs. In this regard, a NGS gene panel would be the best approach in the molecular diagnosis of HCTDs in South Africa.

1.4 Management of HCTDs

For streamlined management of multisystem disorders, a multi-disciplinary approach is required (Ojha and Prasad, 2016). For this reason, efficient management of the HCTDs requires a team of medical specialists (Dietz, 2017, Loeys and Dietz 2018, Byers, 2019). This includes clinical geneticists, ophthalmologists, cardiologists and orthopaedic surgeons, for example. MFS requires great health care expenditure (Achelrod *et al*., 2014), this may be due to the systemic nature of the disorder. Surveillance of MFS patients therefore involves annual ophthalmologic and echocardiography examination. Frequent echocardiography examinations are required when an individual has an aortic dilatation that exceeds 4.5cm in adults, for an example (Dietz, 2017). An echocardiogram is an expensive test (Pyeritz, 2012) and repeated echocardiograms are impractical in a South African setting where the majority of our patients do not have access to specialist care. The benefits of a timely diagnosis of HCTDs, particularly MFS include the possibility that early treatment will prevent serious complications of the disease (Stengl *et al.,* 2020). In South Africa, this may result in the decline of the use of screening approaches such as echocardiograms, which cost about R1 500 and would have to be repeated regularly for the diagnosis of MFS (Child *et al.,* 2007).

1.5. Purpose of the Study

1.5.1 Research Question

The present study intended to study the genetics of HCTDs in South African patients by designing a HCTD-NGS gene panel that incorporated genes determined to be causative of HCTDs and to use it to provide diagnostic services to South African patients presenting with clinical features suggestive of one of the HCTDs investigated in this study.

The main question is whether an NGS panel based on known HCTDs-causal genes would improve the diagnostic yield in patients with a clinical or suspected diagnosis of an HCTD in South Africa.

1.5.2 Aim

The aim of this study was to determine the applicability of a next generation sequencing panel approach to improving the molecular diagnosis of heritable connective tissue disorders in South African patients.

1.5.3 Objectives

The aim of the present project was met by addressing the following objectives:

1. To recruit and sequence the DNA of patients with a clinical diagnosis of either MFS, ELS, SGS, CCA, vEDS, LDS or homocystinuria.
2. To design an NGS panel with genes known and proposed to be implicated in the causation of HCTDs.
3. To analyze and classify the variants obtained on sequencing, validate the results obtained on the NGS platform using Sanger sequencing and assess family segregation, where possible.

2. PARTICIPANTS AND METHODS

2.1 Ethical Clearance

This research project has been reviewed and approved by the University of the Witwatersrand, Human Research Ethics Committee (HREC) (Protocol: M191184). See Appendix A for ethics clearance certificate.

2.2 Subjects

2.2.1 Patient Recruitment

Twelve families with a clinical diagnosis of either MFS, ELS, SGS, CCA, vEDS, LDS or homocystinuria were recruited to the study by medical geneticists from the Division of Human Genetics, NHLS) and the University of Witwatersrand (Wits). These were families who attended a Genetics Clinic hosted by the Division of Human Genetics and met the inclusion criteria for the study.

We faced some complications due to the Covid-19 outbreak in South Africa with regards to patient recruitment and had to augment our recruitment strategy to only invite existing patients from the Genetics Clinic to participate in the study, rather than recruiting patients from a broader base. We therefore restricted patient recruitment to participants with a suspected or clinical diagnosis of a HCTD who were patients that had formerly attended a Genetics Clinic and were seen by medical geneticists for follow up visits in the Division of Human Genetics, NHLS, and Wits University.

2.2.2 Inclusion and Exclusion Criteria

Clinical assessment of the participants by medical geneticists was performed as part of routine care and clinical diagnostic criteria were applied where appropriate (Owlia, 2006). These included the revised Ghent criteria for MFS (Loeys *et al.,* 2010) and the revised Villefranche criteria for vEDS (Malfait *et al.,* 2017). Patients were invited to participate in the study if they had a clinical suspicion or diagnosis of one of the HCTDs investigated in this study. A clinical tick sheet was used (Appendix B), to allow standardised data collection. Participants with a confirmed diagnosis of a different CT disorder from the HCTDs investigated in the present study, and those who had

confirmed chromosome abnormalities were excluded from the study. There were no exclusions based on age, ethnicity or gender.

Upon inviting patients to participate in the study, an information sheet defining the purpose and objectives of the study was provided to the participants or legal guardians. Subsequent to obtaining informed consent, and assent where appropriate, from participants or legal guardians, blood samples were drawn. Additional consent documentation for medical photographs was also acquired from participants or legal guardians for permission to use their photographs. The informed consent form used to obtain consent for patients' medical photographs was adapted from the American College of Medical Genetics (ACMG) (Cunniff *et al.,* 2000).

2.3 Sample Preparation

The figure below (Figure 2.1) summarises the methods and reagents employed in this study in preparation of the samples and data generation.

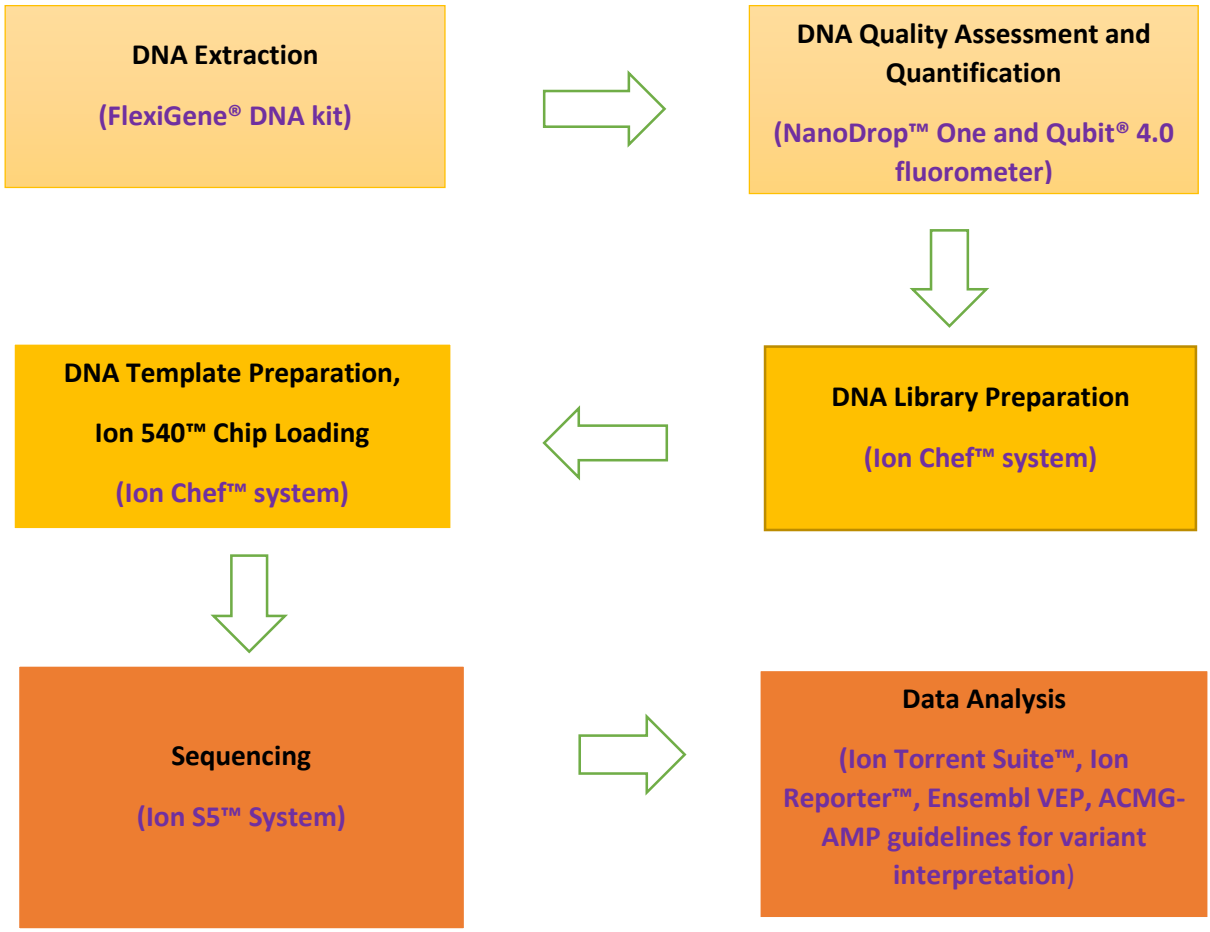


Figure 2.1. Summary of the methods.

DNA was extracted using the FlexiGene® DNA kit and the quality and quantity of the DNA were assessed using the NanoDrop™ One and Qubit® 4.0 fluorometer. Automated DNA library, template preparation and chip loading were carried out using the Ion Chef™ system, subsequently, sequencing was performed by the Ion S5™ System. The Ion™ reporter, Ion Torrent Suite™ softwares and the Ensembl variant effect predictor tool were used for variant calling and variant annotation, respectively. The identified variants were classified according to the American College of Medical Genetics and Genomic and the Association for Molecular Pathology (ACMG-AMP) guidelines for variant interpretation.

2.3.1 Genomic DNA Extraction

DNA was extracted from 3ml whole blood using the FlexiGene® DNA kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions. The extracted DNA was then assessed for quantity and quality using the NanoDrop™ One spectrophotometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). This system employs the use of A260/280 and the A260/230 ratios as DNA purity measures (Desjardins and Conklin, 2010). Pure nucleic acid yields A260/280 and A260/230 ratios of ~1.8 and 1.8 – 2.2, respectively and low values are indicative of contamination. A Qubit® 4.0 fluorometer (Invitrogen™, Waltham, Massachusetts, USA) was also used to assess the quantity of double stranded DNA in each sample prior to NGS library preparation.

2.3.2 NGS-Custom Panel Design

The identification of putative genes involved in the causation of HCTDs was achieved through a detailed literature search. The Genetic Testing Registry (GTR®) database, accessible through the National Center for Biotechnology Information (NCBI), (https://www.ncbi.nlm.nih.gov/gtr/all/tests/?term=190181[mim]) was also investigated. This was with the aim of identifying other genes implicated in HCTDs from HCTDs panels offered by commercial companies and academic laboratories that are registered for molecular diagnostic testing through genomic technology and offer NGS panels for HCTDs. Additional genes from these institutes were added on the Ion AmpliSeq™ On-Demand panel (Thermo Fisher Scientific, Waltham, Massachusetts, USA) that was designed. These include the *MYH11, ACTA2* and *TNXB* genes. The reason for the choice of these genes is because they have been proposed to be implicated in either syndromic aortic aneurysm (AA) or the musculoskeletal manifestations of MFS (Pinard *et al.,* 2019). In the African population, the genetic aetiology of HCTDs is not thoroughly investigated, hence, the inclusion of the additional genes was an attempt to look for novel gene variation that may be implicated in the African population.

A gene panel was designed using the Ion AmpliSeq™ designer tool software (Thermo Fisher Scientific, Waltham, Massachusetts, USA) (https://ampliseq.com/protected/startPage.action). The design of the NGS custom panel was achieved by inputting the genes of interest into the Ion AmpliSeq™ designer search tool. Only the genes available on the Ion AmpliSeq™ On-Demand catalogue were included on the panel for the reason that they are pre-validated. This however meant that we had to exclude the *TNXB* gene, implicated in EDS as it was not yet available on the AmpliSeq On-demand catalogue.

Mendelian disorders are predominantly caused by pathogenic variants in protein-coding regions (Ombrello *et al.,* 2014), and we therefore opted to sequence the exonic sequences of our target genes. Splicing is the biological process where pre-mRNA is processed into mature RNA (Anna and Monika, 2018). This is possible through the recognition of *cis* intronic and exonic elements, also called consensus splice site signals and splicing regulatory sequences, respectively. Mutations affecting the *cis* elements necessary for splicing may change the standard splicing pattern, resulting in genetic disorders (Cartegni *et al.,* 2002). Five base-pairs (5bps) of the exon-intron boundaries were therefore also sequenced to screen for variants that may disrupt these consensus splice sites.

Due to the Covid-19 pandemic in South Africa in 2020, the designed HCTDs panel could not be delivered on time. As a result, the Inherited Disease Project (IDP) Ion AmpliSeq™ On-Demand panel (Thermo Fisher Scientific, Waltham, Massachusetts, USA), comprising 485 genes was used to sequence the samples. This panel was designed and validated in-house to target genes implicated in a variety of Mendelian disorders including HCTDs. It has been established for possible use as the primary basis for molecular diagnostic testing of inherited diseases at the NHLS, Wits University, Division of Human Genetics. Selective analysis was then performed on the 13 genes (Table 2.1) proposed and described as strong candidates for HCTDs.

Table 2.1. The sizes and biologically relevant transcripts of the genes identified as known and putative disease-causing genes in HCTDs.

Gene Name	Associated HCTD(s)	Size (kb)	Number of Exons	Transcript Length (bps)
ACTA2	Skeletal features of MFS	17.70	9 (8 coding)	1756
ADAMTSL4*	ELS	11.56	20 (18 coding)	4319
CBS	Homocystinuria	23.19	17 (15 coding)	2583
COL1A1	vEDS	18.34	51 (all coding)	6727
COL3A1	vEDS	38.43	51 (all coding)	5543
FBN1	MFS, ELS	237.54	66 (65 coding)	11756
FBN2	CCA	280.34	66 (65 coding)	10724
MYH11	Syndromic AA	71	43 (41 coding)	6886
SKI	SGS	81.89	7 (all coding)	5613
SMAD2	LDS	121.70	11 (10 coding)	34626
SMAD3	LDS	129.57	9 (all coding)	6464
TGFβ2*	LDS	99.28	7 (all coding)	5868
TGFβ3	LDS	22.99	7 (all coding)	2522
TGFβR1	LDS	49.48	9 (all coding)	6492
TGFβR2	LDS	87.63	7 (all coding)	4621

*The genes marked in red are implicated in ELS and LDS, respectively, however they were not on the IDP panel, and therefore were not tested. They would have been included on the custom panel.

2.3.3 DNA Library and Template preparation

Prior to DNA library preparation, each sample was diluted to approximately 1ng/µl. Automated DNA library preparation was performed using the Ion AmpliSeq™ Chef DL8 kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the Ion AmpliSeq™ Library preparation protocol in the Ion Chef™ System user guide (Revision F.0) (Thermo Fisher Scientific, Waltham, Massachusetts, USA).

The Ion Chef™ system (Thermo Fisher Scientific, Waltham, Massachusetts, USA) allows DNA library preparation of 8 samples per run, and it achieves this by using polymerase chain reaction (PCR) to amplify each barcoded sample with each primer pool. The resulting DNA pooled libraries were then normalised to a final concentration of approximately 100pM. This was achieved by the Ion Library Equalizer™ technology which is built into the Ion Chef™ system (Thermo Fisher Scientific, Waltham, Massachusetts, USA). To sequence 16 samples on the Ion 540™ chip, the DNA libraries from two 8-sample Ion Chef™ preparations were pooled together. The pooled DNA libraries were then diluted to 50pM before proceeding to template generation.

Automated template preparation was performed on the Ion Chef™ system (Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the Ion 540™ Kit Chef User guide (Revision E.0) (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The Ion Chef™ system (Thermo Fisher Scientific, Waltham, Massachusetts, USA) employs the emulsion PCR amplification technique to prepare DNA templates for sequencing. Template preparation is the clonal amplification of each amplicon on an Ion Sphere Particle (ISP) using emulsion PCR, thereby resulting in ISPs with multiple copies of the same amplicon. The primary reason for performing template preparation is to make multiple copies of the same amplicon in close proximity to ensure that during sequencing, a large enough signal is created, and thus detected.

The Ion Chef™ system then executed chip loading. This is a process of depositing ISPs into the wells of the Ion 540™ chip (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The bottom of the Ion 540™ chip wells has sensory layers for the detection of pH change during sequencing. The loaded Ion 540™ chip was then ready for sequencing.

2.3.4. Sequencing

An initialization step was performed before sequencing using the Ion S5™ Sequencing Kit (Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to the Ion 540 Kit™ Chef user guide (Thermo Fisher Scientific, Waltham, Massachusetts, USA). This step was to prepare the run reagents and ensure that they were all in the correct pH range. Sequencing of the loaded Ion 540™ chip was then performed on the Ion S5™ Sequencer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The sequencing technique used was that which allowed sequencing of 200 base-read DNA libraries. The Ion Torrent™ sequencing technology (Thermo Fisher Scientific, Waltham, Massachusetts, USA) employs the semiconductor sequencing method. This is a variation of sequencing by synthesis method, allowing building of a complementary growing DNA strand from a leading template strand. This is done through the incorporation of a nucleotide by DNA polymerase to the growing DNA strand.

To achieve the sequencing reaction, the chip was flooded with components of a PCR. These include a single species of deoxyribonucleotides triphosphates (dNTPs; dATP, dCTP, dGTP, dTTP), buffers and DNA polymerase. This is because this sequencing method is based on the detection of H+ ions that are released when there is an incorporation of a nucleotide by DNA polymerase, thereby causing a change in pH and creating a positive voltage, hence calling of a base (Merriman *et al.,*2012). A total of 550 sequencing flows were used for each sequencing chip.

The flow of a trial dNTP solution complementary to the template DNA strand leads to base incorporation in the growing DNA strand, thereby causing the release of H+, resulting in a change in pH. The pH change was then detected by a built in pH meter and it assists in determining if a base was added and how many thereof to the sequence read. Subsequent to flowing each trial nucleotide solution, a wash flow is performed to ensure removal of the preceding trial base (Merriman *et al.,* 2012). The flow of a trial dNTP solution that is not complementary to the leading DNA strand means that there will not be an incorporation of a nucleotide, therefore there will be no release of H+ and no pH change, and as a result, no base will be called. In the event of the template DNA strand having homopolymer reads, flowing of a complementary single species dNTP will allow incorporation of multiple dNTP molecules at a single cycle. This results in the release of more H+ ions and the voltage will be doubled. The chip then records multiple bases.

2.4 Data Analysis

2.4.1 Post-sequencing Metrics

The Ion S5™ (Thermo Fisher Scientific, Waltham, Massachusetts, USA) sequencing run report produced by the Ion Torrent Suite™ software was used to assess the quality of the sequencing data. This was done by reviewing the pre-alignment and alignment metrics. The acceptable threshold for chip loading, polyclonality, usable reads and low quality reads in the sequencing run were >70%, <35%, >55% and <20%, respectively (de Abreu *et al.,* 2016). The average sequence depth of coverage of 30X for each targeted region is considered optimum to precisely call a base in germline testing (Jennings *et al.,* 2017). Hence, bases with a sequencing coverage of >30X were presumed to be reliable.

2.4.2 Read Alignment and Variant calling

Alignment of the reads and variant calling were performed by pre-installed plug-ins in the Ion Torrent Suite™ software 5.14 (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The sequence reads were aligned to the human reference genome, Genome Reference Consortium Human Build 37 (GRCh37/hg19), generating the binary alignment map (BAM) files. Variants were called by the VariantCaller plug-in and stored in variant call format (VCF). The BAM and VCF files contain alignment information and genetic sequence variations between the reference genome and that of the patient, respectively (Steward *et al.*, 2017).

2.4.3. Variant Interpretation

The summary of the workflow employed in the interpretation of the variants using various tools is described in Figure 2.2.

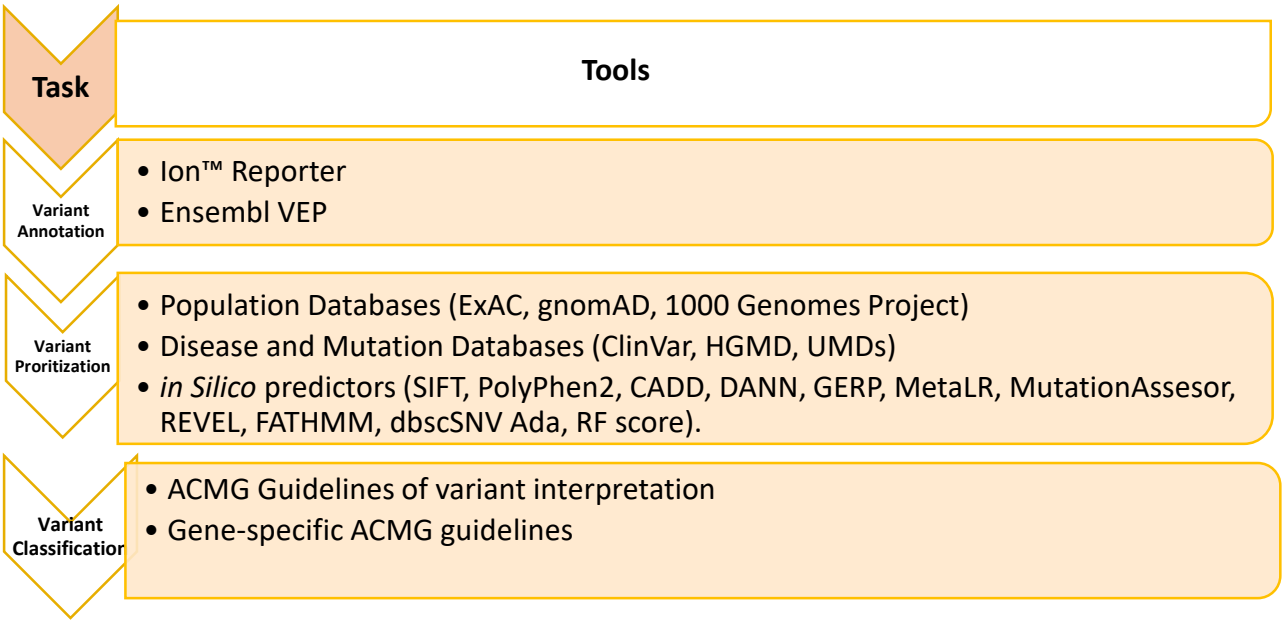


Figure 2.2. A diagram demonstrating an overview of the data interpretation workflow to identify putative disease-causing variants.

2.4.3.1 Variant Annotation

Annotating a DNA variant is the process of giving functional meaning to it (McCarthy *et al.,* 2014). All variants in VCF format for all patients were annotated using the Ion Reporter™ Software 5.14.1.0 (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and Ensembl variant effect predictor (VEP) based on the human reference sequence assembly GRCh37/hg19.

Ensembl VEP is an online tool that offers an automated efficient approach to variant annotation for variants in coding and non-coding regions based on different transcripts (McLaren *et al.,* 2016). It gives information on the class of variant, the affected gene symbol and the location of the variant (Sefid and Gamieldien, 2017). It then investigates various population, variant, disease and mutation databases to determine the frequency for each variant, functional consequence and clinical significance of the variants and these are used as filters when prioritizing variants. It also uses preinstalled plug-ins to predict the effect of the variant on proteins, splicing and regulatory regions. For variants that have been previously described and cited in articles, it also gives PubMed identifiers.

2.4.3.2 Transcript Choice

Genes have alternative transcripts and express many simultaneously. The choice of transcript set is significant when annotating and prioritizing variants (McCarthy *et al.,* 2014). It is for these reasons that the literature and the Genotype-Tissue Expression (GTEx) database (GTEx Consortium, 2020) were assessed to determine alternative biologically relevant transcripts. GTEx provides the different transcripts for each gene, the tissues where they are expressed as well as their expression levels.

In assessing GTEx, the genes; *FBN1*, *COL3A1, COL1A1, ACTA2, TGFβR1, TGFβR2, CBS*, *SMAD2, SMAD3, SKI* and *MYH11* had multiple alternative transcripts, including the main isoform expressed in either the aorta, cultured fibroblasts, the uterus and other relevant tissues. For a standardised approach to variant prioritisation, we restricted our analysis to canonical transcripts to prioritise variants in genes with multiple isoforms expressed in appropriate tissues (Table 2.2). Sequence variants in *FBN2* and *TGFβ3, were* prioritized on the basis of biologically relevant isoforms; *FBN2_201* and *TGFβ3_001,* respectively.

2.4.3.3 Variant Prioritization

An integrated assessment method was used for prioritizing the variants. This included assessing the data from population, disease and mutation databases, as well as using pathogenicity predicting computational tools (Figure 2.3). This information was obtained through Ensembl VEP. In investigating the variants for pathogenicity, we employed a knowledge-driven approach based on the suspected diagnosis, as described by Sefid and Gamieldien (2017). This was done by first evaluating the sequence variants affecting the gene(s) known to have a role in the pathogenesis of the suspected HCTD per patient (MacArthur *et al.,* 2014). Where gene prioritisation did not aid in the identification of putative disease-causing variants, we expanded the analysis to include all of the 13 target genes for this study. The same was done in cases where the patient had an unspecified HCTD diagnosis.

The population databases; 1000 Genomes Project (http://browser.1000genomes.org) (The 1000 Genomes Project Consortium, 2015), Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/) (Karczewski *et al.,* 2017) and genome Aggregation Database (gnomAD) (http://gnomad.broadinstitute.org/) (Karczewski *et al.,*2017) were evaluated to determine how frequently the variants had been observed previously in population data. Rare germline variants, with a minor allele frequency (MAF) of less than 1% (<0.01) (MacArthur *et al.,* 2014), reported on the basis of canonical or biologically relevant transcripts (Table 2.2), were prioritised for further analysis using the ACMG-AMP criteria of variant interpretation (Richards *et al.,* 2015).

The disease and mutation databases; ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) (Landrum *et al.,* 2018), human gene mutation database (HGMD) (http://www.hgmd.org) (Stenson *et al.,* 2003) and universal mutation databases (UMD) (Collod-Béroud *et al.,* 2003 and Frédéric *et al.,* 2009) were then investigated. This was to determine whether or not the variants we observed were associated with the inherited disorders of the CTs under study and to identify the variants observed in this study that had been previously described in multiple unrelated individuals and classified as pathogenic. Variants reported as benign in reputable databases such as these were filtered out.

Table 2.2. The transcripts used in variant prioritization to identify putative pathogenic variants.

Gene Symbol	Location	Canonical Transcript
ACTA2	10q22	NM_001141945.2
CBS	21q22	NM_000071.3
COL1A1	17q21	NM_000088.4
COL3A1	2q32	NM_000090.3
FBN1	15q21	NM_000138.4
FBN2	5q23	NM_001999.4
MYH11	16p13	NM_001040114.1
SKI	1p36	NM_003036.4
SMAD2	18q21	NM_001003652.4
SMAD3	15q22	NM_ 005902.4
TGFβ3	14q24	NM_003239
TGFβR1	9q22	NM_004612.4
TGFβR2	3p24	NM_ 001024847

The transcripts marked in blue are representative of non-canonical transcripts.

Multiple *in silico* predictive tools that were incorporated into Ensembl VEP (such as Sorting Intolerant From Tolerant (SIFT), Polymorphism Phenotyping version 2 (PolyPhen2), Combined Annotation Dependent Depletion (CADD), Deleterious annotation of Genetic Variants using Neural Network (DANN), MetaLR, MutationAssessor, MutationTaster, Rare Exome Variant Ensemble Learner (REVEL), Functional Analysis Through Hidden Markov Models (FATHMM), Genomic Evolutionary Rate Profiling (GERP) and dbscSNV AdaBoost (Ada) and random forest (RF)) were used in combination for prediction of variant pathogenicity. The tools are listed with their references in Table 2.3.

Table 2.3. The pathogenicity predicting *in silico* tools used in the study.

***In silico* pathogenicity predicting tool**	**Reference**
CADD	Rentzsch *et al.,* 2019
DANN	Quang, Chen and Xie, 2015.
dbscSNV	Jian, Boerwinkle, and Liu, 2014.
FATHMM	Shihab *et al.,* 2013
GERP	Cooper *et al.,* 2005
MetaLR	Dong *et al.,* 2015
MutationAssessor	Reva, Antipin and Sander, 2011.
MutationTaster	Schwarz *et al.,* 2014
PolyPhen2	Adzhubei, Jordan and Sunyaev, 2013
REVEL	Ioannidis *et al.,* 2016
SIFT	Kumar, Henikoff and Ng, 2009

These tools predict pathogenicity based on the type of variant, the variant effect at nucleotide level and on the protein structure and function (Richards *et al.*, 2015). The prediction of variant pathogenicity is also on the basis of whether the variant affects a functional domain and/or an evolutionary conserved region (MacArthur *et al.*, 2014). Variants predicted to be non-deleterious by computational pathogenicity predicting tools were filtered out.

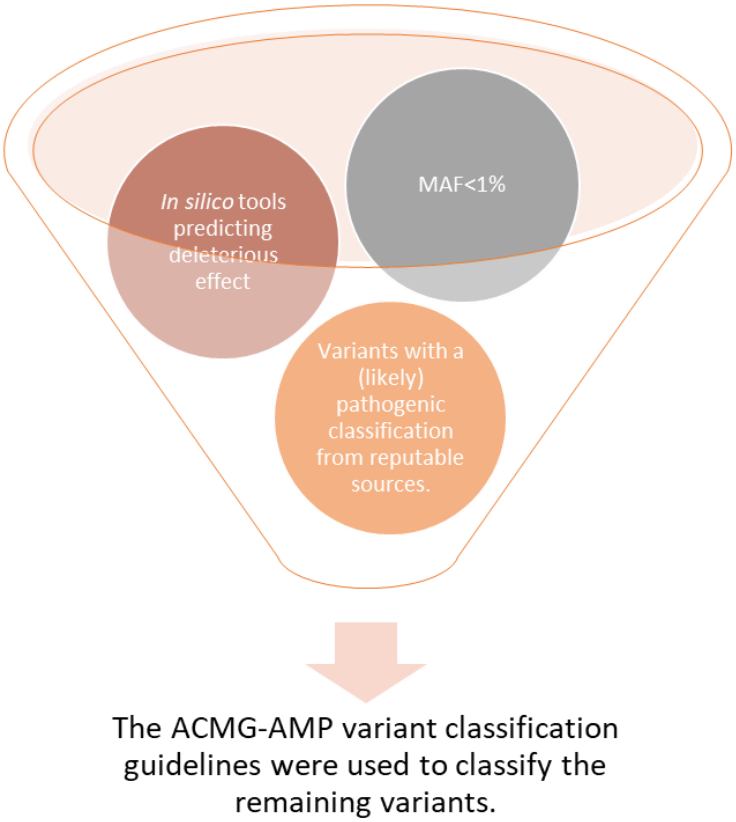


Figure 2.3. Flow diagram indicating the filtering pipeline for the identification of variants implicated in the causation of HCTDs. The information from various databases was used in combination. The remaining variants were then classified according to the ACMG-AMP variant classification guidelines.

2.4.3.4 Variant Classification

The prioritized variants were then interpreted according to the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) variant interpretation guidelines (Richards *et al.,* 2015). These are widely used guidelines in the classification of variants in diagnostic laboratories (Ghosh, Oak and Plon, 2017), providing a systematic approach to classifying variants. They have been used in numerous other studies investigating HCTDs (Mannucci *et al.,* 2019, Xu *et al.,* 2019).

The classification of the variants was based on the five tier system of classification; pathogenic (P), likely pathogenic, variant of uncertain significance (VUS), likely benign and benign (B). The evidence supporting variant classification is based on the combined factors such as *in silico* tools, functional data, inheritance models, population frequency data, segregation patterns and the patients' phenotype or family history (Richards *et al.,* 2015). The ACMG-AMP guidelines support the incorporation of the evidence to assist in the classification of variants, rather than the use of single pieces of evidence. Variant classification was then based on different levels of evidence; very strong (VS), strong (S), moderate (M) and supporting (P). For genes with available gene-specific guidelines, such as the *FBN1*-specific modified 2015 ACMG-AMP criteria, these were also used (Muiño-Mosquera *et al.,* 2018, Baudhuin *et al.,* 2019).

2.4.4 Post-alignment Quality Control

Variants classified as (likely) pathogenic according to the ACMG-AMP guidelines were further examined for quality. The BAM files were viewed in Integrative Genomics Viewer 2.8.9 (IGV) (Thorvaldsdóttir, Robinson, and Mesirov, 2013) to determine if variants were sufficiently covered, hence a base could be confidently called. This is important because it helps identify poorly covered regions and aids in the identification of variant calls that may be sequencing artefacts.

2.5. Validation of Variants by Sanger Sequencing

Sanger sequencing is described as the gold standard for DNA sequencing, hence it is often used for validation of NGS-obtained variants before they are returned to participants (Beck *et al.,* 2016). Sanger sequencing validated variants with a pathogenic or likely pathogenic classification according to the ACMG-AMP guidelines of variant interpretation. Once variants were identified, Sanger sequencing primers (Table 2.4) in this study were designed using Primer3 v.0.4.0 (Untergasser *et al.,* 2012). They were then submitted to NCBI's Primer-BLAST (NCBI Resource Coordinators, 2014) to ensure they were specific to our gene regions of interest. Subsequent to confirming primer specificity, the primers were purchased from Inqaba BioTech™ (Inqaba Biotechnical Industries, Pretoria, South Africa).

Table 2.4. Sanger sequencing primers for validation of putative disease-causing variants.

Gene	Variant	Forward Primer (5'-3')	Reverse Primer (5'-3')	Product Size (bp)
FBN1	c.5726T>C	TTGACAAGTCCCATAACCAATTG	CAACATCAAGAAGTTCTCAGCCT	386
FBN1	c.3794G>A	GCTCTCTTTGGAATGCTGGT	GGCTAAGTTTATTTGACTGCGG	381
FBN1	c.640G>A	GAGAGCCACGTTCCTAATGC	ACAGGGACAGTGATGCATGA	499
FBN1	c.3037G>A	CTGACATTTAAGAGGGCCACC	CCCTCTGACAGATATCCGCC	491
FBN1	c.6670dupA	ATATTCCAATTCCCAGCCTTCT	GGAGACCACTTGATCCATCCA	366
FBN1	c.23_34del	CGGGGATTTGTCTCTGTGTT	CCACACCAAAGGAGGGAAC	400

PCR optimization was performed to determine the most optimal annealing temperature of the primers by means of gradient PCR, using standard PCR reagents. Post PCR, the PCR products were separated using 2% agarose gel electrophoresis. The gels were then visualized using the Omega Fluor™ Gel Documentation System (Vacutec, Johannesburg, South Africa). From the gel images, the optimum annealing temperature (Table 2.5), specific to each primer pair was determined and consequently used to PCR amplify the target regions of our patients' DNA.

Prior to Sanger sequencing, the amplicons were purified in the MultiScreen® PCRμ96 Filter Plates (Merck Millipore, Cork, Ireland) with double-distilled water (ddH2O) using the Millipore Millivac® Maxi SD1P014M04 vacuum pump (Merck Millipore, Cork, Ireland). The cleaned amplicons were then cycle sequenced using the BigDye™ Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo Fisher Scientific, Waltham, Massachusetts, United States) with the sequencing thermal cycler conditions indicated in Table 2.6.

Amplification of the target regions by cycle sequencing utilized amplicons as the DNA template, and the variant-specific forward and reverse primers as sequencing primers, DNA polymerase, traditional dNTPs and the fluorescence-tagged, chain terminating di-deoxynucleotidetriphosphates (ddNTPs). Di-deoxynucleotidetriphosphates lack the 3'-hydroxyl group, crucial for the formation of phosphodiester bonds between two nucleotides. Phosphodiester bonds are formed between the 3'-hydroxyl group of one nucleotide and the 5'-phosphate group of another. The MultiScreen® PCRμ96 Filter Plates (Merck Millipore, Cork, Ireland) were then again used for post sequencing clean-up of the samples, with the Millipore Injection solution (Merck Millipore, Cork,

Ireland), also using the Millipore Millivac® Maxi vacuum pump (Merck Millipore, Cork, Ireland).

Table 2.5. Optimized primer-specific PCR conditions for each variant.

Variant	PCR Conditions					
	Denaturation Temperature (10 minutes)	Denaturation Temperature (1 minute)	Annealing Temperature (1 minute)	Extension Temperature (1 minute)	Extension Temperature (5 minutes)	Hold ∞
c.5726T>C	95°c	95°c	64.3°c	72°c	72°c	4°c
c.3794G>A	95°c	95°c	64.3°c	72°c	72°c	
c.640G>A	95°c	95°c	65°c	72°c	72°c	
c.3037G>A	95°c	95°c	61.8°c	72°c	72°c	
c.6670dupA	95°c	95°c	58.8°c	72°c	72°c	
c.23_24del	95°c	95°c	55°c	72°c	72°c	

Table 2.6. Cycle sequencing thermal cycler conditions

Temperature	Time	Cycle(s)
96°c	1 minute	1
96°c	30 seconds	25
50°c	15 seconds	
60°c	4 minutes	
4-12°c	Hold	1

Following post-sequencing sample purification, the samples were denatured, rendering them single-stranded. They were then separated by size using the fluorescence-based Genetic Analyzer 3130xl (Applied Biosystems®, Thermo Fisher Scientific, Waltham, Massachusetts, United States). The fluorescence-tagged ddNTPs were then detected by a laser and read as the nucleotide sequence for the DNA fragments. The resulting electropherograms were then analysed using the sequence alignment and analysis software BioEdit (Hall, 1999).

3. RESULTS

3.1 Patient Characterization

Twelve families (9 African, 1 Caucasian and 2 of mixed ancestry) with a phenotype suggestive of one of the seven HCTDs that were of interest to the present study were invited to participate in this study (Table 3.1). There were 3 affected individuals in one family, 2 in a second and the other 10 families were represented by one patient each for a total of 15 patients. Eleven had variable clinical features suggesting MFS with only five of these meeting the Ghent criteria for the disorder (Loeys *et al.,* 2010). One patient had a clinical phenotype overlapping MFS and vEDS. The clinical features of three other patients suggested vEDS, CCA and a HCTD that could not be specifically classified. There were no patients with a suspected or clinical diagnosis of SGS, LDS and homocystinuria.

Table 3.1. Patients' clinical features in different systems suggesting inherited disorder(s) of connective tissue.

Patient	Sex	Age	Suspected/ Clinical Diagnosis	Meets Criteria (At initial Visit)	D/A	Family History	Clinical Features				
							Ocular	Cardiac	Systemic Score	Skin	Other
HCTDs_09	M	16	MFS	No	A	No	My	N/A	7	Striae	N/A
HCTDs_13*	F	5	MFS	No	A	Yes	N/A	N/A	3	N/A	N/A
HCTDs_14*	F	33	MFS	No	A	Yes	N/A	MVP	7	N/A	N/A
HCTDs_21	M	32	MFS/vEDS	No	A	N/A	N/A	AV, MV	7	Excessive scarring	Joint hypermobility
HCTDs_22	F	11	MFS	Yes, Ghent	A	No	EL	AD	4	N/A	N/A
HCTDs_23	F	37	MFS	No	A	No	X	X	8	Striae	N/A
HCTDs_24*	F		MFS	Yes, Ghent	A	Yes	EL	X	7	Striae	N/A
HCTDs_25*	M	12	MFS	Yes, Ghent	A	Yes	EL, My	AD	7	N/A	Hyperactive
HCTDs_26*	F	7	MFS	Yes, Ghent	A	Yes	EL	N/A	9	N/A	Arachnodactyly
HCTDs_27	M	22	MFS	No	A	Yes	N/A	MVPR	4	N/A	Learning problem
HCTDs_28	M	7	HCTD		A	N/A	N/A	AR		N/A	Global delay
HCTDs_29	M	9	MFS	Yes, Ghent	A	N/A	EL, My	AD, MVPR	>7	N/A	Developmental delay (DD), Brain anomaly
HCTDs_30	F	3	CCA		A	N/A	N/A	N/A		N/A	Mild DD, Limb, joint Contractures
HCTDs_31	F	28	vEDS	No	D	No	N/A	ARUP, VR	6	Easy and severe bruising, fibrosis, thin skin	Bleeding into neck, torn ligaments, joint hypermobility, large spontaneous haematoma
HCTDs_32	F	12	MFS	No	A	No	N/A	MVP	7	N/A	Learning and Hip problem

A-Alive, AD-Aortic dilatation, AR-Aortic regurgitation, ARUP-, Aortic rupture, AV- Aortic valve, D-Deceased, DD-Developmental delay, EL-Ectopia lentis, F-Female, M-Male, My-Myopia, MV, Mitral valve, MVP-Mitral valve prolapse, MVPR-Mitral valve prolapse with regurgitation, N/A-information not available, VR-Valve regurgitation, X- unspecified involvement of system. *Related patients from two families are represented with the blue and red colour, respectively.

3.2 NGS Data Quality Control

Two sequencing runs were performed in this study, with 12 patients sequenced in the first NGS sequencing run and the remaining 3 sequenced in the second NGS sequencing run.

The quality of the sequencing data was determined by evaluating post-sequencing metrics, pre- and post-read alignment. Chip loading, polyclonality, low quality reads and usable reads in this study were observed to be 94%, 27%, 6% and 68%, respectively, for the first sequencing run. For the second sequencing run, these were respectively observed to be 94%, 24%, 6% and 71%. The sequencing runs had an average read length of >225bp for all samples. This therefore indicates that the data used in this study were of good quality. The ISP loading density, which also indicated optimum data quality, further demonstrated the quality of the data and this was represented by the wide distribution of the red colour (Figure 3.1 and Figure 3.2).

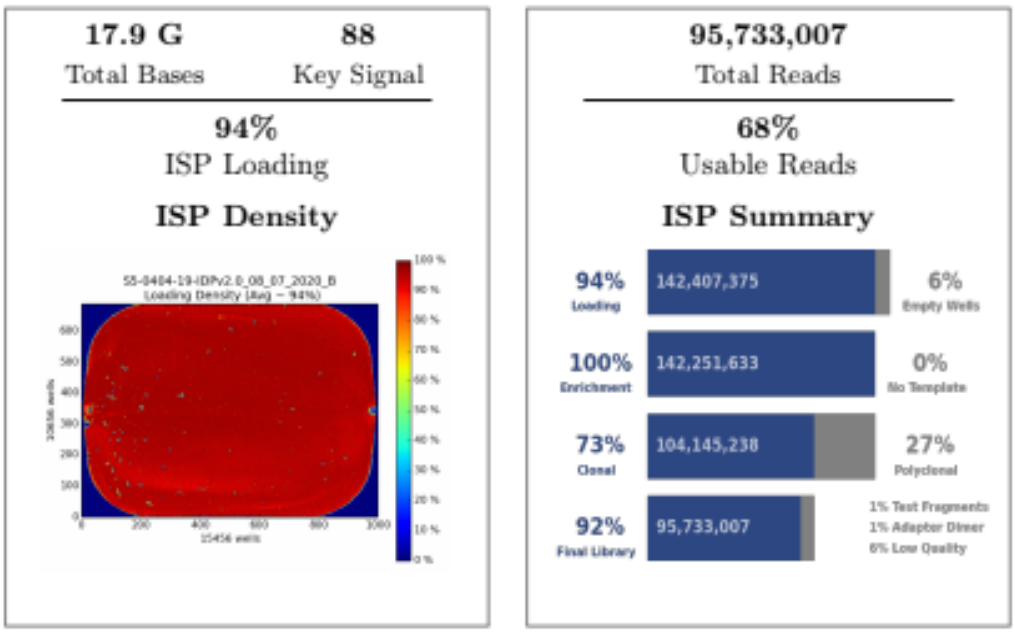


Figure 3.1. Twelve patients' samples were sequenced in the first NGS-sequencing run. These images were obtained from the first sequencing run report and show the post-sequencing metrics assessed to determine the quality of the run: high chip loading (94%) and ISP density (red colour) were observed. Low quality, primer dimers and polyclonal reads were filtered out, resulting in 68% usable reads.

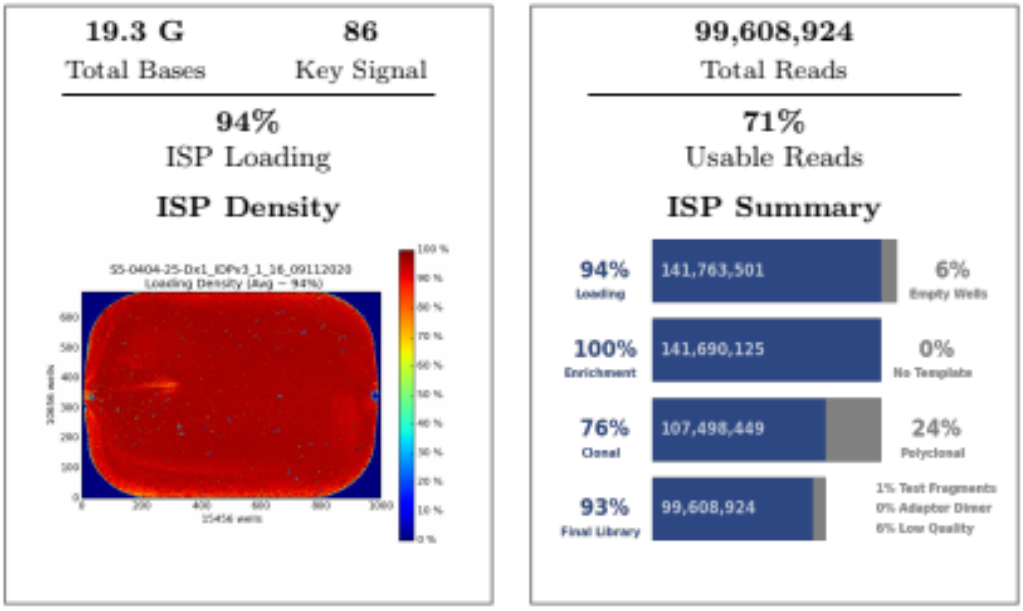


Figure 3.2. Images from the second NGS-sequencing run report. These images indicate the data quality of the sequencing run where three of our patients were sequenced. The data obtained were of good quality, and optimum chip loading (94%) and ISP density (red colour) were observed. Low quality, primer dimers and polyclonal reads were filtered out, resulting in 71% usable reads.

The mean sequencing depth of coverage for the targeted genomic regions was observed to be 350-480X (Table 3.2), as determined by the CoverageAnalysis pre-installed plug-in in Ion Reporter™ (Thermo Fisher Scientific, Waltham, MA, USA). Coverage was also confirmed by manual inspection of Integrative Genomics Viewer (IGV) plots (Thorvaldsdóttir, Robinson, and Mesirov, 2013), by visualising the BAM files of the patients. The variant allele fraction (VAF) of identified variants were similarly evaluated using Ion Reporter™ software (Thermo Fisher Scientific, Waltham, MA, USA) to identify suspicious or low confidence genotype calls.

True heterozygous variants of diploid organisms are required to have an allele balance of approximately 0.5 (Muyas *et al.,* 2019). This means that the reads of both the wild-type and the alternative allele must each be nearly 50% to unquestionably define the variants as true positives. The coverage of the identified heterozygous variants in this study was >190x, and they are shown together with the allele ratios in Table 3.3. This is another quality metric evaluated to determine the NGS-data quality.

Table 3.2. The mean depth of coverage observed for all patients in both NGS sequencing runs and the number of variants identified in each patient prior to variant filtering.

Patient	Sequencing Run	Overall Coverage Depth	Coverage Uniformity (%)	Number of Identified Variants before filtering (called by Ion Reporter™)
HCTDs_09	Second	469.9	95.18	56
HCTDs_13	Second	484.7	95.01	92
HCTDs_14	Second	352.8	94.51	89
HCTDs_21	First	418.7	95.43	80
HCTDs_22	First	359.9	95.39	97
HCTDs_23	First	380.9	95.42	76
HCTDs_24	First	441.4	94.72	90
HCTDs_25	First	366.8	94.85	84
HCTDs_26	First	440.1	95.60	89
HCTDs_27	First	397.2	95.37	92
HCTDs_28	First	412.8	96.00	70
HCTDs_29	First	371.1	95.51	92
HCTDs_30	First	416.3	95.11	108
HCTDs_31	First	406.5	95.35	63
HCTDs_32	First	377.3	95.56	79

3.3 NGS Variants Identified

Subsequent to the extensive filtering and prioritization processes, described in section 2.4.3.3, the remaining variants were classified according to the ACMG-AMP variant interpretation guidelines, indicated in Appendix C (Richards *et al.,* 2015). This section describes the variants prioritized and further investigated in the participants of this study, the *in silico* pathogenicity predictions tools used and the ACMG-AMP criteria codes applied to reach the variant classification.

3.3.1 Putative Disease-Causing Variants

This section will define the variants described as putative disease-causing in each patient or family based on data obtained from population, disease, mutation and protein databases, e.g. gnomAD, ExAC, 1000 Genomes Project, ClinVar, HGMD, UMD-FBN1, LOVD-COL3A1 and UniProtKB. It will also describe the phenotype associated with the variants if they had previously been identified in the literature as well as the clinical data of our patients harbouring the variant(s).

The section will also indicate the IGV images for variants classified as pathogenic or likely pathogenic (Figures 3.4, 3.7, 3.9, 3.13, 3.15 and 3.20) and the gel electrophoresis images (Figures 3.5 and 3.10). Post-PCR amplification, some of the variants identified as pathogenic or likely pathogenic were then validated by Sanger sequencing. Only one DNA sequence electropherogram (either the forward or reverse strand) for the validated variants of some patients is shown below because the other was of insufficient quality to review (Figures 3.6; 3.8; 3.11; 3.14). The electropherograms for the forward and reverse strand of one patient were both of adequate quality to analyse, hence both are indicated in Figures 3.16 and 3.17, respectively. One likely pathogenic variant identified in patients HCTDs_13 and HCTDs_14 could not be validated due to time constraints.

Table 3.3. The coverage and allele fraction of the (likely) pathogenic variants identified in this study.

Patient	Heterozygous Variant	Gene	Allele Coverage	Allele Count	Allele Ratio
HCTDs_13	c.23_34del	*FBN1*	395	AGATCGCCCTGGG=210 G=185	AGATCGCCCTGGG=0.53 G=0.47
HCTDs_14	c.23_34del	*FBN1*	375	AGATCGCCCTGGG=190 G=185	AGATCGCCCTGGG=0.51 G=0.49
HCTDS_21	c.6670dupA	*FBN1*	399	C=214 CA=185	C=0.54 CA=0.46
HCTDS_22	c.640G>A	*FBN1*	400	G=203 A=197	G=0.51 A=0.49
HCTDs_23	c.3794G>A	*FBN1*	400	G=208 A=192	G=0.52 A=0.48
HCTDS_24	c.5726T>C	*FBN1*	197	T=107 C=90	T=0.54 C=0.46
HCTDS_25	c.5726T>C	*FBN1*	217	T=111 C=106	T=0.51 C=0.49
HCTDS_26	c.5726T>C	*FBN1*	299	T=140 C=159	T=0.47 C=0.53
HCTDS_29	c.3037G>A	*FBN1*	272	G=146 A=126	G=0.54 A=0.46

3.3.1.1 Family HCTDs_2456.

The HCTDs_2456 family is one with three patients enrolled in the study, HCTDs_24, HCTDs_25 and HCTDs_26. HCTDs_24 is the mother to HCTDs_25 and HCTDs_26, a male and female, respectively, as indicated in Figure 3.3. Patients HCTDs_24, HCTDs_25 and HCTDs_26 have classic MFS, as they fulfil the Ghent criteria for MFS (Yang *et al.,* 2018). All 3 patients fulfilled the Ghent criteria because they have a family history of MFS and either aortic dilatation, ectopia lentis and a systemic score of ≥7, as shown in table 1.2 (Loeys *et al.,* 2010).

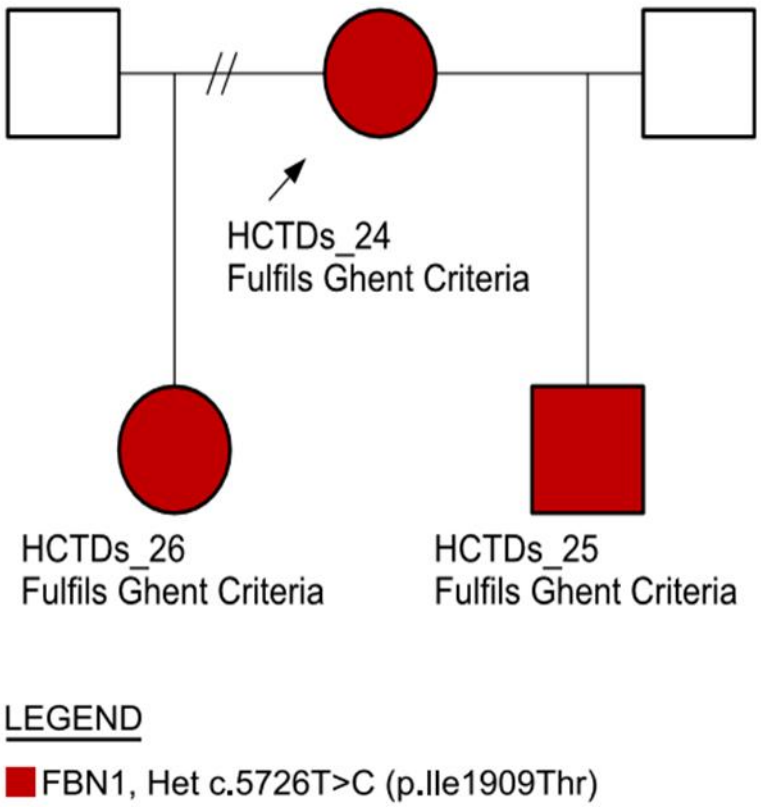


Figure 3.3. Pedigree of family HCTDs_2456.
The likely pathogenic c.5726T>C *FBN1* heterozygous missense variant was identified in the HCTDs_2456 family.

Mutations resulting in monogenic disorders in multiple affected family members segregate with the disease phenotype and therefore the disorder is presumed to be caused by the same pathogenic variant (Gilissen *et al.,* 2012). We used a family-based filtering approach as advised by Ott *et al*. (2015) when analysing the sequencing data for this family. This involved excluding all variants that are not shared by the affected family members to identify the putative disease-causing variant. Hence, all variants shared by at least two family members were prioritised for further investigation.

Patients HCTDs_24, HCTDs_25 and HCTDs_26 had 3, 2 and 2 variants following the segregation-based variant filtering strategy employed in addition to the broad variant filtering approach used in this study (Table 3.4). One missense and two intronic variants were detected. The most probable disease-causing variant identified was the *FBN1* (NM_000138.4): c.5726T>C (p.Ile1909Thr) missense variant present in all 3 family members.

Table 3.4. The three prioritised variants for further investigation in family HCTDs_2456.

Patients	Gene	Variant	Functional Domain/ Conserved Region	*In Silico* tools predicting a Benign effect	*In Silico* tools predicting a pathogenic effect	ClinVar Accession Number	dbSNP ID	HGMD ID	ACMG-AMP codes applied	ACMG-AMP Classification
HCTDs_24; HCTDs_25	*COL3A1*	c.1347+25G>C	Variable region	DANN			rs111708979		PM2, BP4	VUS
HCTDs_24; HCTDs_25; HCTDs_26	*FBN1*	c.5726T>C (p.Ile1909Thr)	cb-EGF-like domain		SIFT, CADD, MetaLR, REVEL, FATHMM	VCV000200189	rs794728333	CM013934	PM1, PM2, PM5, PM7, PP2,PP3, PP4, PP5	Likely Pathogenic
HCTDs_24; HCTDs_26	*MYH11*	c.3672+5T>G	Not strongly conserved	DANN			rs794727439		PM2, BP4	VUS

Investigation of the prioritised variants revealed that all three patients of the HCTDs_2456 family are heterozygous for the *FBN1* (NM_000138.4): c.5726T>C (p.Ile1909Thr) variant, found in exon 47 of the gene. The ACMG-AMP criteria code PP3 can be applied as the variant has deleterious annotations (Sefid and Gamieldien, 2017) from multiple *in silico* tools, SIFT, CADD, MetaLR, REVEL and FATHMM. The variant has been reported in ClinVar (accession number VCV000200189) and HGMD (CM013934) (ACMG-AMP criteria code PP5). A missense variant affecting the same codon as the c.5726T>C variant, with a different nucleotide change; c.5726T>G (p.Ile1909Ser) has also been observed in HGMD (CM098703) (Stenson *et al.,* 2003). This variant was identified as pathogenic (ACMG-AMP criteria code PM5).

The c.5726T>C variant is absent from control databases (gnomAD, ExAC, 1000 Genomes Project) (ACMG-AMP criteria code PM2). This variant is located in the cb-EGF-like domain (McLaren *et al.,* 2016) (ACMG-AMP criteria code PM1). The *FBN1* gene has significantly less benign missense variation than expected as determined by the evaluation of the ExAC missense constraint score and the ACMG-AMP criteria code PP2 was therefore applied as recommended by Baudhuin *et al*. (2019).

A positive family history and co-segregation of the variant with the disease phenotype is suggestive of the application of the ACMG-AMP criteria codes PP4 and PP1, respectively. The *FBN1*-specific ACMG-AMP guideline by Muiño-Mosquera *et al*. (2018) has upgraded the segregation-based ACMG-AMP PP1 criterion to moderate (PM7) should a variant segregate with a Marfan phenotype in 3-4 family members. This is why the ACMG-AMP PM7 criteria code was applied. We therefore classified the *FBN1* c.5726T>C missense variant as likely pathogenic.

The coverage of the putative disease-causing *FBN1* c.5726T>C variant identified in the HCTDs_2456 family was manually inspected using IGV plots (IGV images are depicted in Figure 3.4) and the variant was validated using Sanger sequencing. We were thus able to confirm that all three affected family members are heterozygous for the *FBN1* c.5726T>C missense variant. The Sanger sequencing electropherograms for patients HCTDs_24, HCTDs_25 and HCTDs_26 are depicted in Figure 3.6 in either the forward or reverse strand.

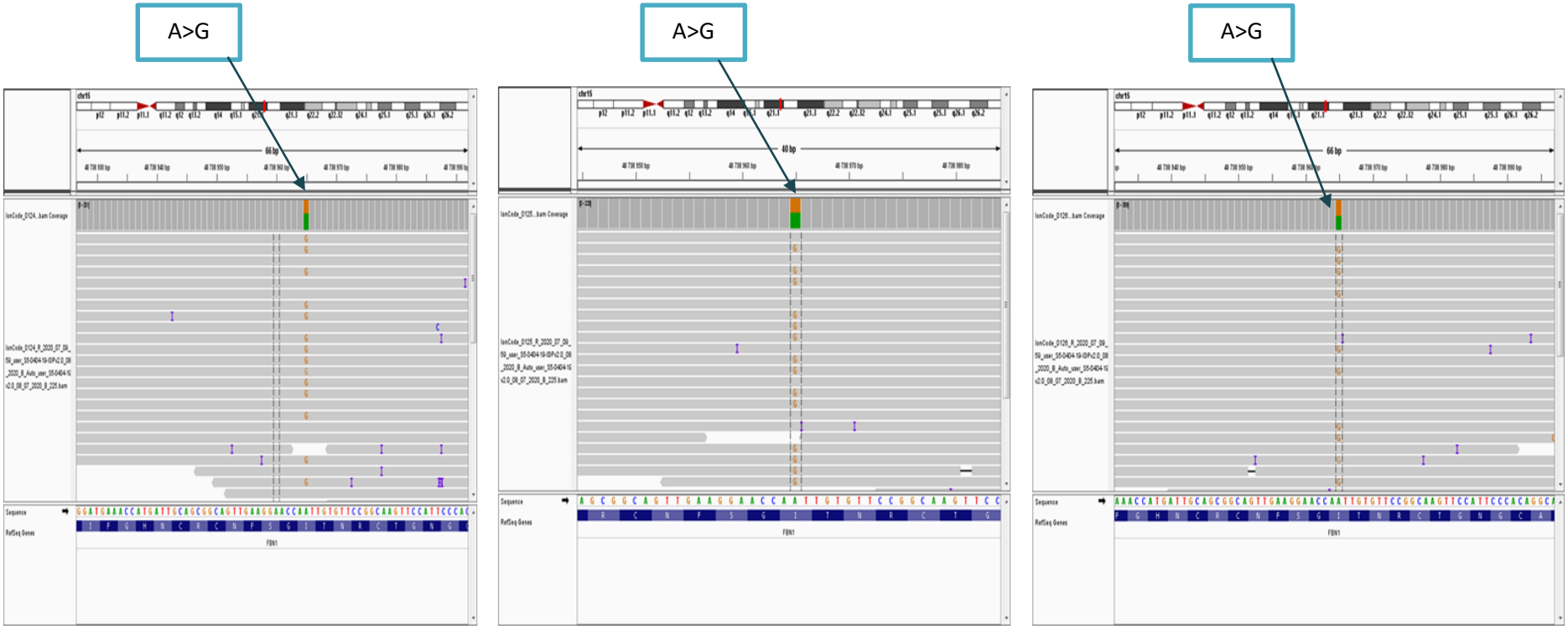


Figure 3.4. IGV images for variant c.5726T>C in exon 47 of the *FBN1* gene, identified in patients HCTDs_24 (**A**), HCTDs_25 (**B**) and HCTDs_26 (**C**). IGV reports this nucleotide change as the complimentary base (A>G); this is why the SNV is seen in the image as G.

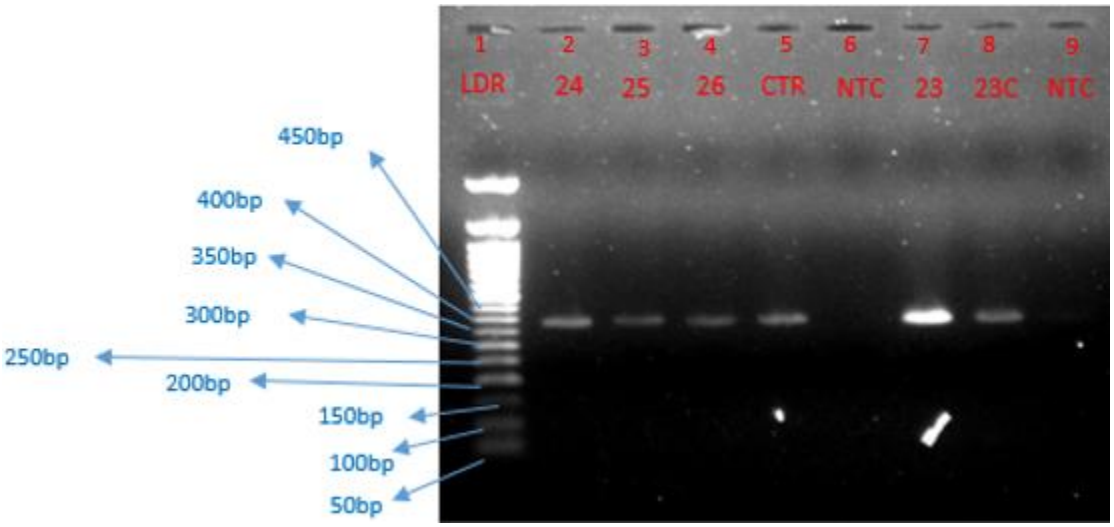


Figure 3.5. The PCR products run in 2% agarose gel for patients HCTDs_24, HCTDS_25, HCTDs_26 and HCTDs_23. Wells 1-9 were loaded with ladder (LDR), HCTDs_24 DNA, HCTDs_25 DNA, HCTDs_26 DNA, control DNA amplified with family HCTDs_2456's primers, non-template control (NTC), patient HCTDs_23 DNA, control DNA amplified with HCTDs_23 primers and non-template control, respectively. The product size of the HCTDs_2456 family is 386bp, while that of patient HCTDs_23 is 381bp and these were expected based on our primer design.

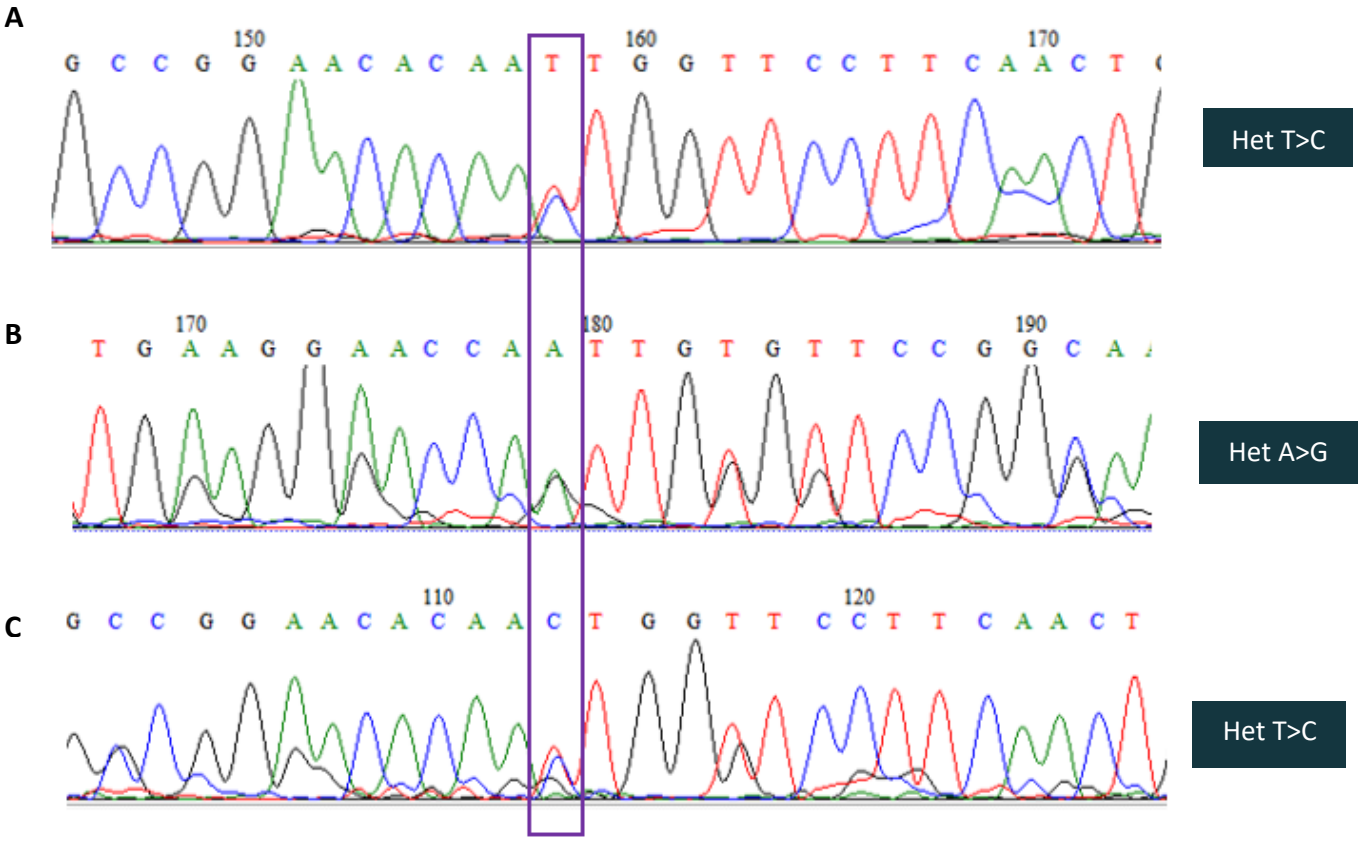


Figure 3.6. The sequence electropherograms for the *FBN1* heterozygous missense variant c.5726T>C in patients HCTDs_24 and HCTDs_26 is shown in the reverse complimentary strand (**A** and **C)**. The forward sequence chromatograph of patient HCTDs_25 is presented in image **B**.

In addition to the *FBN1* mutation identified, a heterozygous *COL3A1* (NM_000090.3): c.1347+25G>C variant, with a VUS classification was also identified in patients HCTDs_24 and HCTDs_25. This variant is absent from control databases, however DANN predicted a non-damaging effect to the gene product because it occurred in a non-conserved region (ACMG-AMP criteria codes PM2 and BP4, respectively). One other candidate variant was identified in HCTDs_24 and HCTDs_26, the heterozygous *MHY11* (NM_001040114.1): c.3672+5T>G variant, also classified as a VUS owing to the application of the ACMG-AMP criteria codes PM2 and BP4.

The identified intronic variants in this family are not deemed to be disease-causing and were not investigated further because a likely pathogenic variant was found segregating with the disease phenotype in this family. The identified likely pathogenic variant has been reported before in patients with a clinical phenotype that matches that found in our patients (Wang *et al.,* 2013).

3.3.1.2 Patient HCTDs_23

Table 3.5 The shortlisted variant in patient HCTDs_23.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	VarSome classification	ACMG-AMP codes applied	Current study ACMG-AMP classification
FBN1	c.3794G>A (p.Cys1265Tyr)	cb-EGF-like domain	SIFT, PolyPhen2, CADD, MetaLR, REVEL, FATHMM			CM085430	Likely Pathogenic	PS3, PM1, PM2, PM5, PP2, PP3	Pathogenic

Patient HCTDs_23 is a female with a tall stature, dysmorphic features, and aortic complications, suspected to have MFS. A heterozygous *FBN1* (NM_000138.4): c.3794G>A (p.Cys1265Tyr) variant in exon 30 was identified (Table 3.5). Exon 30 forms part of the "neonatal region", and mutations in this region are associated with neonatal MFS, a severe form of the disorder (Franken, 2016). ClinGen recommends a missense z-score of 3.09, however the *FBN1* gene has an ExAC missense z-score of 5.33 (ACMG-AMP criteria code PP2) (Baudhuin *et al.,* 2019). The identified variant has not been reported in population databases (gnomAD, ExAC, 1000 Genomes Project) (ACMG-AMP criteria code PM2). It also has not been reported in ClinVar,

however, it was described in HGMD (CM085430). Based on the HGMD database, four other missense changes reported as either likely pathogenic or pathogenic, affecting the same amino acid residue (codon 1265) have been previously observed (Stenson *et al.,* 2003). These are substituting cysteine residues with either glycine (CM098683), arginine (CM980729), serine (CM154832) or result in a termination codon (CM164448); c.3794T>G (p.Cys1265Gly), c.3794T>C (p.Cys1265Arg), c. 3794G>C (p.Cys1265Ser) and c.3794T>A (p.Cys1265Ter), respectively (ACMG-AMP criteria code PM5) (Stenson *et al.,* 2003).

Multiple *in silico* tools: SIFT, PolyPhen2, CADD, MetaLR, REVEL and FATHMM yielded consistent pathogenic predictions, supporting that the identified variant is possibly deleterious to the gene or its product (ACMG-AMP criteria code PP3). The c.3794G>A variant is located in a functional domain (cb-EGF-like domain) and alters a cysteine residue (ACMG-AMP criteria code PM1). Cysteine residues are crucial for the structural integrity of the fibrillin-1 protein, encoded by the *FBN1* gene (Baudhuin *et al.,* 2019), with *FBN1* cysteine eliminating mutations associated with more severe aortic complications than mutations introducing a cysteine residue (Stengl *et al.,* 2020). It is for this reason that the ACMG-AMP criteria code PM1 rule was upgraded to a strong criterion for pathogenicity (Xu *et al.,* 2019). Therefore, this variant was determined to be pathogenic.

The evidence supporting pathogenicity of the heterozygous *FBN1* c.3794G>A missense variant identified in patient HCTDs_23 was considered adequate to define it as putative disease-causing and was therefore inspected on IGV for coverage data and confirmed using Sanger sequencing. The IGV image showing the coverage and the DNA sequence electropherogram are indicated in Figures 3.7 and 3.8, respectively.

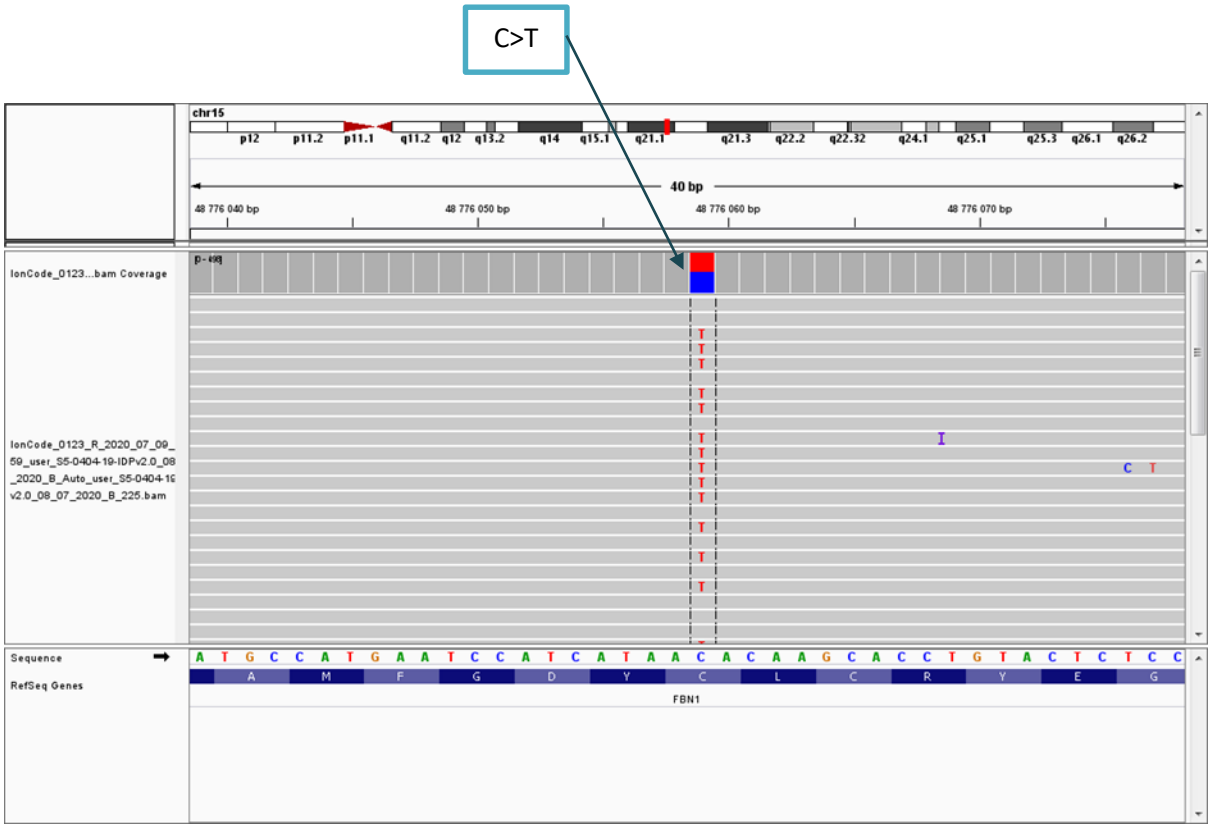


Figure 3.7. IGV image from patient HCTDs_23 indicating a change in nucleotide, c.3794G>A, in the *FBN1* gene, exon 31, reported as C>T in IGV.

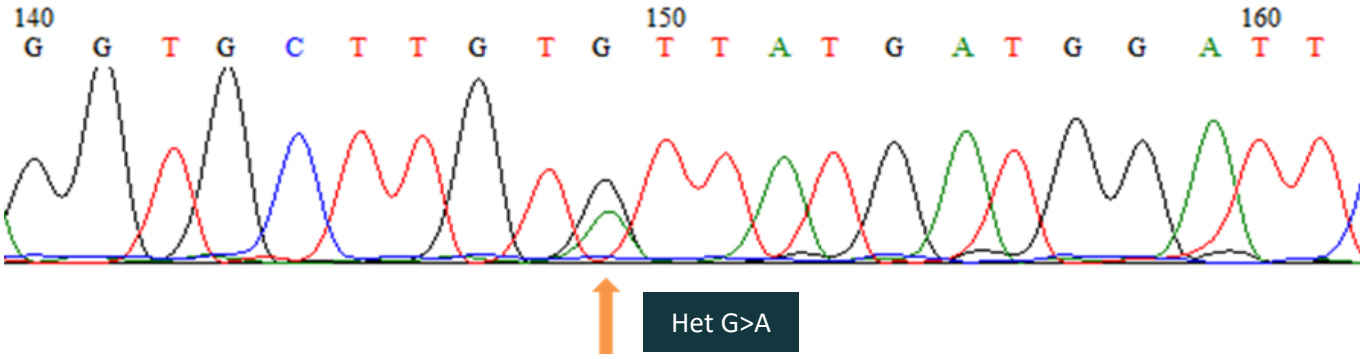


Figure 3.8. This electropherogram depicts the c.3794G>A heterozygous missense variant, found in the *FBN1* gene of patient HCTDs_23 with a pathogenic classification and validated using Sanger sequencing. The arrow shows the mutation position.

3.3.1.3 Patient HCTDs_22

Five variants were prioritised for further investigation in HCTDs_22 (as illustrated in Table 3.6) after the extensive filtering process employed in this study.

Patient HCTDs_22 is a female presenting with complications in the ocular, skeletal and cardiovascular systems with a clinical diagnosis of MFS. A heterozygous *FBN1* (NM_000138.4): c.640G>A (p.Gly214Ser) variant was identified in exon 7 of the gene. This variant has been repeatedly described in ClinVar (accession number VCV000199956) as pathogenic (ACMG-AMP criteria code PP5). It has not been reported in population databases such as gnomAD, ExAC and 1000 Genomes Project (ACMG-AMP criteria code PM2). Multiple *in silico* tools such as SIFT, PolyPhen2, CADD, MetaLR, REVEL and FATHMM are supportive of a damaging effect to the gene product, owing to the location of the variant being in the mutational hotspot; hybrid domain 1 (ACMG-AMP criteria codes PP3 and PM1, respectively).

The c.640G>A variant has been reported in the HGMD database (CM045409), together with another missense variant affecting the same amino acid residue, however caused by a different missense change; c.640G>C (p.Gly214Arg), with accession number CM094367 (Stenson *et al.,* 2003) (ACMG-AMP criteria code PM5). A 2bp substitution variant; c.640_641delGGinsTC (p.Gly214Ser), also in exon 7 of the *FBN1* gene has been reported in UniProtKB (UniProt Consortium, 2009). This variant was described as pathogenic and it produced the same change in amino acid as the variant identified in this patient (ACMG-AMP criteria code PS1).

Table 3.6. The shortlisted variants for patient HCTDs_22.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tools predicting a Benign effect	In Silico tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
CBS	c.1552+98C>T	Variable region	DANN			rs112883646		PM2, BP4	VUS
CBS	c.737-63T>C	Variable region	DANN			rs113750019		BP4	VUS
CBS	c.451+124C>T	Variable region	DANN			rs112120780		PM2, BP4	VUS
FBN1	c.640G>A (p.Gly214Ser)	Hybrid domain-1		SIFT, PolyPhen2, CADD, MetaLR, REVEL, FATHMM	VCV000199956	rs794728162	CM045409	PS1, PM1, PM2, PM5, PP2, PP3, PP5	Pathogenic
FBN1	c.6038-38C>T	Variable region	DANN			rs113682441		PM2, BP4	VUS

The ACMG-AMP criteria code PP2 was also applied for the *FBN1* gene, due to its ExAC missense z score of 5.33 (Baudhuin *et al.,* 2019). The c.640G>A variant has been frequently observed in the literature, segregating with family members affected with MFS or ELS (Dong *et al.,* 2012, Sui *et al.,* 2004). This provides more evidence supporting variant pathogenicity. Based on the ACMG-AMP variant classification guideline, this variant was therefore classified as pathogenic (Richards *et al.,* 2015).

The *FBN1* c.640G>A variant is the most probable disease-causing variant identified in patient HCTDs_22. An inspection of the variant coverage was carried out using IGV, and the IGV image is depicted in Figure 3.9. The variant was then validated using Sanger sequencing. The DNA sequence electropherogram is indicated in Figure 3.11. We therefore confirm that patient HCTDs_22 is heterozygous for the c.640G>A *FBN1* variant.

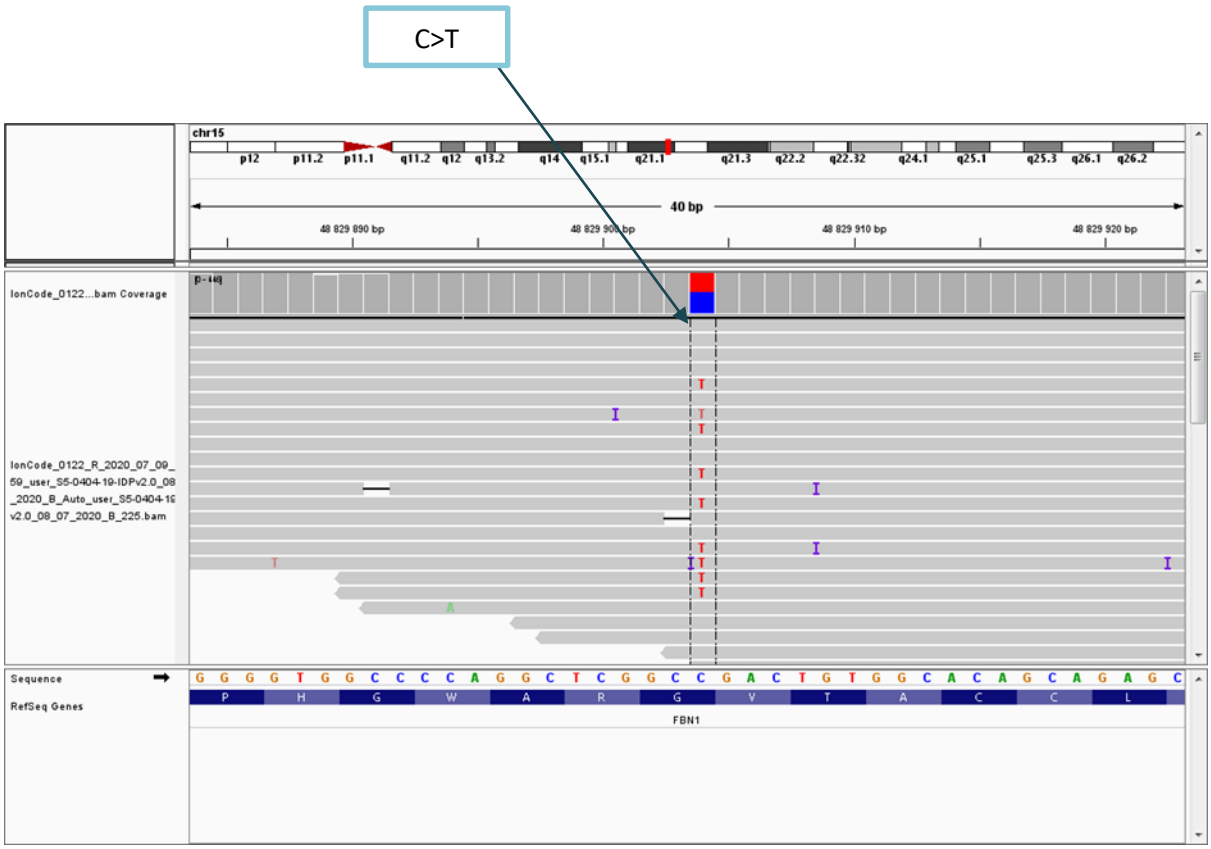


Figure 3.9. IGV image showing the c.640G>A variant in exon 7 of the *FBN1* gene. This is shown as the nucleotide change C>T in IGV and was observed in patient HCTDs_22.

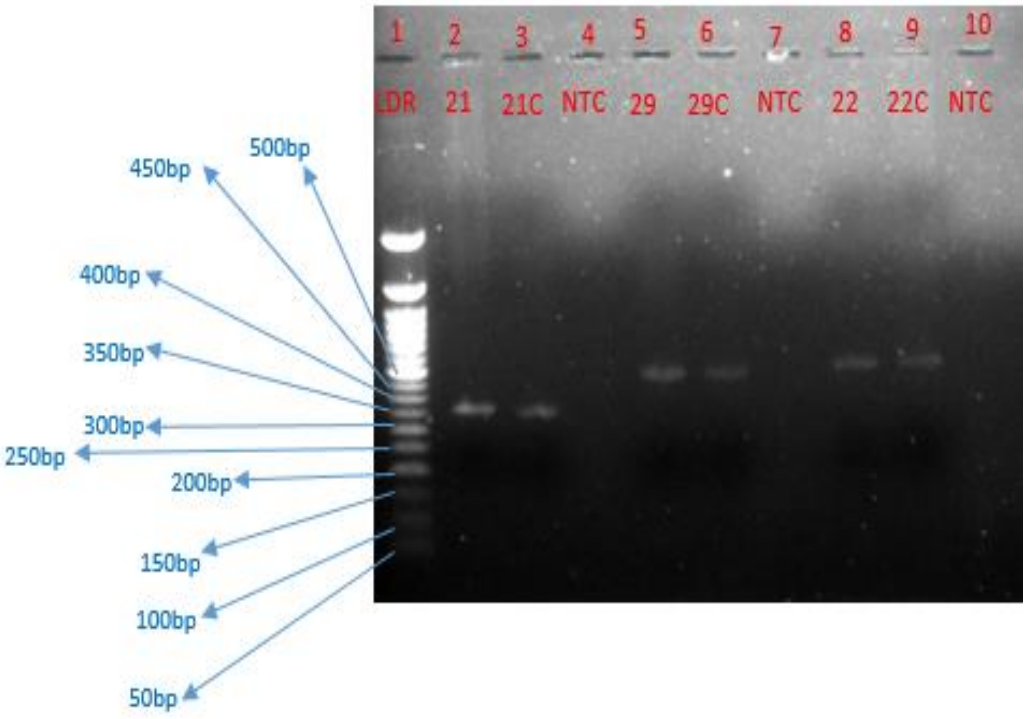


Figure 3.10. The PCR products run in 2% agarose gel electrophoresis. Wells 1-10 were loaded with DNA ladder (LDR), HCTDs_21 DNA, control DNA amplified with HCTDs_21 primers (21C), non-template control (NTC), HCTDs_29 DNA, control DNA amplified with HCTDs_29 primers (29C), non-template control, HCTDs_22 DNA, control DNA amplified with HCTDs_22 primers (22C) and non-template control, respectively. The product size of HCTDs_21, HCTDs_29 and HCTDs_22 is 366bp, 491bp and 499bp, respectively as expected based on our primer design.

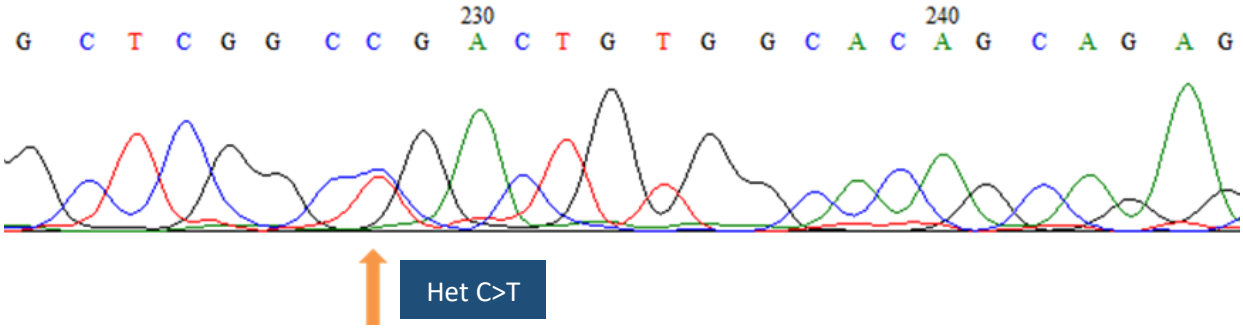


Figure 3.11. The heterozygous missense variant c.640G>A in the *FBN1* gene, seen as C>T in the forward sequence strand identified in HCTDs_22, indicated by the arrow.

Four heterozygous intronic variants were also identified in HCTDs_22. These are *CBS* (NM_000071.3): c.1552+98C>T; c.737-63T>C and c.451+124C>T and the *FBN1* (NM_000138.4): c.6038-38C>T variant. These variants occur in non-conserved regions and the *in silico* tool DANN predicted a non-damaging effect on splicing. This is why the ACMG-AMP criteria code BP4 code was assigned for all the variants. The variants; c.1552+98C>T, c.451+124C>T and c.6038-38C>T were absent from control databases (ACMG-AMP criteria code PM2). All the intronic variants were thus classified as VUSs. The intronic variants identified in patient HCTDs_22 were not deemed to be significant to the patient's phenotype because we identified the *FBN1* c.640G>A variant as the most likely putative disease-causing variant in this patient.

3.3.1.4 Patient HCTDs_21

Patient HCTDs_21 is a male who presented with cardiac valve lesions, excessive scarring and joint hypermobility with a differential diagnosis of MFS or vEDS. One variant remained as a candidate after the filtering steps employed (Table 3.7).

Table 3.7. The shortlisted variant in the *FBN1* gene identified patient HCTDs_21.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
FBN1	c.6670dupA (p.Thr2224AsnfsTer6)	cb-EGF-like domain	GERP, MutationTaster				PVS1, PM1, PM2, PP3	Pathogenic

The heterozygous *FBN1* (NM_000138.4): c.6670dupA (p.Thr2224AsnfsTer6) variant is a single-base insertion, causing a change in the reading frame of the codons. It introduced a stop codon, thereby resulting in a truncated protein. The Mutalyzer 2.0.32 software was used to predict the effect of the c.6670dupA insertion variant on the FBN1 protein, in relation to the reference protein sequence, and the predicted change is depicted in Figure 3.12. The ACMG-AMP criteria code PVS1 was applied because this variant causes a frameshift due to the base (A) insertion and it is found in a gene where loss-of-function is a recognised pathogenic mechanism for the associated disorder (Ma, Wang and Wei, 2016).

The c.6670dupA variant is located in exon 55 of the *FBN1* gene, affecting the cb-EGF-like domain (ACMG-AMP criteria code PM1). The variant is not reported in ClinVar, HGMD, UMD-FBN1 or in control databases such as gnomAD, ExAC and 1000 Genomes Project (ACMG-AMP criteria code PM2). The MutationTaster and GERP *in silico* pathogenicity predicting tools predicted a damaging effect to the gene product (ACMG-AMP criteria code PP3). As a result, this variant was classified as pathogenic.

The evidence supporting pathogenicity of the *FBN1* c.6670dupA frameshift variant was considered sufficient to describe it as possibly disease-causing, hence manual inspection of the variant coverage using IGV and verification using Sanger sequencing were carried out. The IGV image showing the c.6670dupA coverage and the DNA sequence electropherogram are indicated in Figures 3.13 and 3.14, respectively.

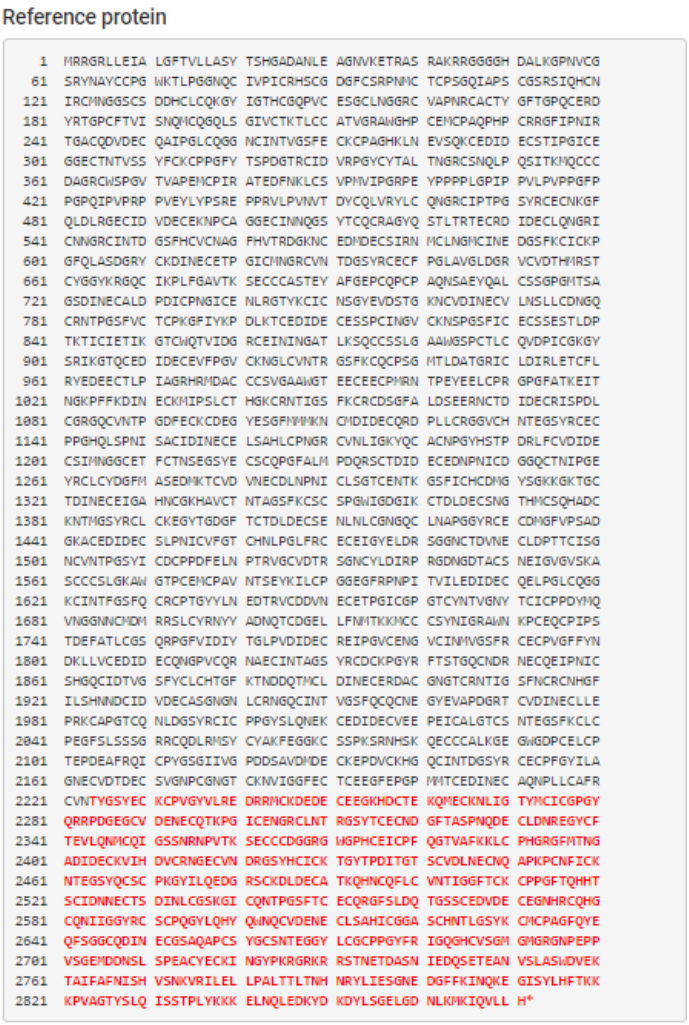


Figure 3.12. An image from Mutalyzer 2.0.32 depicting a shortened protein as a result of the heterozygous c.6670dupA pathogenic frameshift variant in the *FBN1* gene that introduced a stop codon.

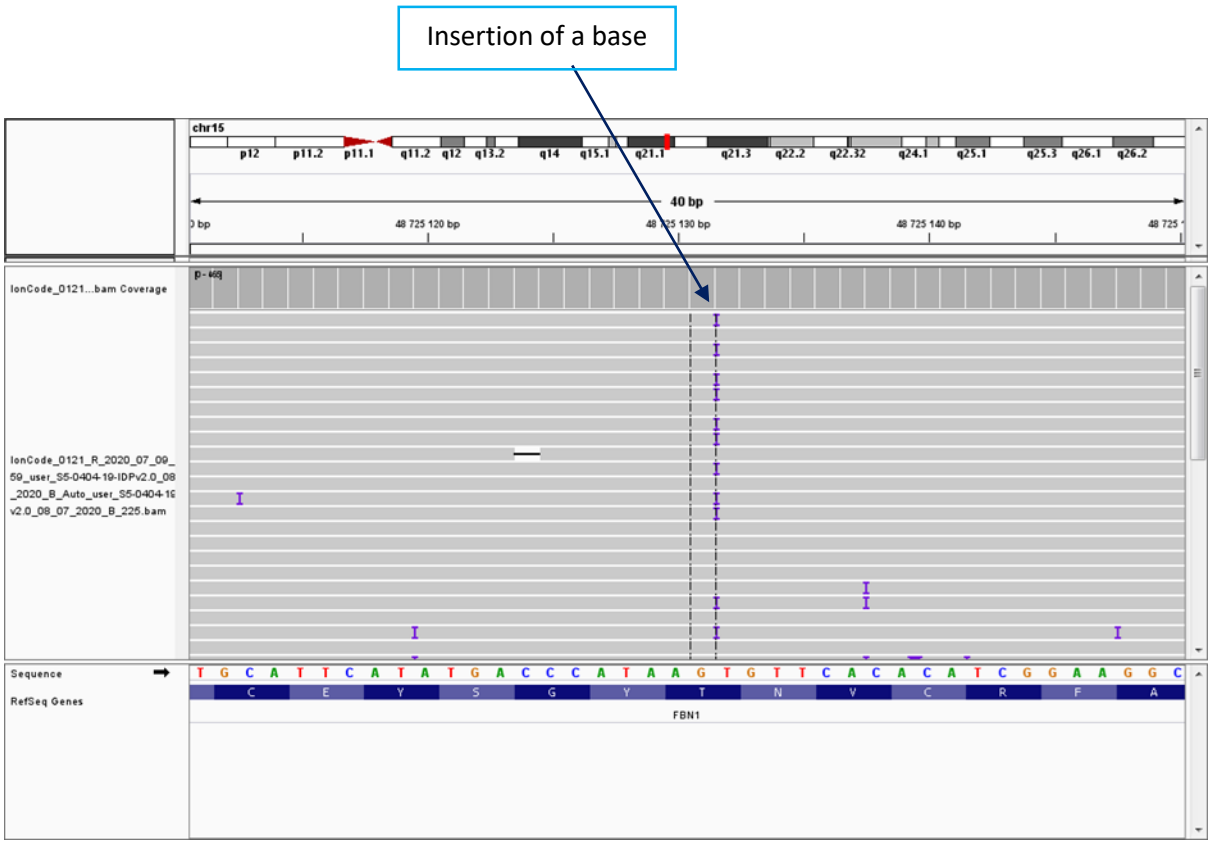


Figure 3.13. An IGV image depicting an insertion of an A base in exon 55 of the *FBN1* gene, seen in patient HCTDs_21.

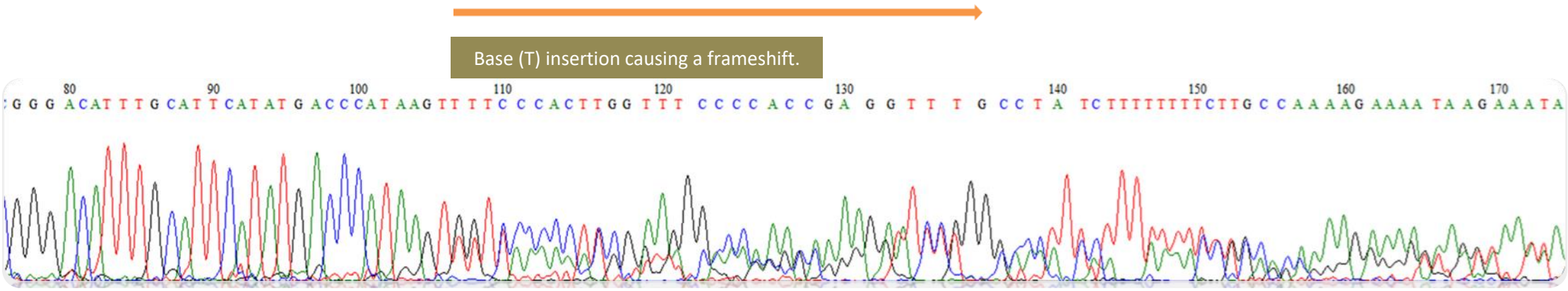


Figure 3.14. An electropherogram depicting the insertion of the base T (c.6670dupA) in the forward strand of the *FBN1* gene, causing a change in the reading frame of the codons (frameshift) from the position of the insertion in patient HCTDs_21.

3.3.1.5 Patient HCTDs_29

There were three variants that passed the filtering steps in patient HCTDs_29; 1 missense and 2 intronic variants, as indicated in Table 3.8.

Patient HCTDs_29 is a male with lens subluxation, joint hyperlaxity and low grade communicating hydrocephalus and fulfils the Ghent nosology due to the manifestation of ectopia lentis and aortic root dilatation. A heterozygous missense *FBN1* (NM_000138.4): c.3037G>A (p.Gly1013Arg) variant was identified in this patient. This variant is located in the “neonatal region”, in exon 25. This variant is also located in a functional domain (cb-EGF) (ACMG-AMP criteria code PM1) and is absent from control databases (gnomAD, ExAC, 1000 Genomes Project) (ACMG-AMP criteria code PM2). *In silico* tools such as SIFT, PolyPhen2, CADD, MetaLR, REVEL and FATHMM predict a damaging effect to the gene product (ACMG-AMP criteria code PP3). ClinVar (accession number: VCV000177648) and HGMD (CM950441) reported this variant as pathogenic (ACMG-AMP criteria code PP5).

The HGMD database also described another missense variant resulting in the same amino acid change in the same codon, c.3037G>C (p.Gly1013Arg, CM983865) described as pathogenic (Stenson *et al.,* 2003) (ACMG-AMP criteria code PS1). According to Baudhuin et al. (2019), the ACMG-AMP criteria code PP2 rule should be applied for *FBN1* missense variants because of its ExAC missense z-score of 5.33. This variant was therefore classified as pathogenic.

The *FBN1* c.3037G>A variant was described to be the putative disease-causing variant in patient HCTDs_29 based on the ACMG-AMP variant classification guideline (Richards *et al.,* 2015). The coverage of the variant was then manually inspected using IGV (Figure 3.15). Validation of the variant by Sanger sequencing was also carried out and the DNA sequence electropherograms of the forward and the reverse strand are depicted in Figures 3.16 and 3.17, respectively. We therefore confirm that patient HCTDs_29 is heterozygous for the *FBN1* c.3037G>A pathogenic variant.

Table 3.8. The three variants shortlisted in patient HCTDs_29.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tools predicting a Benign effect	In Silico tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
COL3A1	c.2607+35C>A	Variable region	DANN			rs115882631		PM2, BP4	VUS
FBN1	c.3037G>A (p.Gly1013Arg)	cb-EGF-like domain		SIFT, PolyPhen2, CADD, MetaLR, REVEL, FATHMM	VCV000177648	rs140593	CM950441	PS1, PM1, PM2, PP2, PP3, PP5	Pathogenic
FBN1	c.3838+16A>C	Not strongly conserved	DANN					PM2, BP4	VUS

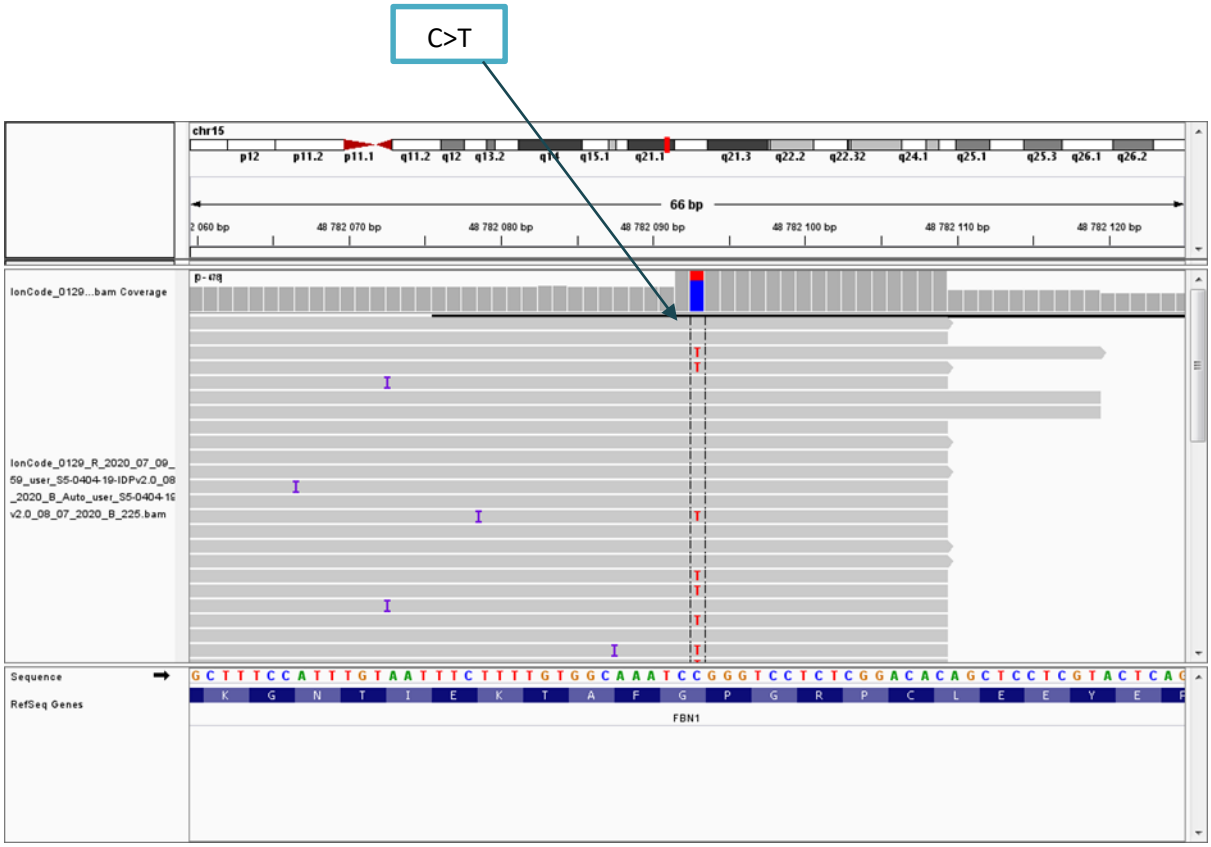


Figure 3.15. This IGV image indicative of the nucleotide change C>T. The c.3037G>A variant was identified in patient HCTDs_29 and it occurs in exon 25 of the *FBN1* gene.

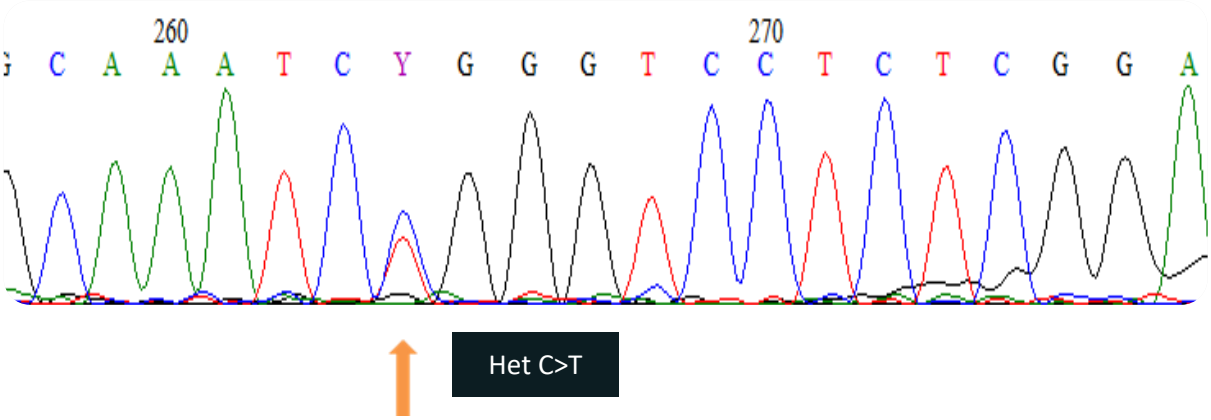


Figure 3.16. An electropherogram showing the heterozygous *FBN1* c.3037G>A missense variant identified in HCTDs_29. The identified variant is depicted as c.3037C>T in the forward strand.

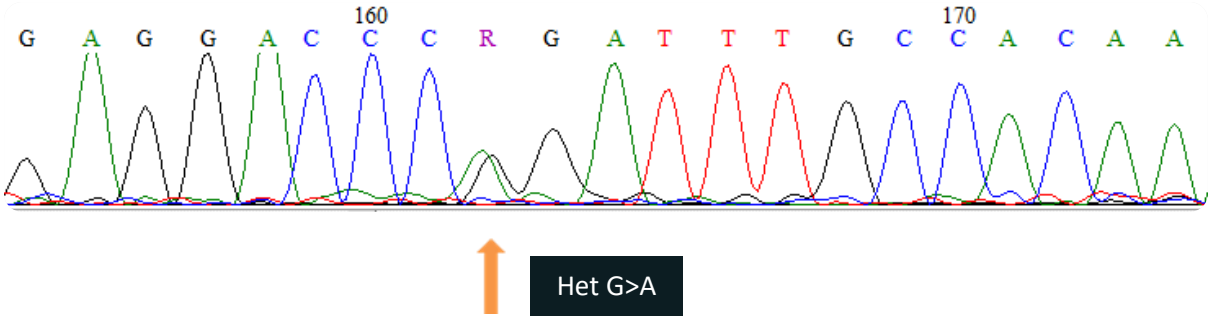


Figure 3.17. The electropherogram depicts the c.3037G>A heterozygous missense *FBN1* variant in the reverse strand identified in patient HCTDs_29.

The heterozygous *FBN1*(NM_000138.5): c.3838+16A>C variant was identified in the *FBN1* gene of patient HCTDs_29, with a benign prediction from DANN (ACMG-AMP criteria code BP4) and with no pathogenic predictions. This variant is located in a non-conserved region, although it is absent from control databases (ACMG-AMP criteria code PM2). This supported the VUS classification.

The heterozygous *COL3A1*(NM_000090.4): c.2607+35C>A (dbSNP ID; rs115882631) intronic variant, also identified in patient HCTDs_29 is located 35 bases from the splice site junction. Intronic variants at positions further from the exon-intron boundary do not typically disrupt splicing (Hellen, 2009). The GERP *in silico* tool predicted a non-damaging effect on splicing (ACMG-AMP criteria code BP4). This variant was however absent from control databases (ACMG-AMP criteria code PM2). For these reasons a VUS classification was given to this variant.

The *FBN1* c.3037G>A variant was identified as possibly causative of our patient's phenotype, therefore, the heterozygous c.3838+16A>C and c.2607+35C>A intron variants were not considered to be relevant.

3.3.1.6 Family HCTDs_143

The HCTDs_143 family is one with two patients enrolled in the current study, HCTDs_14 and HCTDs_13. HCTDs_14 is the mother to the female patient HCTDs_13, both presenting with clinical features indicative of MFS, as depicted in Figure 3.18.

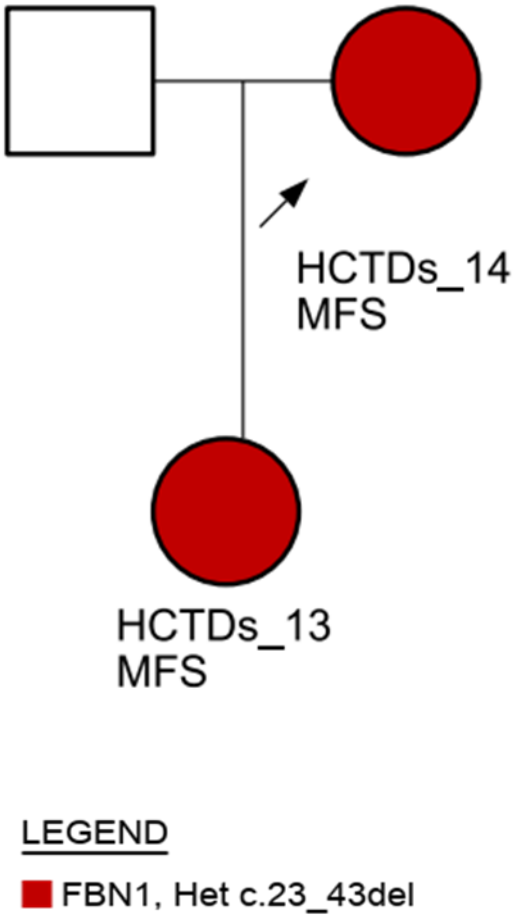


Figure 3.18. The pedigree for HCTDs_13 and HCTDs_14 with a suspected diagnosis of MFS with an identified heterozygous likely pathogenic variant.

Nine variants passed the filtering approach employed in the study. Two in HCTDs_14 and seven in patient HCTDs_13. However, only one variant was described as a candidate for the HCTDs_143 family after filtering the variants using the family segregation-based approach, as described by Ott *et al*. (2015) (Table 3.9).

Table 3.9. The identified variant following the family-segregation variant filtering strategy.

Patient (s)	Gene	Variant	Functional Domain/ Conserved Region	In Silico tool (s) predicting a Benign effect	In Silico tool (s) predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
HCTDs_13; HCTDs_14	*FBN1*	c.23_34del (p.Glu8_Leu11del)	Signal peptide domain		GERP				PM1, PM2, PM4, PP1, PP4	Likely Pathogenic

The identified variant is the heterozygous *FBN1*(NM_000138.5): c.23_34del (p.Glu8_Leu11del) and is a 12bp in-frame deletion occurring in exon 2 of the gene. The ACMG-AMP criteria codes PP1 and PP4 were applied due to the co-segregation of the variant with the disease-phenotype, suggesting a positive family history.

Mutalyzer 2.0.32 was used to predict the effect of the c.23_34del in-frame variant on the FBN1 protein, and four amino acids (Glu-Ile-Ala-Leu) were predicted to be deleted due to this 12bp deletion, thus changing the length of the protein (ACMG-AMP criteria code PM4). The deletion of the four amino acids as predicted by Mutalyzer is indicated in Figure 3.19. This variant has not been reported in ClinVar, HGMD or UMD-FBN1. It is also not reported in control databases such as gnomAD, ExAC and 1000 Genomes Project (ACMG-AMP criteria code PM2). This variant affects the signal peptide domain, which is critical to the fibrillin-1 protein (ACMG-AMP criteria code PM1) (Baudhuin *et al.,* 2019). Therefore, this in-frame deletion was classified as likely pathogenic.



Figure 3.19. A screenshot from Mutalyzer 2.0.32 showing the reference protein and the affected amino acids (EIAL: Glu-Ile-Ala-Leu) as a result of the 12bp in-frame deletion in the *FBN1* gene.

The *FBN1* c.23_34del in-frame deletion variant co-segregated with the Marfan phenotype in this family, suggesting it may be disease-causing, hence, coverage of the variant was inspected using IGV and the IGV image is shown in Figure 3.20. Due to time constraints, the variant was not validated by Sanger sequencing.



Figure 3.20. These IGV images depict the coverage of the *FBN1* heterozygous in-frame deletion; c.23_34del, in patients HCTDs_14 (A) and HCTDs_13 (B) as reported in VEP. This variant occurs in exon 2 of 66.

3.3.1.7 Patient HCTDs_32

HCTDs_32 had two variants considered as possibly disease-causing after the filtering process (Table 3.10). Patient HCTDs_32 is a female with multiple tall family members, has a mitral valve prolapse, an unspecified hip problem and is suspected to have MFS.

Table 3.10. Shortlisted variants identified in HCTDs_32 post variant filtering.

Gene	Variant	Functional Domain/ Conserved Region	*In Silico* tools predicting a Benign effect	*In Silico* tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
COL1A1	c.2721A>C (p.Lys907Asn)	Triple-helical region		PolyPhen2, CADD, MetaLR, REVEL, FATHMM, DANN,				PM2, PP2, PP3	VUS
MYH11	c.3672+5T>G	Not strongly conserved	DANN					PM2, BP4	VUS

NGS sequencing detected the heterozygous *COL1A1*(NM_000088.4): c.2721A>C (p.Lys907Asn) missense variant in exon 39 of the gene. This variant has not been reported in the mutation database HGMD. It is also absent from control databases (ACMG-AMP criteria code PM2). Five *in silico* tools (PolyPhen2, CADD, MetaLR, REVEL and FATHMM) are in agreement that the variant has a damaging effect to the gene and its product (ACMG-AMP criteria code PP3) (Richards *et al.,* 2015). The gnomAD *COL1A1* missense z-score is 3.53, indicating that missense variants in this gene are a common mechanism for disease development (ACMG-AMP criteria code PP2) (Richards *et al.,* 2015). Thus, this variant was classified as a VUS.

The heterozygous intron *MYH11*(NM_001040114.1): c.3672+5T>G variant was also identified in this patient. This variant has been also identified in patients HCTDs_24 and HCTDs_26 in this study cohort, however it was excluded from their shortlisted variants subsequent to employing the family-based variant filtering approach. The variant was classified as a VUS owing to the application of the ACMG-AMP criteria codes BP4 and PM2 due to the benign prediction from DANN irrespective of its absence from the gnomAD database.

3.3.1.8 Patient HCTDs_31

One candidate variant passed the broad variant filtering steps employed in this study (Table 3.11).

Table 3.11. The candidate variant identified in the *COL3A1* gene of patient HCTDs_31.

Gene	Variant	Functional Domain/ Conserved Region	*In Silico* tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
COL3A1	c.2553+5G>A		dbscSNV RF, Ada				PM2, PP3, PP4	VUS

Patient HCTDs_31 was a female with multiple torn ligaments, bruises and abnormal scarring with a suspected diagnosis of vEDS. Molecular analysis detected the *COL3A1* (NM_000090.4): c.2553+5G>A splice variant in intron 36 of the *COL3A1* gene. This variant is absent from control databases (ACMG-AMP criteria code PM2), ClinVar, HGMD and COL3A1-LOVD. The *in silico* tool dbscSNV RF and Ada scores are in agreement that there is an increased probability that this variant affects splicing (ACMG-AMP criteria code PP3). The patient's phenotype was highly specific for vEDS (ACMG-AMP criteria code PP4) (Richards *et al.,* 2015). There was no other evidence supporting pathogenicity, and this is why the variant was therefore classified as a VUS.

3.3.1.9 Patient HCTDs_28

One variant was considered to be a candidate in this patient (Table 3.12). Patient HCTDs_28 is a female with dysmorphic features, global developmental delay, joint laxity and severe scoliosis. Her clinical features were suggestive of a HCTD, but the clinical team was unable to define a specific HCTD.

Table 3.12. Table showing the identified candidate variant in patient HCTDs_28.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tools predicting a Benign effect	In Silico tools predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
COL3A1	c.798+38C>A	Not strongly conserved region	DANN			rs189627079		PM2, BP4	VUS

The heterozygous *COL3A1*(NM_000090.4): c.798+38C>A variant was identified in this patient. It was absent from control databases (ACMG-AMP criteria code PM2), but without pathogenic predictions from the *in silico* tools. The computational tool DANN predicted a benign effect (ACMG-AMP criteria code BP4). We therefore classified this variant as a VUS.

3.3.1.10 Patient HCTDs_27

One variant remained after our variant filtering pipeline for this patient (Table 3.13). Patient HCTDs_27, with a suspected diagnosis of MFS presented with minor dysmorphic features and a mitral valve prolapse.

Table 3.13. The variant identified in patient HCTDs_27.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tool predicting a Benign effect	In Silico tool predicting a pathogenic effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
CBS	c.1223+67_1223+68insG		GERP					PM2, BP4	VUS

The *CBS* (NM_001178008.2): c.1223+67_1223+68insG heterozygous intron variant was identified in this patient. It was classified as a VUS, because of the benign prediction from GERP (ACMG-AMP criteria code BP4) and its absence from control databases (ACMG-AMP criteria codes PM2).

3.3.1.11 Patient HCTDs_30

Only one candidate variant was further investigated for pathogenicity after the filtering procedure (Table 3.14).

Table 3.14. The identified variant in patient HCTDs_30 detected in the *COL1A1* gene.

Gene	Variant	Functional Domain/ Conserved Region	In Silico tool predicting a benign effect	ClinVar	dbSNP	HGMD	ACMG-AMP codes applied	ACMG-AMP classification
COL1A1	c.2451+61G>C		DANN		rs372327842		PM2, BP4	VUS

Patient HCTDs_30 is a male with multiple joint and limb contractures with a suspected diagnosis of CCA. The *COL1A1*(NM_000088.4): c.2451+61G>C heterozygous intron variant, occurring 61bp from the exon-intron boundary was identified in HCTDs_30. The *in silico* tool DANN predicted a non-damaging effect (ACMG-AMP criteria code BP4) even though the variant was absent from control databases (ACMG-AMP criteria code PM2). The application of these ACMG-AMP criteria codes supported a VUS classification.

3.3.1.12 Patient HCTDs_09

Consequent to variant prioritization and filtering based on the MAF, variant data obtained from disease and mutation databases such as ClinVar, HGMD, universal mutation and protein databases, no candidate variants were described for this patient. A manual inspection of the BAM file using IGV also did not reveal any candidate variants.

3.4 Diagnostic Yield

Pathogenic or likely pathogenic variants were identified in 9/15 (60%) of the patients studied in this cohort. However, for the purposes of the diagnostic yield calculation, each family was calculated as a single entity, yielding a diagnostic rate of 50% (6/12).

Table 3.15. Table with the (likely) pathogenic variants in the *FBN1* gene identified in our study cohort.

Patient	Suspected/ Clinical Diagnosis	Gene	Variant	Amino Acid change	Zygosity	Mutation Type	Sex	Previously Reported	ACMG-AMP Classification
HCTDs_13	MFS	*FBN1*	c.23_34del	p.Glu8_Leu11del	Heterozygous	In-frame deletion	Female		Likely Pathogenic
HCTDs_14	MFS	*FBN1*	c.23_34del	p.Glu8_Leu11del	Heterozygous	In-frame deletion	Female		Likely Pathogenic
HCTDS_21	MFS/vEDS	*FBN1*	c.6670dupA	p.Thr2224AsnfsTer6	Heterozygous	Frameshift	Male		Pathogenic
HCTDS_22	MFS (Meets Ghent Criteria)	*FBN1*	c.640G>A	p.Gly214Ser	Heterozygous	Missense	Female	ClinVar, HGMD	Pathogenic
HCTDs_23	MFS	*FBN1*	c.3794G>A	p.Cys1265Tyr	Heterozygous	Missense	Female	HGMD	Pathogenic
HCTDS_24	MFS (Meets Ghent Criteria)	*FBN1*	c.5726T>C	p.Ile1909Thr	Heterozygous	Missense	Female	ClinVar, HGMD	Likely Pathogenic

HCTDS_25	MFS (Meets Ghent Criteria)	*FBN1*	c.5726T>C	p.Ile1909Thr	Heterozygous	Missense	Male	ClinVar, HGMD	Likely Pathogenic
HCTDS_26	MFS (Meets Ghent Criteria)	*FBN1*	c.5726T>C	p.Ile1909Thr	Heterozygous	Missense	Female	ClinVar, HGMD	Likely Pathogenic
HCTDS_29	MFS (Meets Ghent Criteria)	*FBN1*	c.3037G>A	p.Gly1013Arg	Heterozygous	Missense	Male	ClinVar, HGMD	Pathogenic

4. DISCUSSION

4.1 Project Rationale

Inherited disorders of the connective tissues are a heterogeneous group of disorders affecting one or more components of the connective tissue. They are caused by pathogenic variants in the genes encoding the proteins in the ECM, or the proteins involved in their post-translation modification. MFS, ELS, CCA, SGS, vEDS, LDS and homocystinuria were the HCTDs of interest to the present study as they are the ones identified by clinicians to be the most frequently observed in genetic clinics hosted by the Division of Human Genetics.

HCTDs affect multiple systems owing to the pleiotropic nature of the implicated genes with a wide phenotypic spectrum and extensive clinical overlap between the different HCTDs. As a result, making a definitive diagnosis is difficult. An additional factor that can complicate the diagnosis of HCTDs is the age-dependent nature of the physical manifestations of HCTDs, which is particularly relevant to MFS (Ergoren *et al.,* 2019). In South Africa, HCTDs are diagnosed using the available international clinical diagnostic criteria. It is however uncertain if the diagnostic criteria are directly applicable to the South African population (Child *et al.,* 2007) due to the wide phenotypic heterogeneity associated with HCTDs. Moreover, certain diagnostic criteria, and in particular the Ghent criteria only applies to individuals of a certain age group (>20 years) and may not be necessarily specific to patients younger than 20 years. In this regard, the revised Ghent nosology for diagnosing MFS (Loeys *et al.,* 2010) has given more weight to molecular diagnostic testing, aiding in the diagnosis of MFS.

To the best of our knowledge, a limited number of molecular studies have been undertaken to determine the genetic aetiology of HCTDs and to further improve their diagnostic yield in South Africa. However, in these cases, the molecular analysis was not performed in South Africa. For example, Child *et al*. (2007) reports nine South African cases with possible MFS seen at the Groote Schuur Hospital Genetic clinic, University of Cape Town, with the *FBN1* mutation analysis carried out at St Georges Hospital, London using conventional techniques.

Clinical HCTDs studies carried out in South Africa have reported on the phenotype of MFS cases (Naidoo *et al.,* 2018) and the use of the MFS diagnostic criteria to make a clinical diagnosis. The present study aimed to perform molecular screening on patients with HCTDs using local NGS infrastructure to generate knowledge that can be used to improve the diagnostic yield of inherited disorders of the connective tissues in a South African cohort.

4.2 Diagnostic Yield

4.2.1 HCTDs Diagnostic Yield

The reported overall diagnostic yield of HCTDs NGS panels worldwide is relatively low. A mutation detection rate of 17.1% was observed in German patients in study by Renner *et al*. (2019). Overwater *et al*. (2018) and Campens *et al*. (2015) reported a diagnostic yield of 8.1% and 13% respectively. The 8.1% diagnostic yield is the lowest compared to previous studies that have identified pathogenic variants in 10.3%-35.5% of the individuals with HCTDs (Overwater *et al.,* 2018). These studies were testing patients with features suggestive of HCTDs including MFS, LDS, EDS, CCA and TAAD. In the present study's patient cohort, 15 patients with a clinical or suspected diagnosis of one of the seven HCTDs of concern were invited to participate, in which 6 *FBN1* pathogenic or pathogenic variants were identified in 12 family entities, making our mutation detection rate 50%.

4.2.2 *FBN1* Diagnostic Yield

In different populations, *FBN1* diagnostic testing yields have been reported throughout the years (Table 4.1). The difference in the mutation detection rate when screening the *FBN1* gene across populations may be due to the difference in the clinical selection of patients. According to Radonic *et al*. (2011), a *FBN1* pathogenic variant is identified through gene sequencing in approximately 70-93% of individuals with a clinical diagnosis of MFS, diagnosed using the revised Ghent criteria. A study by Loeys *et al*. (2004) reported this figure to be 91% based on a large cohort while Yang *et al*. (2018) reports it as 90%.

Molecular analysis of the *FBN1* gene in Italian patients with clinical features suggestive of MFS reported a mutation detection rate of 81% (Attanasio *et al.,* 2008) and 86.3% (Mannucci *et al.,* 2019). Studies from Belgium and Korea performed molecular analysis in patients with classic MFS and identified *FBN1* pathogenic variants in 91% (Loeys *et al.,* 2004) and 89.7% (Yoo *et al.,* 2010) of their patient cohort, respectively. These studies used NGS and other technologies such as MLPA. According to Loeys *et al*. (2001), the high incidence of patients fulfilling the Ghent nosology in a patient cohort is a good predictor of the presence of pathogenic *FBN1* variants. Hence, the high mutation detection rate observed in these studies may be attributed to the increased number of patients (>80%) in their cohorts fulfilling the Ghent nosology.

The relatively low diagnostic yield often observed in alternative studies screening the *FBN1* gene may be attributed to the increased number of people who do not fulfil the Ghent nosology in their cohort and this may be because they have a Marfan-related phenotype such as MVP-MASS syndrome (Loeys *et al*., 2001). These are features on the HCTD continuum, but they do not fulfil the diagnostic criteria. This was demonstrated in Rommel *et al*. (2005) and Zhurayev *et al*. (2016), with their diagnostic yields reported as 33% and 67.5% in the German and Ukrainian populations, respectively. Gao *et al*. (2019) reported the overall diagnostic yield in patients with classic and incomplete MFS to be 92% and 11.1%, respectively. The yield observed in non-classic MFS patients in Gao *et al*. (2019) are in concordance with the observed 12% and 10% (likely) pathogenic variants identified in *FBN1,* in patients with a suspected diagnosis of MFS in other studies; Loeys *et al*. (2001) and Rommel *et al*. (2005). These figures give a better demonstration of the effect of the Ghent nosology in the *FBN1* diagnostic yield.

Table 4.1. Published mutation screening *FBN1* yields in different population groups.

Study Population	**Ghent Nosology**		**Overall *FBN1* mutation yield (%)**	**Reference**
	Classic MFS patients	Non- classic patients		
Australia	49%	51%	86%	Biggin *et al.,* 2004
Belgium	100%	0%	91%	Loeys *et al.,* 2004
China	57%	43%	57%	Gao *et al.,* 2019
Italy	81%	19%	86.3%	Mannucci *et al.,* 2019
Italy	86%	14%	81%	Attanasio *et al.,* 2008
Korea	100%	0%	89.7%	Yoo *et al.,* 2010
Ukraine	70%	30%	67.5%	Zhurayev *et al.,* 2016
Study population	No. of Patients with *FBN1* pathogenic variants		Total number of patients with *FBN1* pathogenic variants	Reference
	No. of Classic MFS patients with a *FBN1* pathogenic variant	No. of Non-classic patients with a *FBN1* pathogenic variant		
China	11	1	12	Gao *et al.,* 2019
Germany	37	24	61	Rommel *et al.,* 2005
Korea	22	14	36	Seo *et al.,* 2018

4.2.3 *FBN1* Mutation-Detection Rate in Present Cohort

Of the 15 patients in this study, 11/15 (73%) had a clinical or suspected diagnosis of MFS, with 5 of these patients fulfilling the Ghent criteria (45%) at the time of the initial clinical visit (Loeys *et al.,* 2010). One patient had a differential diagnosis of MFS or vEDS and the other 3 had suspected diagnoses of CCA, vEDS and a non-specific HCTD. All the identified pathogenic and likely pathogenic variants in this study were in the *FBN1* gene, implicated in the causation of MFS, MVP and MASS (Murphy-Ryan Psychogios and Lindor, 2010). Sixty-percent (9/15) of our patient cohort are young individuals (<18 years), which may explain why only 5 patients met the Ghent criteria for a clinical diagnosis at the time of the initial clinical visit.

Our overall *FBN1* mutation detection rate was 60% (9/15) and may be accounted for as 2 and 3 members from the HCTDs_143 and HCTDs_2456 families were tested respectively, in which the same *FBN1* likely pathogenic variant was identified in each family member. Our mutation detection rate was therefore 50% (6/12). Of the 5 patients fulfilling the Ghent criteria, a pathogenic or likely pathogenic variant was identified in 100% (5/5) of them, while in the remaining 10 patients without a clinical diagnosis of MFS, a molecular diagnosis was achieved in 40% (4/10). Of the six mutation-negative patients and not fulfilling the Ghent criteria, four were <18 years old.

4.3 Identified Putative Disease-Causing Variants

Six different pathogenic or likely pathogenic variants were identified in nine probands of this study cohort, four (4) pathogenic and two (2) likely pathogenic. Four (66%) of the variants were missense, one was a frameshift and there was one in-frame deletion (indicated in Figure 4.1). Of the six putative disease-causing variants, four had been previously reported in other studies and two were reported for the first time in this study.

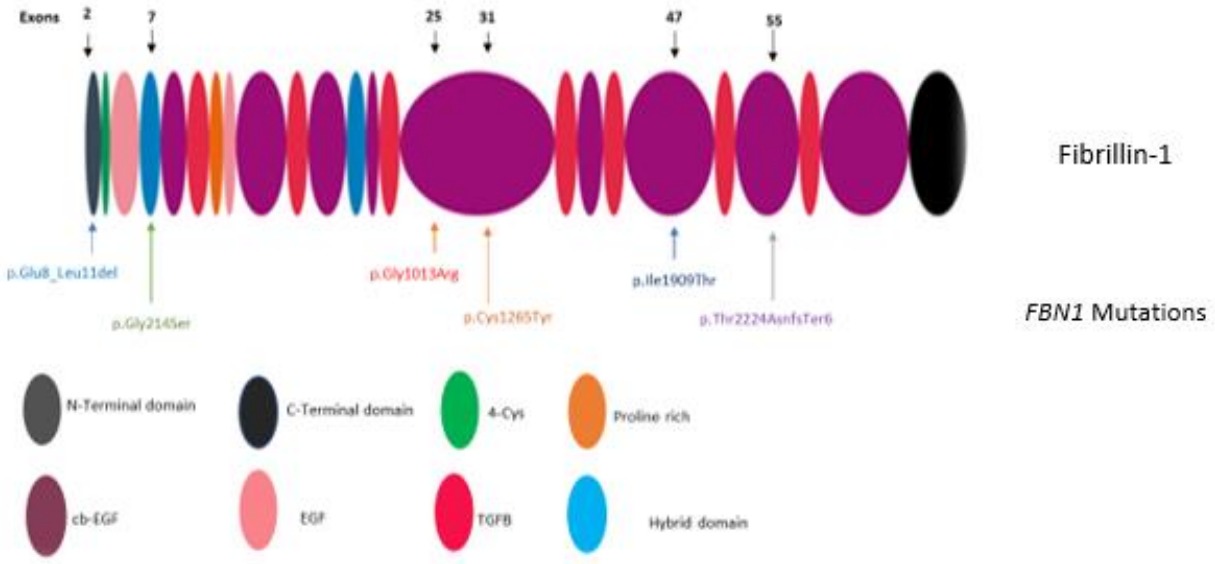


Figure 4.1. A diagram showing the distribution of the *FBN1* likely pathogenic or pathogenic variants identified in the current study. The fibrillin1 protein has the epidermal growth factor (EGF), calcium-binding epidermal growth factor (cb-EGF), transforming growth factorβ binding protein (TGFβ) and the hybrid domains, which are crucial for the structural integrity of the protein. The cb-EGF-like domain of the *FBN1* gene also protects the fibrllin1 protein from proteolytic degradation. The N- and C-terminal regions are required for the secretion of the fibrillin1 protein from the cells and microfibril assembly. Exons 24-32 span the "neonatal region" of the *FBN1* gene and mutations in this region are associated with a severe MFS phenotype.

4.3.1 *FBN1* Missense Variants

4.3.1.1 *FBN1*: c.5726T>C (p.Ile1909Thr)

The heterozygous c.5726T>C variant (p.Ile1909Thr), in exon 47 of the *FBN1* gene was identified in 3 related patients in this cohort; HCTDs_24, HCTDs_25 and HCTDs_26. HCTDs_25 and HCTDs_26 are children to HCTDs_24. HCTDs_24 is an African female who presented with ectopia lentis and cardiovascular complications. She also had a systemic score of 7 and fulfilled the Ghent criteria for MFS (Loeys *et al.,* 2010).

Her two children were born from different fathers and both had a systemic score>7. HCTDs_25 is a hyperactive 12-year-old male with ectopia lentis, myopia, mild scoliosis and aortic root dilatation. He also fulfilled the Ghent criteria for MFS (Loeys *et al.,* 2010), while HCTDs_26 is a female of 7 years also with ectopia lentis, scoliosis and pectus deformity, also fulfilling the Ghent nosology (Loeys *et al.,* 2010). The co-segregation of *FBN1* variants with the disease phenotype was previously considered supporting evidence of variant pathogenicity (Richards *et al.,* 2015), however it is now a moderate criterion supportive of variant pathogenicity, dependent on the number of MFS affected family members harbouring the putative disease-causing variant (Muiño-Mosquera *et al.,* 2018). More weight supporting variant pathogenicity was given to the c.5726T>C variant as there are more than 2 individuals with a MFS phenotype in the family, all harbouring the *FBN1* mutation.

The *FBN1* c.5726T>C variant is absent from control databases (gnomAD, ExAC and 1000 Genomes Project) and had a likely pathogenic classification in ClinVar (VCV000200189). It was also reported in HGMD (CM013934) to be associated with MFS. An alternative variant, c.5726T>G altering the same amino acid residue, however substituting isoleucine with serine was also reported in a patient with MFS, indicating that this particular amino acid residue may be clinically significant. Multiple computational pathogenicity predictive tools have also produced pathogenic predictions, owing to the reason that the identified variant affected a functional domain. Baudhuin *et al*. (2019) described the affected cb-EGF like domain as a crucial residue for protein function. The c.5726T>C variant has been frequently observed in MFS patients from different populations (Loeys *et al.,* 2001, Loeys *et al.,* 2004, Wang *et al.,* 2013, Baetens *et al.,* 2011). Wang *et al*. (2013) reported this variant to be associated

with major involvement of the cardiac and skeletal systems, as well as ectopia lentis, in accordance with what we observed in this family.

4.3.1.2 *FBN1*: c.3794G>A (p.Cys1265Tyr)

A heterozygous *FBN1* c.3794G>A (p.Cys1265Tyr) variant was detected in exon 30 of patient HCTDs_23. The 37-year-old is an African female with a suspected diagnosis of MFS with a systemic score of 8, but she did not fulfil the Ghent criteria. She also had dysmorphic features, a tall stature, eye problems requiring surgery and aortic complications necessitating multiple hospital admissions from childhood. The identified variant was not present in control databases (gnomAD, 1000 Genomes Project and ExAC) or in ClinVar, but was predicted to have a pathogenic effect by multiple computational pathogenicity predicting tools.

Exon 30 is encompassed in the “neonatal region” of the *FBN1* gene, spanning exons 24-32, which is associated with a severe phenotype of MFS (Stark *et al.,* 2020). Patients with neonatal MFS are usually diagnosed early and have aortic complications within their first year of life (Liu *et al.,* 2018). The encoded fibrillin1 protein is cysteine-rich (Robinson and Godfrey, 2000) and the c.3794G>A variant affects a cysteine residue which is important for the appropriate structural formation of the protein. It has been widely noted that variants either altering or creating cysteine residues are known to be the frequent cause of a Marfan-phenotype (Schrijver *et al.,* 1999).

Over 20 years ago, Montgomery *et al*. (1998) detected the c.3793T>C (p.Cys1265Arg) (rs137854474) *FBN1* cysteine substitution in a severely affected family with MFS. Other *FBN1* missense variants substituting this cysteine have been discovered ever since, including the c.3793T>G (p.Cys1265Gly) (CM098683) (Stheneur *et al.,* 2009), c.3794G>C (p.Cys1265Ser) (CM154832) (Baudhuin *et al.,* 2015) and c.3795T>A (p.Cys1265Ter) (rs1555398278*,* CM164448) (Latasiewicz *et al.,* 2016). Alternative cysteine variants have also been reported as pathogenic or likely pathogenic in the UniProtKB database. These include the c.3793_3795delTGTinsAGG (p.Cys1265Arg) and c.3793_3795delTGTinsAGA (p.Cys1265Arg) variants.

The cysteine substitution identified in our patient also affects a critical domain, cb-EGF, and according to the UMD-FBN1 database, cysteine substitutions in the cb-EGF like-domain largely represent pathogenic missense variants implicated in MFS (Collod-Beroud *et al.,* 2003). According to Baudhuin *et al*. (2019), an automatic likely pathogenic classification should be given to *FBN1* variants altering a cysteine residue in either the cb-EGF or the TB domains.

The c.3794G>A variant, detected in HCTDs_23 had been seen in the literature before in a 5-year-old Italian male with severe clinical features; aortic dilatation, mitral valve prolapse, pectus carinatum and ectopia lentis, with a clinical diagnosis of MFS (Attanasio *et al.,* 2008). These clinical features are representative of a severe MFS phenotype. Our patient had similar features. Patient HCTDs_23 also had childhood onset of clinical features suggesting a HCTD, such as poor vision and hypermobility, however she was only diagnosed with MFS as an adult. *FBN1* mutation screening has a role in the early diagnosis of patients with MFS. Patient HCTDs_23 could have benefitted from an early molecular diagnosis, facilitating timely medical, surgical and/or prophylactic intervention to prevent aortic dilatations and eventually ruptures (Yip and Sawatzky, 2014).

4.3.1.3 *FBN1*: c.640G>A (p.Gly214Ser)

The c.640G>A (p.Gly214Ser) heterozygous variant was detected in one patient in this cohort; HCTDs_22. HCTDs_22 is a 11-year-old African female, born from unaffected parents and was clinically diagnosed with MFS. She had a mild mitral valve prolapse, aortic dilatation and ectopia lentis and she fulfilled the Ghent criteria (Loeys *et al.,* 2010). The variant identified affected the hybrid-1 domain, in exon 7 and was predicted to be deleterious by various *in silico* tools. An equivalent pathogenic *FBN1* c.640_641delGGinsTC (p.Gly214Ser) variant was reported in UniProtKB, resulting in the same amino acid change, supporting variant pathogenicity.

The identified variant was not reported in control databases (gnomAD, 1000 Genomes Project and ExAC) but was reported in ClinVar (VCV000199956.6) as pathogenic and was also observed frequently in the literature in patients with MFS (Attanasio *et al.,* 2008, Dong *et al.,* 2012, Yang *et al.,* 2016, Madar *et al.,* 2019, Renner *et al.,* 2019). The variant has been reported in different populations in unrelated patients with MFS

(Attanasio *et al.,* 2008, Yang *et al.,* 2016, Renner *et al.,* 2019). Attanasio *et al*. (2008) reports the clinical features of a female Italian patient with this variant and with a diagnosis of MFS amongst others as; aortic dilatation, mitral valve prolapse and ectopia lentis. Renner *et al*. (2019) reported this variant in a patient suspected to have TAAD and it was classified as disease-causing.

This variant had also been reported in a Chinese family with ELS where all the affected members had the variant while the healthy individuals, as well as the 50 controls did not (Sui *et al.,* 2004). Another Chinese family conversely with MFS also had this variant (Dong *et al.,* 2012). All the healthy family members and the 100 controls did not have the variant. The Sui *et al*. (2004) and Dong *et al*. (2012) studies suggest co-segregation of the variant with affected family members, which is supporting evidence for the variant pathogenicity classification. The difference in the variant associated phenotype between the two studies on the Chinese families may be because of the median age of the patients. The median age for the family with MFS reported by Dong *et al*. (2012) is 37 years, while the family reported by Sui *et al*. (2004) is relatively young. The physical manifestations of MFS become more prominent with age, hence younger individuals may not display some clinical features at a certain age, and as a result, they may lack an accurate clinical diagnosis.

4.3.1.4 *FBN1*: c.3037G>A (p.Gly1013Arg)

The heterozygous missense c.3037G>A (p.Gly1013Arg) (rs140593) variant was found in a 9-year old African male; HCTDs_29, clinically diagnosed with MFS. He presented with ectopia lentis, myopia, aortic dilatation, developmental delay and had a systemic score >7. He met the Ghent nosology at age 7, suggestive of a severe phenotype.

The severe phenotype at a young age in our patient can be attributed to the location of the identified *FBN1* variant. The variant was located in exon 25 of the “neonatal region”, a well-established *FBN1* hotspot associated with a more severe phenotype of MFS. This variant is absent in control databases such as gnomAD and 1000 Genomes Project, however it was reported in the mutation and disease databases; HGMD (CM950441) and ClinVar (VCV000177648), respectively prior to it being identified in our cohort, with a pathogenic classification. The c.3037G>A variant is in a functional domain, cb-EGF, and was predicted to be damaging by multiple *in silico* tools.

This variant is among the frequently reported *FBN1* pathogenic variants known to cause MFS (Rommel *et al.,* 2005). It has been reported repeatedly over the years in literature and comparison of the clinical manifestations of the unrelated patients with this variant is consistent (Loeys *et al.,* 2001, Biggin *et al.,* 2004, Tiecke *et al.,* 2001, Rommel *et al.,* 2005). According to Madar *et al*. (2019), the c.3037G>A variant is considered a minor hotspot for severe clinical features, with early onset. The severe phenotype associated with this variant observed in the other patients was also observed in our patient, however our patient did not have neonatal MFS.

4.3.2 *FBN1* Frameshift Variant

4.3.2.1 *FBN1*: c.6670dupA (p.Thr2224AsnfsTer6)

The *FBN1* frameshift c.6670dupA (p.Thr2224AsnfsTer6) heterozygous variant, classified as pathogenic was identified in exon 55 of a 32-year old male of mixed ancestry (HCTDs_21) presenting with joint hyperlaxity. He did not present with significant clinical features in the ocular system, however he had a systemic score of 7. At the time of his initial clinic visit, he did not meet the Ghent criteria for MFS, and at the time, the differential diagnoses included MFS and vEDS due to his aortic and mitral valve lesions, and excessive scarring. He had an aortic aneurysm repair subsequently. Thus, at age 30, he then presented with aortic dilatation, thus fulfilling the Ghent criteria. The variant identified in HCTDs_21 is absent from gnomAD, 1000 Genomes Project and ExAC, as well as mutation and disease databases (ClinVar, HGMD, UMD-FBN1), suggesting this is a novel mutation.

The identified variant is a duplication of the A nucleotide at position c.6670, resulting in the amino acid change threonine to asparagine, consequently altering the reading frame of the *FBN1* gene. Subsequently, the change in the *FBN1* translational reading frame introduced a premature termination codon, terminating at codon six (Ter6) and as a result the encoded protein was significantly shortened. Haploinsufficiency of *FBN1* mutations is one of the mechanisms underlying MFS (Stengl *et al.,* 2020), hence, the resulting truncated protein, consequent to the introduction of the stop-codon by the c.6670dupA variant is very strong evidence of variant pathogenicity.

A repeated correlation has been reported between *FBN1* frameshift truncating variants and aortic dissection (Baudhuin *et al.,* 2015, Becerra-Munoz *et al.,* 2018, Xu *et al.,* 2020). This is because these variants have a more deleterious effect on the fibrillin1 protein, resulting in decreased elastic fibres and a weak ECM. This is best illustrated by the staining analysis of specimens obtained from patients who underwent aortic surgery, which revealed ruptured elastic fibres and fibrosis in patients with frameshift and nonsense mutations (Xu *et al.,* 2020). The clinical features of our patient overlap that of patients with vEDS, however, the variant identified is supportive of a MFS diagnosis.

4.3.3 *FBN1* In-frame Deletion Variant

4.3.3.1 *FBN1*: c.23_34del (p.Glu8_Leu11del)

The heterozygous in-frame c.23_34del (p.Glu8_Leu11del) *FBN1* variant was identified in 2 related patients in our patient cohort, HCTDs_14 and HCTDs_13, both suspected to have MFS. HCTDs_14 is the mother to HCTDs_13. HCTDs_14 is a 33-year-old African female with mitral valve prolapse and a systemic score>7. Her daughter, HCTDs_13, is a tall 5-year-old with a reduced upper segment to lower segment ratio, and a systemic score of 3. Both patients did not fulfil the Ghent criteria for MFS, however, the co-segregation of the same likely pathogenic variant in the two patients with clinical features suggestive of MFS is evidence confirming the diagnosis of MFS.

An in-frame deletion is a multiple base deletion where the deleted bases are a multiple of three. In-frame mutations are dominant negative variants that maintain the gene's reading frame, however, alter the encoded protein (Stark *et al.,* 2020). Four amino acids, glutamic acid (E), isoleucine (I), alanine (A) and leucine (L), were deleted as a result of the deletion of the 12 bases from nucleotide c.23 to nucleotide c.34 in exon 2 of the *FBN1* gene. *FBN1* deletion mutations have been associated with a severe MFS phenotype (Furtado *et al.,* 2011). Stheneur *et al*. (2011) reported 3 *FBN1* in-frame deletions where all MFS patients with an in-frame deletion died of cardiac complications before age 5. Stark *et al*. (2020) did not report on the clinical features of the 4 patients with *FBN1* in-frame deletions in his study.

The identified c.23_34del in-frame variant is absent from control databases, ClinVar, HGMD and UMD-FBN1, suggesting that it may be novel and has a positive score from GERP signifying that it occurs in a conserved region. The variant is located in the signal peptide. Signal peptides are short peptides of about 16-30 amino acids present in the N-terminal region of the fibrillin-1 protein (FUN). The FUN is a known functional domain in the *FBN1* encoded protein (Baudhuin *et al.,* 2019). The signal peptide is crucial for the translocation of the fibrillin-1 protein from the cell to the ECM and the assembly of microfibrils (Jacobs *et al.,* 2002). It is for all these reasons that this variant was assigned the likely pathogenic classification.

4.4 Mutation-Negative Cases

There were 6/15 (40%) patients suspected to have MFS, CCA, vEDS and an unspecified HCTD in whom a pathogenic or likely pathogenic variant was not identified. HCTDS_09 is a 16-year-old male with a suspected diagnosis of MFS. He is tall, has myopia and a systemic score of 7, while HCTDs_30 is a 3-year-old African male suspected to have CCA because of his joint contractures and mild developmental delay.

HCTDs_32 is a 12-year-old African female patient presenting with a positive wrist thumb sign, joint problems, arachnodactyly and mild scoliosis without a family history of any HCTD but has tall relatives. She also had a hip problem, a systemic score of 7 and was suspected to have MFS. HCTDs_27, an African male aged 22 years presented with minor dysmorphic features, a systemic score of 4 and a family history of unexplained sudden death. He was suspected to have MFS. Patients HCTDs_32 and HCTDs_27 both have a learning problem. HCTDs_28 is a 7-year-old African female with global developmental delay, joint laxity, dysmorphic features and severe scoliosis with a suspected diagnosis of a HCTD that could not be specifically classified.

HCTDs_31 was a female who had a lot of bruises, a thin skin, torn ligaments, fibrosis, and valve regurgitation. She died from aortic rupture and had already had at least 3 spontaneous arterial ruptures. A *COL3A1*: c.2553+5G>A variant was identified in HCTDs_31 which we classified as a VUS, although it was previously classified in a different laboratory as likely pathogenic without supporting evidence described. Her clinical features were strongly suggestive of vEDS, and 2 *in silico* tools predicted a

damaging effect, however, there was no additional data supporting variant pathogenicity.

The reasons that 40% of our patient cohort remained without a molecular diagnosis may perhaps be due to our panel design and/or a lack of deletion and duplication analysis. In the present study, only exonic regions and 5-bases past the exon-intron boundary were sequenced. Deep intronic and regulatory regions were not sequenced, as a result, the detection of pathogenic variants in these regions could not be achieved. HCTDs have vast genetic heterogeneity, and there is extensive genetic diversity between Africans and the well-studied Europeans (Campbell and Tishkoff, 2008). Some of the clinical features of HCTDs are frequently observed in specific ethnic groups. For example, EDS-joint hypermobility is more frequently observed in Africans than in Caucasians (Hakim and Grahame, 2003). However, the degree of the genetic contribution has not yet been established (Hakim *et al.,* 2004), it is thus to be expected that alternative, yet unknown, genes may be implicated in the causation of HCTDs in the African population and these were not included in our NGS-custom panel design.

Approximately 60% of *FBN1* mutations are missense (Zhang *et al.,* 2018, Ouyang *et al.,* 2019), 22% are small insertions and deletions, 11% are splicing mutations and 3% of patients with MFS have *FBN1* gross deletions (Stenson *et al.,* 2003, http://www.hgmd.cf.ac.uk/ac/all.php). The present study only focused on small variant (<50bp) analysis and did not look at copy number variations (CNVs). CNVs are large structural variations; deletions or duplications, often involving large stretches of DNA, often ≥50bp (Stengl *et al.,* 2020). Several other studies investigating HCTDs have achieved higher mutation detection rates by performing MLPA in mutation-negative patients for the detection of CNVs (Yoo *et al.,* 2010, Baetens *et al.,* 2011, Mannucci *et al.,* 2019, Stengl *et al.,* 2020).

4.5 Implications of *FBN1* Testing

MFS threatens multiple body systems and affected individuals have an average life expectancy of 32 years without treatment (Yang *et al.,* 2016). The Ghent nosology is not reliable in making a diagnosis of MFS in affected children (Loeys *et al.,* 2010) owing to the late onset of clinical features and as a result, most affected individuals remain clinically undiagnosed, as indicated in our patient cohort. Thus, the accurate diagnosis of MFS or any HCTD requires extensive knowledge of the clinical features, the differential diagnoses, and the genetic background (Hoffjan, 2012). This then indicates the importance of molecular screening of the *FBN1* gene to provide an early molecular diagnosis to achieve timely and accurate management of MFS.

4.5.1 Diagnostic and Predictive Testing

In our patient cohort, 11 patients had a suspected or clinical diagnosis of MFS, and (likely) pathogenic variants were identified in 8/11 (73%). The provisional diagnosis was confirmed in 5 patients who fulfilled the Ghent criteria for MFS. Patient HCTDs_21, with differential diagnoses of MFS or vEDS was found to harbour a *FBN1* pathogenic variant. Even though he did not fulfil the Ghent criteria at his initial clinical visit, he did at recruitment. In this case, the molecular result aided in achieving a definitive diagnosis of MFS. Thus, in patients who did not fulfil the Ghent criteria but a likely pathogenic or a pathogenic variant was identified, a diagnosis was established based on the molecular result.

Only 25% of *FBN1* mutations are *de novo* and 75% are inherited according to the literature (Dietz *et al.,* 2017), with only 10% of these mutations being recurrent. Identification of a *FBN1* mutation in a family means that asymptomatic individuals may benefit from pre-symptomatic testing. Additional individuals who may benefit from pre-symptomatic testing prior to the manifestation of clinical features are younger individuals who may be affected, but not yet present with a MFS phenotype due to the age-dependent nature of some physical manifestations of MFS. Molecular diagnostic testing is therefore important for the early diagnosis of younger MFS patients.

Early diagnosis of MFS even in adults is significant as physicians will be able to timely treat MFS-related complications, improving the life expectancy and quality of affected individuals (De Backer, 2009). The importance of an early diagnosis is further illustrated in a case report by Aquila *et al*. (2020) where a pregnant woman died after delivery unware that she was affected with MFS, with an echocardiogram indicating aortic dilatation and aortic dissection being the cause of death. Early diagnosis of MFS could have assisted in monitoring the patient for possibly arising aortic complications during the gestation period and provided appropriate management. For the reason that a majority of MFS mortality cases are related to cardiovascular complications, early diagnosis, therefore appropriate management could improve the prognosis of MFS patients (Ammash and Connolly, 2007). This is best demonstrated by the significant increase in the life expectancy of MFS patients over the years following the improvement of MFS treatment (Vanem *et al*., 2020).

Mutation analysis of the *FBN1* gene can also be performed for family planning decision making, and this is because individuals with MFS or those with a family history may also have pre-natal testing once a pathogenic variant has been identified in a family (Ades *et al.,* 2007). Preimplantation genetic diagnosis can be also performed after the identification of a *FBN1* pathogenic variant in a family (Kilpatrick *et al.,* 1996, Arsian-Kirchner *et al.,* 2010), affording couples the possibility of having unaffected offspring as only MFS-free embryos can be implanted. This is however an expensive procedure and is not easily available to many families.

4.5.2 *FBN1* Genotype-Phenotype Correlations

This importance of *FBN1* molecular analysis is further illustrated by the influence of the DNA variant on the patients' phenotype. The *FBN1* mutation identified in HCTDs_21 assisted in understanding the cause of his less distinctive and severe clinical features. Individuals with *FBN1* frameshift mutations have been reported to have an increased risk of aortic complications (Baudhuin *et al.,* 2015).

Aortic events are often the cause of MFS morbidity and mortality. This demonstrates that the type and location of the pathogenic variant has an impact on the severity of MFS. In a study by Xu *et al*. (2019), MFS patients with frameshift and truncating mutations were observed to have a higher incidence of aortic dissections than patients

with *FBN1* missense mutations. Knowledge of the pathogenic variant can thus assist in the identification of individuals who should be more closely monitored for aortic dissections, thereby enhancing personalized patient care (Arnaud *et al*., 2021). This can in turn improve the quality of life of MFS-affected individuals.

The *FBN1* gene has other genotype-phenotype correlations, a well-established one, as well as other reproducible correlations. The well-established genotype-phenotype correlation is the “neonatal region”, associated with neonatal or childhood MFS. Recurrent *FBN1* genotype-phenotype correlations reported in the literature include the association of truncating mutations with an increased risk of aortic dissections (Becerra-Muñoz *et al.,* 2018), a vEDS like phenotype as observed in one of our patients (HCTDs_21), and missense substitutions affecting cysteine residues in the cb-EGF-like domain with ectopia lentis (Grange *et al.,* 2020).

A sex difference in the cardiovascular phenotype has been described and this is termed the “Carter effect” (Grange *et al.,* 2020). This is a phenomenon explaining that males with MFS often have greater cardiovascular involvement than females, and for females to have cardiovascular complications, they must have multiple genetic modifiers inherited together with the pathogenic *FBN1* variant. Owing to the passing down of half of the genetic determinants to the next generation, this then means that males born from mothers with severe MFS should be closely monitored for severe clinical presentations (Grange *et al.,* 2020). This means that knowledge of the origin of the mutation (either paternal or maternal) could assist in the proper surveillance of affected patients. Molecular diagnostic testing can also then assist to provide prognostic information.

4.5.3 Genetic Counselling

FBN1 testing is important at different stages of life; pre-conception, pre-natal, post-natal and adult-life, and genetic counselling services are of significance at all stages. Genetic counselling is the branch of human genetics that provides psychological support to patients, gives counselling on the inheritance and implications of genetic diseases and guides patients in making informed decisions about genetic disorders in their families (Dietz *et al.,* 2017).

Receiving a genetic test result can be overwhelming for a patient. The role of genetic counselling is to communicate what a positive, negative and an inconclusive (VUS) result means for the patient and family members. They then offer guidance based on the nature of the result, while communicating the chances of family members being affected including the siblings and the proband's offspring (Aatre and Day, 2011).

In each pregnancy, the proband inevitably confers 50% risk for MFS to offspring for example. Genetic counselling will provide information on the likelihood of having an unaffected child and the recurrent risks if you are an affected person, with either an unaffected or an affected partner. Genetic counselling is important in clarifying that should a proband have a mild phenotype, that does not confer the same risk for their children and this may be attributed to genetic modifiers. In alternative cases, probands may manifest clinical features suggestive of MFS without a family history as a result of factors such as parental mosaicism (Martínez-Quintana *et al.,* 2017) or *de novo* mutations. This way, genetic counselling helps guide towards reproductive choices. In the case of prenatal diagnosis, it is also the role of genetic counselling to explain the psychological and financial implications of the diagnosis made, as well as give informed choices with regards to the status of the unborn baby.

MFS is inherited in an autosomal dominant manner and has a high penetrance, however clinical variability has been observed in families with MFS, even when they share the same *FBN1* pathogenic variant. *FBN1* predictive-testing is possible when a *FBN1* mutation has been identified in the proband. The parents and siblings of the proband can be tested to determine whether the identified *FBN1* mutation occurred *de novo* or it was inherited. This can also help establish if there is a family history of MFS. If the parents of the probands have the *FBN1* pathogenic variant, there is a 50% chance that the siblings of the proband also have MFS, however the disease severity cannot be determined.

In most families, a disease that is not thoroughly understood may often cause confusion in the family. This is best illustrated by Mclnerney-Leo *et al*. (2020) where people relayed that they thought causal contributions to MFS included having the blue eye-colour trait, being a male individual and the birth order rather than the genetic contribution. A genetic test result can therefore help parents understand the reason for their children's illness and thereby explain it to their relatives.

4.5.4 Management

Different HCTDs have different risk profiles and they are managed differently (Loeys *et al.,* 2010). This highlights the importance of an accurate diagnosis to provide appropriate management. An accurate diagnosis ultimately allows physicians to provide the best patient care. Precision medicine is the branch of medicine concerned with identifying certain traits of an individual related to health such as biochemical and genetic markers. The identification of these traits assists in moving from treating a group of individuals with the same phenotype with the same pharmacotherapy to personalized treatment to achieve optimal patient care and minimal risk (Leopold and Loscalzo, 2018).

A correlation between response to losartan therapy and the type of *FBN1* mutation has been reported. Losartan is an angiotensin receptor blocker used to reduce the stress in the aortic walls of patients with MFS, thereby preventing aortic root dilatation (Elbadawi *et al.,* 2019). According to Franken *et al*. (2015) and Groenink and Mulder (2016), losartan minimized the rate of aortic dilations in patients with *FBN1* haploinsufficient mutations as opposed to those with dominant negative mutations. This is further indicative of the significance of knowledge of a molecular profile to apply personalized treatment towards achieving efficient management of the cardiovascular phenotype of MFS.

4.6 Limitations of the Study

We set out to host special HCTDs clinics to recruit patients with any of the 7 different HCTDs: MFS, ELS, CCA, SGS, vEDS, LDS and homocystinuria. However, due to the Covid-19 outbreak in 2020, we had to restrict recruitment to 15 patients who had already been seen by the doctors in the Division of Human Genetics. Hence, 12/15 (80%) of the patients had a clinical or suspected diagnosis of MFS, while the remaining 3/15 (20%) of our patient cohort had either CCA, vEDS or an undefined HCTD.

As a result, we were restricted in diversifying the patient cohort for a better representation of the frequently seen HCTDs in our genetic clinics. Our limited sample size was also a limitation of the current study because it may be difficult to make extrapolations (Faber and Fonseca, 2014) on the applicability of a HCTD NGS gene

panel in improving the diagnostic yield to the overall African population based on our data.

Due to the use of the IDP panel, the *ADAMTSL4* and *TGFβ2* genes implicated in autosomal recessive ELS and LDS, respectively were not tested. The omission of these genes may have contributed to the number of mutation-negative cases observed in the study. In a large cohort of patients with Marfan-related phenotypes, disease-casing variants have been identified in the *FBN1* gene, and in genes implicated in one of the differential diagnoses of MFS, LDS for an example (Bombardieri *et al*., 2020). In a study by Stengl *et al.* (2020), 84 pathogenic variants were identified in patients with a Marfanoid habitus. Two of these were disease-causing variants in the *TGFβ2* gene. Thus, it is recommended that a gene panel, rather than single gene testing for patients with a phenotype suggestive of MFS is should be used. The contribution of *TGFβ2* in a South African cohort, unfortunately, remains unknown.

Interpreting intronic variants is difficult owing to insufficient functional information (Dos Santos *et al.,* 2018), as a result all rare variants in non-coding regions identified in this study were regarded as VUSs in the absence of additional published information on the particular variant. This may also be due to the limited splice site pathogenicity predicting computational tools to predict the effect of intronic variants on splicing. Wu *et al*. (2021) reported a MFS patient with a pathogenic *FBN1*-intronic mutation further highlighting the importance of functional studies on *FBN1*-intron variation for accurate variant classification. The African population is understudied in genetics research (Tucci and Akey, 2019), as a result, there is insufficient genetic information on Africans. This may be an additional reason why all of the intronic variants were either classified as VUSs or (likely) benign. This points towards the significance of sharing African molecular genetic data on public databases to improve classification of African variants.

An additional limitation to the present study is the inability to validate the *FBN1* c.23_34del variant, classified as likely pathogenic due to time restrictions. It is therefore recommended that Sanger sequencing is carried out to confirm the variant.

4.7 Future Studies

For optimizing *FBN1* pathogenic variant detection, the use of a secondary technology such as MLPA in mutation-negative cases to screen for possible deletions and duplications is recommended (Stengl *et al.,* 2020). For repeated mutation-negative cases, whole exome or whole genome sequencing can be performed to improve mutation detection. This may also aid in the identification of novel genes implicated in the causation of HCTDs in the African population.

4.8 Conclusion

HCTDs are one of the most difficult groups of disorders to diagnose definitively. At the time of clinical presentation, most HCTDs cannot be specifically classified. In South Africa, HCTDs are diagnosed using clinical assessments and clinical diagnostic criteria, where appropriate. This is a problem when a large portion of South Africa struggle to get access to specialist medical care. Molecular diagnostic testing is not available for a majority of our patients. There is also not enough literature on the genetics of HCTDs in the African population.

Early diagnosis of HCTDs is significant because it may decrease health expenditures (Grosse *et al.,* 2010) particularly in a low-resource country like South Africa. Early and accurate diagnosis of a HCTD is also imperative because it enables efficient management of clinical manifestations by physicians and the identification of high-risk individuals (Murphy-Ryan, Psychogios and Lindor, 2010). For example, an affected pregnant woman is at a high risk of arterial dissection either during or after birth (Jessurun *et al.,* 2016) as well as individuals with *FBN1* frameshift or truncating mutations (Baudhuin *et al.,* 2015, Xu *et al.,* 2019), hence an accurate diagnosis will afford affected individuals personalized medicine facilitating optimum patient care.

A targeted NGS-custom panel was designed and used with an aim of providing a molecular diagnosis to 15 patients with HCTDs in this study. To the best of our knowledge, the current study is the one of the few molecular studies in South Africa to perform mutation screening on patients with HCTDs and it reports a high mutation detection rate (50%). We have established several benefits of molecular diagnostic testing when screening for HCTDs in the African population and inferred that mutation

screening of the *FBN1* gene is an appropriate diagnostic approach in the South African population in patients with a phenotype suggestive of MFS.

This is with regard to the current study being able to give a molecular diagnosis to patients who had an unclear clinical diagnosis. Given that a definitive diagnosis is even more difficult to reach in younger individuals due to the age-dependent physical manifestations of MFS, they were also given a molecular diagnosis. The current study also contributed two novel pathogenic *FBN1* variants to the mutation spectrum of MFS. The present study has provided important baseline data for the establishment of molecular testing for MFS, which can be expanded to other HCTDs.

REFERENCES

1000 Genomes Project Consortium, (2015). A global reference for human genetic variation. *Nature*, *526*(7571):68-74.

Aatre, R.D. and Day, S.M., (2011). Psychological issues in genetic testing for inherited cardiovascular diseases. *Circulation: Cardiovascular Genetics*, *4*(1):81-90.

Achelrod, D., Blankart, C.R., Linder, R., Von Kodolitsch, Y. and Stargardt, T., (2014). The economic impact of Marfan syndrome: a non-experimental, retrospective, population-based matched cohort study. *Orphanet Journal of Rare Diseases*, *9*(1):1-12.

Ades, L., (2007). Guidelines for the diagnosis and management of Marfan syndrome. *Heart, Lung and Circulation*, *16*(1):28-30.

Adzhubei, I., Jordan, D.M. and Sunyaev, S.R., (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, *76*(1):7-20.

Aleem, A.A., (2019). DNA Sequencing Resolves Misdiagnosed and Rare Genetic Disorders. In *Biochemical Analysis Tools-Methods for Bio-Molecules Studies*. IntechOpen.

Ammash, N.M. and Connolly, H.M., (2007). Marfan syndrome diagnosis and management: article 2. *SA Heart*, *4*(1):10-17.

Anna, A. and Monika, G., (2018). Splicing mutations in human genetic disorders: examples, detection, and confirmation. *Journal of Applied Genetics*, *59*(3):253-268.

Aquila, I., Sacco, M.A., Cordasco, F. and Ricci, P., (2020). Forensic case of a pregnant woman with Marfan syndrome. *BMJ Case Reports*, *13*(12).

Arnaud, P., Milleron, O., Hanna, N., Ropers, J., Ouali, N.O., Affoune, A., Langeois, M., Eliahou, L., Arnoult, F., Renard, P. and Michelon-Jouneaux, M., (2021). Clinical relevance of genotype–phenotype correlations beyond vascular events in a cohort study of 1500 Marfan syndrome patients with FBN1 pathogenic variants. *Genetics in Medicine*, 23:1296–1304.

Arslan-Kirchner, M., Arbustini, E., Boileau, C., Child, A., Collod-Beroud, G., De Paepe, A., Epplen, J., Jondeau, G., Loeys, B. and Faivre, L., (2010). Clinical utility gene card for: Marfan syndrome type 1 and related phenotypes [FBN1]. *European Journal of Human Genetics*, *18*(9):1071-1071.

Attanasio, M., Lapini, I., Evangelisti, L., Lucarini, L., Giusti, B., Porciani, M.C., Fattori, R., Anichini, C., Abbate, R., Gensini, G.F. and Pepe, G., (2008). FBN1 mutation screening of patients with Marfan syndrome and related disorders: detection of 46 novel FBN1 mutations. *Clinical Genetics*, *74*(1):39-46.

Aubart, M., Benarroch, L., Arnaud, P., Collod-Béroud, G., Jondeau, G. and Boileau, C., (2016). Molecular genetics of the fibrillinopathies. *eLS*:1-13.

Baetens, M., Van Laer, L., De Leeneer, K., Hellemans, J., De Schrijver, J., Van De Voorde, H., Renard, M., Dietz, H., Lacro, R.V., Menten, B. and Van Criekinge, W., (2011). Applying massive parallel sequencing to molecular diagnosis of Marfan and Loeys-Dietz syndromes. *Human Mutation*, *32*(9):1053-1062.

Baranov, V.S., Arsentyev, V.G., Serebryakova, E.A. and Shabalov, N.P., (2018). Genetics of Connective Tissue Diseases: State of the Art and Perspectives. *Russian Journal of Genetics*, *54*(7):788-795.

Baudhuin, L.M., Kluge, M.L., Kotzer, K.E. and Lagerstedt, S.A., (2019). Variability in gene-based knowledge impacts variant classification: an analysis of FBN1 missense variants in ClinVar. *European Journal of Human Genetics*, *27*(10):1550-1560.

Baudhuin, L.M., Kotzer, K.E. and Lagerstedt, S.A., (2015). Increased frequency of FBN1 truncating and splicing variants in Marfan syndrome patients with aortic events. *Genetics in Medicine*, *17*(3):177-187.

Becerra-Muñoz, V.M., Gómez-Doblas, J.J., Porras-Martín, C., Such-Martínez, M., Crespo-Leiro, M.G., Barriales-Villa, R., de Teresa-Galván, E., Jiménez-Navarro, M. and Cabrera-Bueno, F., (2018). The importance of genotype-phenotype correlation in the clinical management of Marfan syndrome. *Orphanet Journal of Rare Diseases*, *13*(1):16.

Beck, T.F., Mullikin, J.C. and NISC Comparative Sequencing Program Biesecker Leslie G lesb@ mail. nih. gov, (2016). Systematic evaluation of Sanger validation of next-generation sequencing variants. *Clinical Chemistry*, *62*(4):647-654.

Behjati, S. and Tarpey, P.S., (2013). What is next generation sequencing? *Archives of Disease in Childhood-Education and Practice*, *98*(6):236-238.

Biggin, A., Holman, K., Brett, M., Bennetts, B. and Adès, L., (2004). Detection of thirty novel FBN1 mutations in patients with Marfan syndrome or a related fibrillinopathy. *Human Mutation*, *23*(1):99-99.

Bombardieri, E., Rohrbach, M., Greutmann, M., Matyas, G., Weber, R., Radulovic, J., Fasnacht Boillat, M., Linka, A., De Pasquale, G., Bonassin, F. and Attenhofer Jost, C.H., (2020). Marfan syndrome and related connective tissue disorders in the current era in Switzerland in 103 patients: Medical and surgical management and impact of genetic testing. *Swiss Medical Weekly*, *150*(1314).

Byers, P.H., Belmont, J., Black, J., De Backer, J., Frank, M., Jeunemaitre, X., Johnson, D., Pepin, M., Robert, L., Sanders, L. and Wheeldon, N., (2017). Diagnosis, natural history, and management in vascular Ehlers–Danlos syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 175(1):40-47.

Campbell, I.M., Kolodziejska, K.E., Quach, M.M., Wolf, V.L., Cheung, S.W., Lalani, S.R., Ramocki, M.B. and Stankiewicz, P., (2011). TGFBR2 deletion in a 20-month-old female with developmental delay and microcephaly. *American Journal of Medical Genetics Part A*, *155*(6):1442-1447.

Campbell, M.C. and Tishkoff, S.A., (2008). African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. *Annu. Rev. Genomics Human Genetics*, *9*:403-433.

Campens, L., Callewaert, B., Mosquera, L.M., Renard, M., Symoens, S., De Paepe, A., Coucke, P. and De Backer, J., (2015). Gene panel sequencing in heritable thoracic aortic disorders and related entities–results of comprehensive testing in a cohort of 264 patients. *Orphanet Journal of Rare Diseases*, *10*(1):1-9.

Cartegni, L., Chew, S.L. and Krainer, A.R., (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nature Reviews Genetics*, *3*(4):285-298.

Chai, C., (2019). Principle of Emulsion PCR and Its Applications in Biotechnology. *Journal of Animal Reproduction and Biotechnology*, *34*(4):259-266.

Chandra, A. and Charteris, D., (2014). Molecular pathogenesis and management strategies of ectopia lentis. *Eye*, 28(2):162-168.

Chandra, A., Aragon-Martin, J.A., Hughes, K., Gati, S., Reddy, M.A., Deshpande, C., Cormack, G., Child, A.H., Charteris, D.G. and Arno, G., (2012). A genotype-phenotype comparison of ADAMTSL4 and FBN1 in isolated ectopia lentis. *Investigative Ophthalmology & Visual Science*, *53*(8):4889-4896.

Chandra, A., Patel, D., Aragon-Martin, J.A., Pinard, A., Collod-Béroud, G., Comeglio, P., Boileau, C., Faivre, L., Charteris, D., Child, A.H. and Arno, G., (2015). The revised ghent nosology; reclassifying isolated ectopia lentis. *Clinical Genetics*, *87*(3):284-287.

Child, A., Comeglio, P., Arno, G. and Beighton, P., (2007). Marfan syndrome in South Africa-a molecular genetic approach to diagnosis. *South African Medical Journal*, *97*(9):845-847.

Collod-Béroud, G., Le Bourdelles, S., Ades, L., Ala-Kokko, L., Booms, P., Boxer, M., Child, A., Comeglio, P., De Paepe, A., Hyland, J.C. and Holman, K., (2003). Update of the UMD-FBN1 mutation database and creation of an FBN1 polymorphism database. *Human Mutation*, *22*(3):199-208.

Cooper, G.M., Stone, E.A., Asimenos, G., Green, E.D., Batzoglou, S. and Sidow, A., (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Research*, *15*(7):901-913.

Cortini, F., Marinelli, B., Romi, S., Seresini, A., Pesatori, A.C., Seia, M., Montano, N. and Bassotti, A., (2017). A new COL3A1 mutation in Ehlers-Danlos syndrome Vascular type with different phenotypes in the same family. *Vascular and Endovascular Surgery*, *51*(3):141-145.

Cunniff, C., Byrne, J.L., Hudgins, L.M., Moeschler, J.B., Olney, A.H., Pauli, R.M., Seaver, L.H., Stevens, C.A. and Figone, C., (2000). Informed consent for medical photographs. *Genetics in Medicine*, *2*(6):353-355.

Damrauer, S.M., Hardie, K., Kember, R.L., Judy, R., Birtwell, D., Williams, H., Rader, D.J. and Pyeritz, R.E., (2019). FBN1 Coding variants and nonsyndromic aortic disease. *Circulation: Genomic and Precision Medicine*, *12*(6):e002454.

Davis, B.E. and Bennett, F.C., (2003). Developmental delay in children younger than 6 years. In S.Berman (Ed.), *Pediatric Decision-Making*, *4 Edition*: 86-89. Philadelphia: Mosby.

de Abreu, F.B., Peterson, J.D., Amos, C.I., Wells, W.A. and Tsongalis, G.J., (2016). Effective quality management practices in routine clinical next-generation sequencing. *Clinical Chemistry and Laboratory Medicine (CCLM)*, *54*(5):761-771.

De Backer, J., (2009). Cardiovascular characteristics in Marfan syndrome and their relation to the genotype. *Verhandelingen-Koninklijke Academie voor Geneeskunde van Belgie*, *71*(6):335-371.

De Paepe, A. and Malfait, F., 2012. The Ehlers–Danlos syndrome, a disorder with many faces. *Clinical Genetics*, *82*(1):1-11.

Deng, H., Lu, Q., Xu, H., Deng, X., Yuan, L., Yang, Z., Guo, Y., Lin, Q., Xiao, J., Guan, L. and Song, Z., (2016). Identification of a novel missense FBN2 mutation in a Chinese family with congenital contractural arachnodactyly using exome sequencing. *PloS One*, 11(5): e0155908.

Di Resta, C., Galbiati, S., Carrera, P. and Ferrari, M., (2018). Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities. *EJIFCC*, 29(1):4-14.

Dietz, H., (2017). Marfan syndrome. In *GeneReviews®[Internet*]. University of Washington, Seattle. 2019-05-09, http://www.ncbi.nlm.nih.gov/books/NBK1335/ .

Dong, C., Wei, P., Jian, X., Gibbs, R., Boerwinkle, E., Wang, K. and Liu, X., (2015). Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular SANTOS*, *24*(8):2125-2137.

Dong, J., Bu, J., Du, W., Li, Y., Jia, Y., Li, J., Meng, X., Yuan, M., Peng, X., Zhou, A. and Wang, L., (2012). A new novel mutation in FBN1 causes autosomal dominant Marfan syndrome in a Chinese family. *Molecular Vision*, *18*:81.

Dos Santos, E.S., Caputo, S.M., Castera, L., Gendrot, M., Briaux, A., Breault, M., Krieger, S., Rogan, P.K., Mucaki, E.J., Burke, L.J. and Bièche, I., (2018). Assessment of the functional impact of germline BRCA1/2 variants located in non-coding regions in families with breast and/or ovarian cancer predisposition. *Breast Cancer Research and Treatment*, *168*(2):311-325.

Doyle, A.J., Doyle, J.J., Bessling, S.L., Maragh, S., Lindsay, M.E., Schepers, D., Gillis, E., Mortier, G., Homfray, T., Sauls, K. and Norris, R.A., (2012). Mutations in the TGF-β repressor SKI cause Shprintzen-Goldberg syndrome with aortic aneurysm. *Nature Genetics*, 44(11):1249-1254.

Dunn, P., Albury, C.L., Maksemous, N., Benton, M.C., Sutherland, H.G., Smith, R.A., Haupt, L.M. and Griffiths, L.R., (2018). Next generation sequencing methods for diagnosis of epilepsy syndromes. *Frontiers in Genetics*, 9:1-11.

Elbadawi, A., Omer, M.A., Elgendy, I.Y., Abuzaid, A., Mohamed, A.H., Rai, D., Saad, M., Mentias, A., Rezq, A., Kamal, D. and Khalife, W., (2019). Losartan for Preventing Aortic Root Dilatation in Patients with Marfan Syndrome: A Meta-Analysis of Randomized Trials. *Cardiology and Therapy*, *8*(2):365-372.

Ergoren, M.C., Turkgenc, B., Teralı, K., Rodoplu, O., Verstraeten, A., Van Laer, L., Mocan, G., Loeys, B., Tetik, O. and Temel, S.G., (2019). Identification and characterization of a novel FBN1 gene variant in an extended family with variable clinical phenotype of Marfan syndrome. *Connective Tissue Research*, *60*(2):146-154.

Faber, J. and Fonseca, L.M., (2014). How sample size influences research outcomes. *Dental Press Journal of Orthodontics*, *19*(4):27-29.

Franken, R., (2016). *Marfan syndrome: Getting to the Root of the Problem*. Universiteit van Amsterdam [Host].

Franken, R., den Hartog, A.W., Radonic, T., Micha, D., Maugeri, A., van Dijk, F.S., Meijers-Heijboer, H.E., Timmermans, J., Scholte, A.J., van den Berg, M.P. and

Groenink, M., (2015). Beneficial outcome of losartan therapy depends on type of FBN1 mutation in Marfan syndrome. *Circulation: Cardiovascular Genetics*, *8*(2):383-388.

Franken, R., Heesterbeek, T.J., de Waard, V., Zwinderman, A.H., Pals, G., Mulder, B.J. and Groenink, M., (2014). Diagnosis and genetics of Marfan syndrome. *Expert Opinion on Orphan Drugs*, *2*(10):1049-1062.

Frédéric, M.Y., Monino, C., Marschall, C., Hamroun, D., Faivre, L., Jondeau, G., Klein, H.G., Neumann, L., Gautier, E., Binquet, C. and Maslen, C., (2009). The FBN2 gene: new mutations, locus-specific database (Universal Mutation Database FBN2), and genotype-phenotype correlations. *Human Mutation*, *30*(2):181-190.

Furtado, L.V., Wooderchak-Donahue, W., Rope, A.F., Yetman, A.T., Lewis, T., Plant, P. and Bayrak-Toydemir, P., (2011). Characterization of large genomic deletions in the FBN1 gene using multiplex ligation-dependent probe amplification. *BMC Medical Genetics*, *12*(1):1-7.

Fusco, C., Morlino, S., Micale, L., Ferraris, A., Grammatico, P. and Castori, M., (2019). Characterization of Two Novel Intronic Variants Affecting Splicing in FBN1-Related Disorders. *Genes*, *10*(6):442.

Gabriel, L.A., Wang, L.W., Bader, H., Ho, J.C., Majors, A.K., Hollyfield, J.G., Traboulsi, E.I. and Apte, S.S., (2012). ADAMTSL4, a secreted glycoprotein widely distributed in the eye, binds fibrillin-1 microfibrils and accelerates microfibril biogenesis. *Investigative Ophthalmology & Visual Science*, *53*(1):461-469.

Gao, L., Tian, T., Zhou, X., Fan, L., Wang, R. and Wu, H., (2019). Detection of ten novel FBN1 mutations in Chinese patients with typical or incomplete Marfan syndrome and an overview of the genotype-phenotype correlations. *International journal of Cardiology*, *293*:186-191.

Ghosh, R., Oak, N. and Plon, S.E., (2017). Evaluation of in silico algorithms for use with ACMG/AMP clinical variant interpretation guidelines. *Genome Biology*, *18*(1):1-12.

Gilissen, C., Hoischen, A., Brunner, H.G. and Veltman, J.A., 2012. Disease gene identification strategies for exome sequencing. *European Journal of Human Genetics*, *20*(5):490-497.

Gong, B., Yang, L., Wang, Q., Ye, Z., Guo, X., Yang, C., Hao, F., Shi, Y., Huang, Y., Qu, C. and Yang, Z., (2019). Mutation screening in the FBN1 gene responsible for Marfan syndrome and related disorder in Chinese families. *Molecular Genetics & Genomic Medicine*, *7*(4):e00594.

Gordon, K.J. and Blobe, G.C., (2008). Role of transforming growth factor-β superfamily signaling pathways in human disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, *1782*(4):197-228.

Grahame, R. and Kazkaz, H., 2014. The rheumatological heritable disorders of connective tissue. *Medicine*, *42*(5):275-278.

Grange, T., Aubart, M., Langeois, M., Benarroch, L., Arnaud, P., Milleron, O., Eliahou, L., Gross, M.S., Hanna, N., Boileau, C. and Gouya, L., (2020). Quantifying the genetic basis of marfan syndrome clinical variability. *Genes*, *11*(5):574.

Greally, M.T., (2013). Shprintzen-Goldberg syndrome. *In GeneReviews®[Internet].* University of Washington, Seattle. 2019-05-09, http://www.ncbi.nlm.nih.gov/books/NBK1335/.

Groenink, M. and Mulder, B.J., (2016). How to treat Marfan syndrome: an update.

Grosse, S.D., Rogowski, W.H., Ross, L.F., Cornel, M.C., Dondorp, W.J. and Khoury, M.J., (2010). Population screening for genetic disorders in the 21st century: evidence, economics, and ethics. *Public Health Genomics*, *13*(2):106-115.

Groth, K.A., Gaustadnes, M., Thorsen, K., Østergaard, J.R., Jensen, U.B., Gravholt, C.H. and Andersen, N.H., (2016). Difficulties in diagnosing Marfan syndrome using current FBN1 databases. *Genetics in Medicine*, *18*(1):98-102.

GTEx Consortium. "The GTEx Consortium atlas of genetic regulatory effects across human tissues." *Science* 369, no. 6509 (2020): 1318-1330.

Guo, D.C., Regalado, E.S., Minn, C., Tran-Fadulu, V., Coney, J., Cao, J., Wang, M., Yu, R.K., Estrera, A.L., Safi, H.J. and Shete, S.S., (2011). Familial thoracic aortic aneurysms and dissections: identification of a novel locus for stable aneurysms with a low risk for progression to aortic dissection. *Circulation: Cardiovascular Genetics*, *4*(1):36-42.

Guo, X., Song, C., Shi, Y., Li, H., Meng, W., Yuan, Q., Xue, J., Xie, J., Liang, Y., Yuan, Y. and Yu, B., (2016). Whole exome sequencing identifies a novel missense FBN2 mutation co-segregating in a four-generation Chinese family with congenital contractural arachnodactyly. *BMC Medical Genetics*, *17*(1):91.

Hakim, A. and Grahame, R., (2003). Joint hypermobility. *Best Practice & Research Clinical Rheumatology*, *17*(6):989-1004.

Hakim, A.J., Cherkas, L.F., Grahame, R., Spector, T.D. and MacGregor, A.J., (2004). The genetic epidemiology of joint hypermobility: a population study of female twins. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, *50*(8):2640-2644.

Hall, T.A., (1999), January. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic Acids Symposium Series*, 41(41):95-98). [London]: Information Retrieval Ltd., c1979-c2000.

Hara, H., Takeda, N., Fujiwara, T., Yagi, H., Maemura, S., Kanaya, T., Nawata, K., Morita, H. and Komuro, I., 2019. Activation of TGF-β signaling in an aortic aneurysm in a patient with Loeys-Dietz syndrome caused by a novel loss-of-function variant of TGFBR1. *Human Genome Variation*, *6*(1):1-4.

Hartman, P., Beckman, K., Silverstein, K., Yohe, S., Schomaker, M., Henzler, C., Onsongo, G., Lam, H.C., Munro, S., Daniel, J. and Billstein, B., (2019). Next generation sequencing for clinical diagnostics: five year experience of an academic laboratory. *Molecular Genetics and Metabolism Reports*, *19*:100464.

Hellen, B., (2009). Splice site tools: a comparative analysis report. *Mancester: National Genetics Reference Laboratory*:1-12.

Hoffjan, S., (2012). Genetic dissection of Marfan syndrome and related connective tissue disorders: an update 2012. *Molecular Syndromology*, *3*(2):47-58.

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D. and Cannon-Albright, L.A., (2016). REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, *99*(4):877-885.

Iskandar, Z., Mordi, I., Lang, C.C., Huang, J.T. and Choy, A.M., (2020). Biomarkers of aortopathy in Marfan Syndrome. *Cardiology in Review*, *28*(2):92-97.

Jacobs, A.M., Toudjarska, I., Racine, A., Tsipouras, P., Kilpatrick, M.W. and Shanske, A., (2002). A recurring FBN1 gene mutation in neonatal Marfan syndrome. *Archives of Pediatrics & Adolescent Medicine*, *156*(11):1081-1085.

Jennings, L.J., Arcila, M.E., Corless, C., Kamel-Reid, S., Lubin, I.M., Pfeifer, J., Temple-Smolkin, R.L., Voelkerding, K.V. and Nikiforova, M.N., (2017). Guidelines for validation of next-generation sequencing–based oncology panels: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *The Journal of Molecular Diagnostics*, *19*(3):341-365.

Jessurun, C.A., Bom, D.A. and Franken, R., (2016). An update on the pathophysiology, treatment and genetics of Marfan syndrome. *Expert Opinion on Orphan Drugs*, *4*(6):605-612.

Jian, X., Boerwinkle, E. and Liu, X., (2014). In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Research*, *42*(22):13534-13544.

Judge, D.P. and Dietz, H.C., (2005). Marfan's syndrome. *The Lancet*, *366*(9501):1965-1976.

Karczewski, K. and Francioli, L., (2017). The genome aggregation database (gnomAD). MacArthur Lab.

Karczewski, K.J., Weisburd, B., Thomas, B., Solomonson, M., Ruderfer, D.M., Kavanagh, D., Hamamsy, T., Lek, M., Samocha, K.E., Cummings, B.B. and Birnbaum, D., (2017). The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Research*, *45*(D1):D840-D845.

Karki, G. (2018). Connective tissue: characteristics, functions and types. [Image] Available at: https://www.onlinebiologynotes.com/connective-tissue-characteristics-functions-types/ [Accessed 13 Oct. 2020].

Khodakov, D., Wang, C. and Zhang, D.Y., (2016). Diagnostics based on nucleic acid sequence variant profiling: PCR, hybridization, and NGS approaches. *Advanced Drug Delivery Reviews*, *105*:3-19.

Kilpatrick, M.W., Harton, G.L., Phylactou, L.A., Levinson, G., Fugger, E.F., Schulman, J.D., Black, S.H. and Tsipouras, P., (1996). Preimplantation genetic diagnosis in Marfan syndrome. *Fetal Diagnosis and Therapy*, *11*(6):402-406.

Kitahara, H., Aeba, R., Takaki, H. and Shimizu, H., 2016. Palliative mitral valve repair during infancy for neonatal Marfan syndrome. *The Annals of Thoracic Surgery*, *101*(5):1987-1988.

Kromberg, J.G., Sizer, E.B. and Christianson, A.L., (2013). Genetic services and testing in South Africa. *Journal of Community Genetics*, *4*(3):413-423.

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

Kusindarta, D.L. and Wihadmadyatami, H., (2018). The role of extracellular matrix in tissue regeneration. *Tissue Regeneration*, *75728*:6.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W. and Karapetyan, K., (2018). ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*:1062-1067.

Latasiewicz, M., Fontecilla, C., Millá, E. and Sánchez, A., (2016). Marfan syndrome: ocular findings and novel mutations—in pursuit of genotype–phenotype associations. *Canadian Journal of Ophthalmology*, *51*(2):113-118.

Lee, S.J., Lee, D.H., Yoo, H.W., Koo, S.K., Park, E.S., Park, J.W., Lim, H.G. and Jung, S.C., (2005). Identification and functional analysis of cystathionine beta-synthase gene mutations in patients with homocystinuria. *Journal of Human Genetics,* 50(12):648-654.

Leopold, J.A. and Loscalzo, J., (2018). Emerging role of precision medicine in cardiovascular disease. *Circulation Research*, *122*(9):1302-1315.

Liu, L.H., Lin, S.M., Lin, D.S. and Chen, M.R., (2018). Losartan in combination with propranolol slows the aortic root dilatation in neonatal Marfan syndrome. *Pediatrics & Neonatology*, *59*(2):211-213.

Loeys, B., Nuytinck, L., Delvaux, I., De Bie, S. and De Paepe, A., (2001). Genotype and phenotype analysis of 171 patients referred for molecular study of the fibrillin-1 gene FBN1 because of suspected Marfan syndrome. *Archives of Internal Medicine*, *161*(20):2447-2454.

Loeys, B.L. and Dietz, H.C., (2018). Loeys-Dietz syndrome. *In GeneReviews®*[*Internet].* University of Washington, Seattle. 2019-05-09, https://www.ncbi.nlm.nih.gov/sites/books/NBK1133/.

Loeys, B.L., Dietz, H.C., Braverman, A.C., Callewaert, B.L., De Backer, J., Devereux, R.B., Hilhorst-Hofstee, Y., Jondeau, G., Faivre, L., Milewicz, D.M. and Pyeritz, R.E., (2010). The revised Ghent nosology for the Marfan syndrome. *Journal of Medical Genetics,* 47(7):476-485.

Ma, M., Li, Z., Wang, D.W. and Wei, X., (2016). Next-generation sequencing identifies novel mutations in the FBN1 gene for two Chinese families with Marfan syndrome. *Molecular Medicine Reports*, *14*(1):151-158.

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A. and Barrett, J.C., (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature*, *508*(7497):469-476.

Madar, L., Szakszon, K., Pfliegler, G., Szabó, G.P., Brúgós, B., Ronen, N., Papp, J., Zahuczky, K., Szakos, E., Fekete, G. and Oláh, É., (2019). FBN1 gene mutations in 26 Hungarian patients with suspected Marfan syndrome or related fibrillinopathies. *Journal of Biotechnology*, *301*:105-111.

Maeda, J., Kosaki, K., Shiono, J., Kouno, K., Aeba, R. and Yamagishi, H., (2016). Variable severity of cardiovascular phenotypes in patients with an early-onset form of Marfan syndrome harboring FBN1 mutations in exons 24–32. *Heart and Vessels*, *31*(10):1717-1723.

Málaga, D.R., Brusius-Facchin, A.C., Siebert, M., Pasqualim, G., Saraiva-Pereira, M.L., De Souza, C.F., Schwartz, I.V., Matte, U. and Giugliani, R., (2019). Sensitivity, advantages, limitations, and clinical utility of targeted next-generation sequencing panels for the diagnosis of selected lysosomal storage disorders. *Genetics and Molecular Biology*, *42*(1):197-206.

Malfait, F., Francomano, C., Byers, P., Belmont, J., Berglund, B., Black, J., Bloom, L., Bowen, J.M., Brady, A.F., Burrows, N.P. and Castori, M., (2017). The 2017 international classification of the Ehlers–Danlos syndromes. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, 175(1):8-26.

Malfait, F., Symoens, S., De Backer, J., Hermanns-Lê, T., Sakalihasan, N., Lapière, C.M., Coucke, P. and De Paepe, A., (2007). Three arginine to cysteine substitutions in the pro-alpha (I)-collagen chain cause Ehlers-Danlos syndrome with a propensity to arterial rupture in early adulthood. *Human Mutation*, *28*(4):387-395.

Mannucci, L., Luciano, S., Salehi, L.B., Gigante, L., Conte, C., Longo, G., Ferradini, V., Piumelli, N., Brancati, F., Ruvolo, G. and Novelli, G., (2020). Mutation analysis of the FBN1 gene in a cohort of patients with Marfan Syndrome: A 10-year single center experience. *Clinica Chimica Acta*, *501*:154-164.

Martínez-Quintana, E., Caballero-Sánchez, N., Rodríguez-González, F., Garay-Sánchez, P. and Tugores, A., (2017). Novel Marfan syndrome-associated mutation in the FBN1 gene caused by parental mosaicism and leading to abnormal limb patterning. *Molecular Syndromology*, *8*(3):148-154.

Martínez-Quintana, E., Rodríguez-González, F., Garay-Sánchez, P. and Tugores, A., (2014). A novel fibrillin 1 gene mutation leading to Marfan syndrome with minimal cardiac features. *Molecular Syndromology*, *5*(5):236-240.

Matkar, P.N., Chen, H.H., Leong-Poi, H. and Singh, K.K., (2015). Overview of Marfan Syndrome: knowns and unknowns. *Journal of Controversies in Biomedical Research*, *1*(1):51-66.

McCarthy, D.J., Humburg, P., Kanapin, A., Rivas, M.A., Gaulton, K., Cazier, J.B. and Donnelly, P., (2014). Choice of transcripts and software has a large effect on variant annotation. *Genome Medicine*, *6*(3):1-16.

McInerney-Leo, A.M., West, J., Meiser, B., West, M., Brown, M.A. and Duncan, E., (2020). Causal Attributions in an Australian Aboriginal family with Marfan syndrome: A qualitative study. *Frontiers in Genetics*, *11*.

McKee, T.J., Perlman, G., Morris, M. and Komarova, S.V., (2019). Extracellular matrix composition of connective tissues: a systematic review and meta-analysis. *Scientific eports*, *9*(1):1-15.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P. and Cunningham, F., (2016). The ensembl variant effect predictor. *Genome Biology*, *17*(1):1-14.

Merriman, B., R&D Team, I.T. and Rothberg, J.M., (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, *33*(23):3397-3417.

Miller, E.M., Patterson, N.E., Zechmeister, J.M., Bejerano-Sagie, M., Delio, M., Patel, K., Ravi, N., Quispe-Tintaya, W., Maslov, A., Simmons, N. and Castaldi, M., (2017). Development and validation of a targeted next generation DNA sequencing panel outperforming whole exome sequencing for the identification of clinically relevant genetic variants. *Oncotarget*, *8*(60):102033.

Mosca, M., Tani, C., Vagnani, S., Carli, L. and Bombardieri, S., (2014). The diagnosis and classification of undifferentiated connective tissue diseases. *Journal of Autoimmunity*, *48*:50-52.

Muiño-Mosquera, L., Steijns, F., Audenaert, T., Meerschaut, I., De Paepe, A., Steyaert, W., Symoens, S., Coucke, P., Callewaert, B., Renard, M. and De Backer, J., (2018). Tailoring the American College of Medical Genetics and Genomics and the Association for Molecular Pathology Guidelines for the Interpretation of Sequenced Variants in the FBN1 Gene for Marfan Syndrome: proposal for a disease-and gene-specific guideline. *Circulation: Genomic and Precision Medicine*, *11*(6):2039.

Murphy-Ryan, M., Psychogios, A. and Lindor, N.M., (2010). Hereditary disorders of connective tissue: a guide to the emerging differential diagnosis. *Genetics in Medicine*, *12*(6):344-354.

Muyas, F., Bosio, M., Puig, A., Susak, H., Domènech, L., Escaramis, G., Zapata, L., Demidov, G., Estivill, X., Rabionet, R. and Ossowski, S., (2019). Allele balance bias identifies systematic genotyping errors and false disease associations. *Human Mutation*, *40*(1):115-126.

Naidoo, P., Ranjith, N., Zikalala, Z., Mahoney, S. and Ho, K., (2018). Marfan syndrome: a case report and pictorial essay. *The Pan African Medical Journal*, *30*.

Naidoo, S., (2012). The South African national health insurance: A revolution in health-care delivery. *Journal of Public Health*, *34*(1):149-150.

National Department of Health , (2014), *Department of Health Strategic Plan 2014/15–2018/19,* Health e-News, Pretoria, from https://www.health-e.org.za/2014/09/25/report-department-health-strategic-plan-201415-20189/.

NCBI Resource Coordinators, (2014). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *42*:7-17.

Neuhann, T.M., Stegerer, A., Riess, A., Blair, E., Martin, T., Wieser, S., Kläs, R., Bouman, A., Kuechler, A. and Rittinger, O., (2015). ADAMTSL4-associated isolated ectopia lentis: Further patients, novel mutations and a detailed phenotype description. *American Journal of Medical Genetics Part A*, *167*(10):2376-2381.

Ojha, R. and Prasad, A.N., (2016). Menkes disease: what a multidisciplinary approach can do. *Journal of Multidisciplinary Healthcare*, *9*:371.

Olusanya, B.O., Davis, A.C., Wertlieb, D., Boo, N.Y., Nair, M.K.C., Halpern, R., Kuper, H., Breinbauer, C., De Vries, P.J., Gladstone, M. and Halfon, N., (2018). Developmental disabilities among children younger than 5 years in 195 countries and territories, 1990–2016: a systematic analysis for the global burden of disease study 2016. *The Lancet Global Health*, *6*(10):1100-1121.

Ombrello, M.J., Sikora, K.A. and Kastner, D.L., (2014). Genetics, genomics, and their relevance to pathology and therapy. *Best Practice & Research Clinical Rheumatology*, *28*(2):175-189.

Ott, J., Kamatani, Y. and Lathrop, M., 2011. Family-based designs for genome-wide association studies. *Nature Reviews Genetics*, *12*(7):465-474.

Ouyang, P.B., Zhao, Y., Peng, Y.Q., Zhang, L.S., Cao, J. and Li, Y., (2019). A novel mutation in FBN1 gene in autosomal dominant Marfan syndrome and macular degeneration in a Chinese consanguineous family. *International journal of ophthalmology*, *12*(5):725.

Overwater, E., Marsili, L., Baars, M.J., Baas, A.F., van de Beek, I., Dulfer, E., van Hagen, J.M., Hilhorst-Hofstee, Y., Kempers, M., Krapels, I.P. and Menke, L.A., (2018). Results of next- generation sequencing gene panel diagnostics including copy-number variation analysis in 810 patients suspected of heritable thoracic aortic disorders. *Human Mutation*, 39(9):1173-1192.

Owlia, M.B., (2006). Clinical spectrum of connective tissue disorders. *Journal, Indian Academy of Clinical Medicine*, *7*(3):217-224.

Parikh, V.N. and Ashley, E.A., (2017). Next-generation sequencing in cardiovascular disease: present clinical applications and the horizon of precision medicine. *Circulation*, *135*(5):406-409.

Pope, M.K., Ratajska, A., Johnsen, H., Rypdal, K.B., Sejersted, Y. and Paus, B., (2019). Diagnostics of hereditary connective tissue disorders by genetic next-generation sequencing. *Genetic Testing and Molecular Biomarkers*, *23*(11):783-790.

Pyeritz, R.E., (2012). Evaluation of the adolescent or adult with some features of Marfan syndrome. *Genetics in Medicine*, *14*(1):171-177.

Quang, D., Chen, Y. and Xie, X., (2015). DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics*, *31*(5):761-763.

Radonic, T., de Witte, P., Groenink, M., de Bruin-Bon, R.A., Timmermans, J., Scholte, A.J., van den Berg, M.P., Baars, M.J., van Tintelen, J.P., Kempers, M. and Zwinderman, A.H., (2011). Critical appraisal of the revised Ghent criteria for diagnosis of Marfan syndrome. *Clinical Genetics*, *80*(4):346-353.

Renner, S., Schüler, H., Alawi, M., Kolbe, V., Rybczynski, M., Woitschach, R., Sheikhzadeh, S., Stark, V.C., Olfe, J., Roser, E. and Seggewies, F.S., (2019). Next-generation sequencing of 32 genes associated with hereditary aortopathies and related disorders of connective tissue in a cohort of 199 patients. *Genetics in Medicine*, *21*(8):1832-1841.

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J. and Kircher, M., (2019). CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, *47*(D1):886-894.

Reva, B., Antipin, Y. and Sander, C., (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Research*, *39*(17):118-118.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E. and Voelkerding, K., (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5):405-423.

Rippe, M., De Backer, J., Kutsche, K., Mosquera, L.M., Schüler, H., Rybczynski, M., Bernhardt, A.M., Keyser, B., Hillebrand, M., Mir, T.S. and Berger, J., (2016). Mitral valve prolapse syndrome and MASS phenotype: Stability of aortic dilatation but progression of mitral valve prolapse. *IJC Heart & Vasculature*, *10*:39-46.

Robinson, P.N. and Godfrey, M., (2000). The molecular genetics of Marfan syndrome and related microfibrillopathies. *Journal of Medical Genetics*, *37*(1):9-25.

Rommel, K., Karck, M., Haverich, A., von Kodolitsch, Y., Rybczynski, M., Müller, G., Singh, K.K., Schmidtke, J. and Arslan-Kirchner, M., (2005). Identification of 29 novel and nine recurrent fibrillin-1 (FBN1) mutations and genotype–phenotype correlations in 76 patients with Marfan syndrome. *Human Mutation*, *26*(6):529-539.

Sacharow, S.J., Picker, J.D. and Levy, H.L., (2017). Homocystinuria caused by cystathionine beta-synthase deficiency. In *GeneReviews®[Internet]*. University of Washington, Seattle. https://www.ncbi.nlm.nih.gov/books/NBK1524/

Sakai, L.Y., Keene, D.R., Renard, M. and De Backer, J., (2016). FBN1: The disease-causing gene for Marfan syndrome and other genetic disorders. *Gene*, *591*(1):279-291.

Salam, A., Simpson, M.A., Stone, K.L., Takeichi, T., Nanda, A., Akiyama, M. and McGrath, J.A., (2014). Next generation diagnostics of heritable connective tissue disorders. *Matrix Biology*, *33*:35-40.

Sarr, S.A., Djibrilla, S., Aw, F., Bodian, M., Babaka, K., Ngaidé, A.A., Dioum, M. and Ba, S.A., (2017). Marfan syndrome and cardiovascular complications: results of a family investigation. *BMC Cardiovascular Disorders*, *17*(1):193.

Saudi Mendeliome Group falkuaya@ kfshrc. edu. sa, (2015). Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biology*, *16*:1-14.

Schepers, D., Doyle, A.J., Oswald, G., Sparks, E., Myers, L., Willems, P.J., Mansour, S., Simpson, M.A., Frysira, H., Maat-Kievit, A. and Van Minkelen, R., (2015). The SMAD-binding domain of SKI: a hotspot for de novo mutations causing Shprintzen–Goldberg syndrome. *European Journal of Human Genetics*, *23*(2):224.

Schrijver, I., Liu, W., Brenn, T., Furthmayr, H. and Francke, U., (1999). Cysteine substitutions in epidermal growth factor–like domains of fibrillin-1: distinct effects on biochemical and clinical phenotypes. *The American Journal of Human Genetics*, *65*(4):1007-1020.

Schwarz, J.M., Cooper, D.N., Schuelke, M. and Seelow, D., (2014). MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods*, *11*(4):.361-362.

Sefid Dashti, M.J. and Gamieldien, J., 2017. A practical guide to filtering and prioritizing genetic variants. *Biotechniques*, *62*(1):18-30.

Sellos-Moura, M., Glavin, F., Lapidus, D., Evans, K., Lew, C.R. and Irwin, D.E., (2020). Prevalence, characteristics, and costs of diagnosed homocystinuria, elevated homocysteine, and phenylketonuria in the United States: a retrospective claims-based comparison. *BMC Health Services Research*, *20*(1):1-11.

Sepetiene R, Simonyte S. FBN1 and TGFß1: Molecular mechanisms in the pathogenesis of thoracic aortic aneurysms and dissections (2016). *Mol Med Ther,* 1(1):1-7.

Shafique, M., Muzaffar, W. and Ishaq, M., (2016). The eye as a window to a rare disease: ectopia lentis and homocystinuria, a Pakistani perspective. *International Ophthalmology*, *36*(1):79-83.

Shihab, H.A., Gough, J., Cooper, D.N., Stenson, P.D., Barker, G.L., Edwards, K.J., Day, I.N. and Gaunt, T.R., (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*, *34*(1):57-65.

Spits, C., De Rycke, M., Verpoest, W., Lissens, W., Van Steirteghem, A., Liebaers, I. and Sermon, K., (2006). Preimplantation genetic diagnosis for Marfan syndrome. *Fertility and Sterility*, *86*(2):310-320.

Stark, V.C., Hensen, F., Kutsche, K., Kortüm, F., Olfe, J., Wiegand, P., von Kodolitsch, Y., Kozlik-Feldmann, R., Müller, G.C. and Mir, T.S., (2020). Genotype–Phenotype Correlation in Children: The Impact of FBN1 Variants on Pediatric Marfan Care. *Genes*, *11*(7):799.

Steiner, R.D. and Basel, D., (2019). COL1A1/2 osteogenesis imperfecta. *In GeneReviews®[Internet].* University of Washington, Seattle. 2005-01-28, https://www.ncbi.nlm.nih.gov/books/NBK1295/ .

Stengl, R., Bors, A., Ágg, B., Pólos, M., Matyas, G., Molnár, M.J., Fekete, B., Csabán, D., Andrikovics, H., Merkely, B. and Radovits, T., (2020). Optimising the mutation screening strategy in Marfan syndrome and identifying genotypes with more severe aortic involvement. *Orphanet Journal of Rare Diseases*, *15*(1):1-13.

Stenson, P.D., Ball, E.V., Mort, M., Phillips, A.D., Shiel, J.A., Thomas, N.S., Abeysinghe, S., Krawczak, M. and Cooper, D.N., (2003). Human gene mutation database (HGMD®): 2003 update. *Human Mutation*, *21*(6):577-581.

Steward, C.A., Parker, A.P., Minassian, B.A., Sisodiya, S.M., Frankish, A. and Harrow, J., (2017). Genome annotation for clinical genomic diagnostics: strengths and weaknesses. *Genome medicine*, *9*(1):1-19.

Stheneur, C., Collod-Béroud, G., Faivre, L., Buyck, J.F., Gouya, L., Le Parc, J.M., Moura, B., Muti, C., Grandchamp, B., Sultan, G. and Claustres, M., (2009). Identification of the minimal combination of clinical features in probands for efficient mutation detection in the FBN1 gene. *European Journal of Human Genetics*, *17*(9):1121-1128.

Stheneur, C., Faivre, L., Collod-Béroud, G., Gautier, E., Binquet, C., Bonithon-Kopp, C., Claustres, M., Child, A.H., Arbustini, E., Adès, L.C. and Francke, U., (2011). Prognosis factors in probands with an FBN1 mutation diagnosed before the age of 1 year. *Pediatric Research*, *69*(3):265-270.

Sui, R.F., Wei, H.B., Zhao, J.L., Hu, S.Y., Wang, B., Huang, S.Z. and Dong, M., (2004). Novel mutation of fibrillin 1 gene cause ectopia lentis in a Chinese family. *[Zhonghua yan ke za zhi] Chinese Journal of Ophthalmology*, *40*(12):828-831.

Summers, K.M., West, J.A., Hattam, A., Stark, D., McGill, J.J. and West, M.J., (2012). Recent developments in the diagnosis of Marfan syndrome and related disorders. *Medical Journal of Australia*, *197*(9):494-497.

Summers, K.M., West, J.A., Peterson, M.M., Stark, D., McGill, J.J. and West, M.J., (2006). Challenges in the diagnosis of Marfan syndrome. *Medical Journal of Australia*, *184*(12):627-631.

Takeda, N., Hara, H., Fujiwara, T., Kanaya, T., Maemura, S. and Komuro, I., (2018). TGF-β signaling-related genes and thoracic aortic aneurysms and dissections. *International Journal of Molecular Sciences*, *19*(7):2125.

Thorvaldsdóttir, H., Robinson, J.T. and Mesirov, J.P., (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, *14*(2):178-192.

Tiecke, F., Katzke, S., Booms, P., Robinson, P.N., Neumann, L., Godfrey, M., Mathews, K.R., Scheuner, M., Hinkel, G.K., Brenner, R.E. and Hövels-Gürich, H.H., (2001). Classic, atypically severe and neonatal Marfan syndrome: twelve mutations and genotype–phenotype correlations in FBN1 exons 24–40. *European Journal of Human Genetics*, *9*(1):13-21.

Tucci, S. and Akey, J.M., (2019). The long walk to African genomics. *Genome Biology*, *20*(1):1-3.

Tunçbilek, E. and Alanay, Y., (2006). Congenital contractural arachnodactyly (Beals syndrome). *Orphanet Journal of Rare Diseases*, 1:1-20.

UniProt Consortium, (2009). The universal protein resource (UniProt) 2009. *Nucleic Acids Research*, *37*: 169-174.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B.C., Remm, M. and Rozen, S.G., (2012). Primer3—new capabilities and interfaces. *Nucleic Acids Research*, *40*(15):115-115.

Vanakker, O., Callewaert, B., Malfait, F. and Coucke, P., (2015). The genetics of soft connective tissue disorders. *Annual Review of Genomics and Human Genetics*, *16*:229-255.

VanderJagt, K. and Butler, M.G., (2019). Ehlers–Danlos syndrome and other heritable connective tissue disorders that impact pregnancies can be detected using next-generation DNA sequencing.

Vanem, T.T., Rand-Hendriksen, S., Brunborg, C., Geiran, O.R. and Røe, C., (2020). Health-related quality of life in Marfan syndrome: a 10-year follow-up. *Health and Quality of Life Outcomes*, *18*(1):1-8.

Verstraeten, A., Alaerts, M., Van Laer, L. and Loeys, B., (2016). Marfan syndrome and related disorders: 25 years of gene discovery. *Human Mutation*, *37*(6):524-531.

Von Kodolitsch, Y., De Backer, J., Schüler, H., Bannas, P., Behzadi, C., Bernhardt, A.M., Hillebrand, M., Fuisting, B., Sheikhzadeh, S., Rybczynski, M. and Kölbel, T., (2015). Perspectives on the revised Ghent criteria for the diagnosis of Marfan syndrome. *The Application of Clinical Genetics*, *8*:137.

Wang, W.J., Han, P., Zheng, J., Hu, F.Y., Zhu, Y., Xie, J.S., Guo, J., Zhang, Z., Dong, J., Zheng, G.Y. and Cao, H., (2013). Exon 47 skipping of fibrillin-1 leads preferentially to cardiovascular defects in patients with thoracic aortic aneurysms and dissections. *Journal of Molecular Medicine*, *91*(1):37-47.

Weerakkody, R.A., Vandrovcova, J., Kanonidou, C., Mueller, M., Gampawar, P., Ibrahim, Y., Norsworthy, P., Biggs, J., Abdullah, A., Ross, D. and Black, H.A., (2016). Targeted next-generation sequencing makes new molecular diagnoses and expands genotype–phenotype relationship in Ehlers–Danlos syndrome. *Genetics in Medicine*, *18*(11):1119-1127.

Wooderchak-Donahue, W., VanSant-Webb, C., Tvrdik, T., Plant, P., Lewis, T., Stocks, J., Raney, J.A., Meyers, L., Berg, A., Rope, A.F. and Yetman, A.T., (2015). Clinical utility of a next generation sequencing panel assay for Marfan and Marfan-like syndromes featuring aortopathy. *American Journal of Medical Genetics Part A*, *167*(8):1747-1757.

Wu, Y., Sun, H., He, Y. and Zhang, H., (2021). A novel intron mutation in FBN-1 gene identified in a pregnant woman with Marfan syndrome. *Hereditas*, *158*(1):1-5.

Xu, P., Li, R., Huang, S., Sun, M., Liu, J., Niu, Y., Zou, Y., Li, J., Gao, M., Li, X. and Gao, X., (2020). A Novel Splicing Mutation in the FBN2 Gene in a Family with Congenital Contractural Arachnodactyly. *Frontiers in Genetics*, *11*:143.

Xu, S., Li, L., Fu, Y., Wang, X., Sun, H., Wang, J., Han, L., Wu, Z., Liu, Y., Zhu, J. and Sun, L., (2020). Increased frequency of FBN1 frameshift and nonsense mutations in Marfan syndrome patients with aortic dissection. *Molecular Genetics & Genomic Medicine*, *8*(1):1041.

Yang, H., Luo, M., Fu, Y., Cao, Y., Yin, K., Li, W., Meng, C., Ma, Y., Zhang, J., Fan, Y. and Shu, C., (2016). Genetic testing of 248 Chinese aortopathy patients using a panel assay. *Scientific Reports*, *6*(1):1-8.

Yang, H., Ma, Y., Luo, M., Zhao, K., Zhang, Y., Zhu, G., Sun, X., Luo, F., Wang, L., Shu, C. and Zhou, Z., (2018). Identification of gross deletions in FBN1 gene by MLPA. *Human Genomics*, *12*(1):1-7.

Yip, J. and Sawatzky, J.A., (2014). Cardiovascular management of Marfan syndrome: Implications for nurse practitioners. *The Journal for Nurse Practitioners*, *10*(8):594-602.

Yohe, S. and Thyagarajan, B., (2017). Review of clinical next-generation sequencing. *Archives of Pathology & Laboratory Medicine*, *141*(11):1544-1557.

Yoo, E.H., Woo, H., Ki, C.S., Lee, H.J., Kim, D.K., Kang, I.S., Park, P., Sung, K., Lee, C.S., Chung, T.Y. and Moon, J.R., (2010). Clinical and genetic analysis of Korean patients with Marfan syndrome: possible ethnic differences in clinical manifestation. *Clinical Genetics*, *77*(2):177-182.

Zhang, L., Xu, X., Sun, K., Sun, J., Wang, Y., Liu, Y., Yang, N., Tao, C., Cai, B., Shi, G. and Zhang, F., (2019). A de novo mutation in DHD domain of SKI causing spina bifida with no craniofacial malformation or intellectual disability. *American Journal of Medical Genetics Part A,* 179(6):936-939.

Zhang, M., Zhou, Y., Peng, Y. and Jin, L., (2018). Two rare missense mutations in the fibrillin-1 gene associated with atypical cardiovascular manifestations in a Chinese patient affected by Marfan syndrome. *Molecular Medicine Reports*, *18*(1):877-881.

Zhurayev, R., Proost, D., Zerbino, D., Fedorenko, V., Meester, J.A., Van Laer, L. and Loeys, B.L., (2016). Identification of FBN1 gene mutations in Ukrainian Marfan syndrome patients. *Genetics Research*, *98*.

APPENDICES

Appendix A: Ethics Clearance Certificate

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Miss Felicity Lindokuhle Mhlongo et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M191184

NAME: (Principal Investigator)	Miss Felicity Lindokuhle Mhlongo et al
DEPARTMENT:	Human Genetics National Health Laboratory Service
PROJECT TITLE:	Investigating the genetic aetiology of Heritable Connective Tissue Disorders in South Africa using Next Generation Sequencing
DATE CONSIDERED:	29/11/2019
DECISION:	Approved unconditionally
CONDITIONS:	Approval applies to CHBAH site only
SUPERVISOR:	Dr Nadia Carstens
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	18/03/2020

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **November** and will therefore be due in the month of **November** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature

Date 20/03/2020

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B: Connective Tissue Disorder Clinical Tick sheet

Name: _____ **Male** ☐ **Female** ☐

DOB: _____ **Caucasian** ☐ **Black** ☐

Hospital: _____ **Indian** ☐ **Mixed** ☐
ancestry

Hospital Number: _____ **Participant number:** _____

Attending Clinician: _____ **Possible diagnosis:** _____

GROWTH

Height:	<3rd centile	☐	3rd – 97th centile	☐	>97th centile	☐	SD	_____
Weight:	<3rd centile	☐	3rd – 97th centile	☐	>97th centile	☐	SD	_____
OFC:	<3rd centile	☐	3rd – 97th centile	☐	>97th centile	☐	SD	_____

Arm span: height ratio _____

US: LS ratio _____

SKIN/INTEGUMENT ANOMALIES

Hyperlaxity	☐	Abnormal scarring	☐	Varicose veins	☐
Hernias/ Prolapse	☐	Striae	☐		

Other_____

FACIAL FEATURES

General: flattening	Dolicocephaly	☐	Craniosynostosis	☐	Malar ☐
Eyes:	DSPF Hypertelorism	☐	Ectopia lentis	☐	☐
	Enophthalmos	☐			
Mouth:	High arched palate	☐	Cleft palate	☐	Bifid uvula ☐

Retrognathia ☐

Ears: Crumpled helix ☐

Other: _____

CARDIAC ANOMALIES

Aortic dilatation/aneurysm ☐ ASD ☐ VSD ☐ PDA ☐ AR ☐ ☐
MVP

Other aneurysms/dilatation ☐ Other:_____

MUSCULO-SKELETAL ANOMALIES

Chest: Pectus excavatum ☐ Pectus carinatum ☐

Hands: Long fingers ☐ Thumb sign ☐ Wrist sign ☐

Feet: Pes cavus ☐ Pes planus ☐ Hindfoot ☐
deformity

Spine: Kyphosis ☐ Scoliosis ☐

Joints: Laxity ☐ Beighton score _____ ☐
Camptodactyly

Dislocations ☐ _____ DDH ☐

Other:_____

Marfan systemic score:_____

NEURODEVELOPMENT

Normal ☐ Mild ID ☐ Moderate ID ☐ Severe ID ☐
0001256
0010864
0002342

SENSORY DEVELOPMENT

Vision: High myopia ☐ Retinal abnormality ☐ Other:_____

_____

OTHER IMPORTANT FEATURES/STRUCTURAL ANOMALIES

CNS:

Renal:

Hollow organ rupture:

Appendix C: Information Sheet and Consent Forms.

Heritable Connective Tissue Disorders Information Sheet

We (Dr Nadia Carstens and members of the Division of Human Genetics) are doing research on heritable connective tissue disorders. We are from the Division of Human Genetics (University of the Witwatersrand and National Health Laboratory Service) where this research will be done. Research is just the process to learn the answer to a question. In this study we want to understand the genetic factors that cause heritable connective tissue disorders in South Africans.

We would like to invite you to enter your child into this study as a participant, because your child has a heritable connective tissue disorder. In addition may also invite you and your family members to participate as analysis of additional family members my sometimes help with our research. Participation in the study will mean answering some questions about the health and background of you and your children. We will examine your child and a blood sample will be taken. We may also take you medical photographs. The photographs will not have your names on them and you can choose if you do or not want photographs of you to be taken. This will not affect your medical care. A separate consent form for photographs will be provided.

You and your children do not have to join this study. You can also decide to leave the study at any time. If you do not want to be part of this study or decide to leave the study it will not affect your child's regular treatments or medical care in any way. We invite your child to approach the researchers to be re-consented when he/she turns 18 if he/she wishes. He/she can also decide to leave the study at any time.

What are heritable connective tissue disorders?

There are more than 200 heritable disorders of connective tissue that can affect the tissues between the cells of your body that give tissues form and strength. All of these diseases are directly related to problems in genes that are responsible for building connective tissues. Connective tissues are biological tissues that support and join body organs together. For an example the connections between bone-to-bone and bone-to-muscle are connective tissues. The disorders are called "heritable," because they are passed on from parent to child.

Some heritable disorders of connective tissue change the look and growth of skin, bones, joints, heart, blood vessels, lungs, eyes, and ears. Others change how these tissues work. Many, but not all, are rare.

Information about DNA sampling

In order to study your genes we will take a sample of blood from your child and any additional participating family members. Our blood contains cells. All the cells in our bodies contain a chemical called DNA that we inherit from our parents. It carries the instructions for making us living organisms. It contains all our genes. We can take the DNA out of the cells that are in the sample of blood that we take from your child's arm. Once we have isolated the DNA from the blood we can study the genes. In this study we may need to examine every gene in the DNA to try and find the gene that causes the heritable connective tissue disorder. Sometimes we have to send a little bit of the sample overseas to experienced service providers or collaborators to examine these genes, because we don't always have the equipment and machines that are necessary to do this work in the country. All the information

that is discovered with this process will, however, return to South Africa where it can be analysed and used in other genetic studies.

We may also share the genetic or clinical data or the data from your DNA with other scientist in situations where the data can help the larger research community to better understand the genes that cause heritable connective tissue disorders. These data will be anonymized in all cases and we will not release your name(s).

If your family is interested in helping us with further research then a sample of your DNA will be stored in the Division of Human Genetics in the Molecular Genetics Laboratory here in South Africa for an indefinite period of time. It can sometimes take a long time and lots of experiments to fully understand the cause of a particular disease, which is why we ask to keep the samples. You will be asked to sign a separate consent form for this and your children may sign a separate assent form for this. In the future if new tests become available for genes involved in heritable connective tissue disorders we may then use this DNA sample to look further for these genes. We will only use the DNA for tests related to genetic- and heritable connective tissue disorders -related research and it will not be used for any other purpose. Any further tests will be monitored by the Wits Ethics Committee. If samples are to be sent overseas they will be protected by a legal agreement with Wits University. This will also be monitored by this Ethics Committee. You may choose to have the DNA destroyed at any time.

What to expect from the study?

Once you have signed the consent form and your children have signed the assent form we will ask you some questions about your and your children's background and health.

A complete clinical examination will be performed on your affected child. We may ask to access data relevant to your child's diagnosis from his/her clinical file or test results. This information will help us to understand your child's condition better and to interpret the data generated by our research. Unaffected family members may also be examined to confirm that they do not have the condition. This information will be used in the research study, but your family's personal details will be withheld. Then a nursing sister or doctor will take up to two teaspoons (10ml or 2 tubes) of blood from a vein in each participating family member. Venepunctures (drawing blood) are normally done as part of routine medical care and present a slight risk of discomfort. Drawing blood may result in faintness, inflammation of the vein, pain, bruising or bleeding at the puncture site. Your protection is that experienced personnel perform the procedures under sterile conditions.

Confidentiality

Only the investigators of the study will know your and your children's names and personal details. You will each be given a code and number for the study so that no one knows who you are. The blood and DNA samples will not be attached to your names – they will only have their code and number on them.

Contact details

If you have any problems you can contact us at any time on this telephone number:

Dr Nadia Carstens or Prof Amanda Krause – 011 489 9225.

Some families or patients may experience psychological distress should a complication be detected. You can contact the genetic counselling unit at the Division of Human Genetics, NHLS Braamfontein to arrange a session with an HPCSA registered genetic counsellor to discuss these issues. They are experienced and qualified counsellors who are able to offer distress counselling, decision support, address concerns associated with genetic disorders and -testing and explain medical terminology. Appointments can be arranged by contacting Ms Janet Robbie on (011)

489 9223 or Ms Shelley Macaulay on (011) 489 9243.

If you have any complaints or problems with this research you may report them to the Chair of the Human Research Ethics Committee at the University of the Witwatersrand on this telephone number – (011) 717 1234.

CONSENT FORM for adults

"The purpose of this study, procedures to be followed, risks and benefits, have been explained to me. I have been allowed to ask the questions I have, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this consent form and agree to participate in this research study with the understanding that I may withdraw at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one."

_____ _____

Signature of Participant Date

_____ _____

Signature of Person Obtaining Consent Date

CONSENT FORM for children

"The purpose of this study, procedures to be followed, risks and benefits, have been explained to me and my child. I have been allowed to ask the questions I have, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this consent form and agree to allow my child to participate in this research study with the understanding that I may withdraw my child at any time. We have discussed the study with my child to the best of our ability, who agrees to be in the study. We have explained to my child that he/she may approach the researchers to be re-consented when he/she turns 18 if he/she wishes and that he/she may withdraw from the study at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one."

_____ _____

Name of Child Date

_____ _____

Signature of Parent or Legal Guardian Date

_____ _____

Signature of Person Obtaining Consent Date

ASSENT FORM

"The purpose of this study, procedures to be followed, risks and benefits, have been explained to me. I have been allowed to ask the questions I have, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this assent form (or had it read to me) and agree to participate in this research study with the understanding that I may withdraw at any time. I understand that that I may approach the researchers to be re-consented when I turn 18 if I want to and that I may withdraw from the study at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one."

_____ _____

Signature of Child/ thumb print Date

_____ _____

Signature of Parent or Legal Guardian Date

_____ _____

Signature of Person Obtaining Consent Date

CONSENT FORM for ADULTS for DNA storage for possible future use

I, the undersigned, have been fully informed as to the procedure to be followed and have been given a description of the risks and benefits of having my DNA extracted and stored for possible further research.

In signing this form, I agree to have my DNA stored at the Division of Human Genetics in the Molecular Genetics Laboratory.

I understand that most of the DNA that is extracted from my blood will be stored in Johannesburg for an indefinite period of time and may be used for further genetic testing for genetic- or developmental disorder-related research only. This may include sending a portion of the sample overseas for further tests. Any further testing on these samples will be monitored by the Wits Ethics Committee.

I understand that if I have any questions at any time, they will be answered.

I understand that I can choose to have my DNA destroyed at any time.

I understand that confidentiality will be maintained at all times through the use of a code and numbering system.

A signed copy of this consent form will be made available to me if I want one.

The information around the blood and DNA samples taken from me is clear.

YES ☐ NO ☐

The purpose of this consent is for me to inform the study what they can or cannot do with these samples.

YES ☐ NO ☐

I acknowledge that all procedure/tests on the stored blood and DNA samples have been or will be approved by the Human Research Ethics Committee of the University of the Witwatersrand.

YES ☐ NO ☐

I am in agreement that my DNA may be stored and used for the purposes described above.

YES ☐ NO ☐

I am in agreement that the data generated from my DNA may be made available in a public domain without any identifiers.

YES ☐ NO ☐

I agree that a small bit of my DNA may be sent out of the country if the research cannot easily be done in South Africa.

YES ☐ NO ☐

I understand that every time a new study is done on my DNA, permission will be obtained from the ethics committee for the study to make sure that it is used only for the purposes stated above.

YES ☐ NO ☐

I understand that I may not benefit directly from the research done on my DNA and my family's DNA. I understand that the tests will reveal information about my biological ancestors, about my ability to respond to certain drugs and that genetic variants that may affect my risk for some diseases may be identified.

YES ☐ NO ☐

If this research does identify information that would be important or of benefit to me or my family, I would like to be contacted to receive the information.

YES ☐ NO ☐

I understand that my family and I may withdraw from the study at any time.

YES ☐ NO ☐

RESEARCHER:

Printed Name Signature Date

PARTICIPANT:

_____

Printed name Signature Date

WITNESS:

Printed Name Signature Date

CONSENT FORM for CHILDREN for DNA storage for possible future use

I, the undersigned, have been fully informed as to the procedure to be followed and have been given a description of the risks and benefits of having my child's blood sampled for the purpose of DNA extraction and storage for possible further research.

In signing this form, I agree to have my child's DNA stored at the Division of Human Genetics in the Molecular Genetics Laboratory.

I understand that most of the DNA that is extracted from my child's blood will be stored in Johannesburg for an indefinite period of time and may be used for further genetic testing for genetic- or developmental disorder-related research only. This may include sending a portion of the sample overseas for further tests. Any further testing on these samples will be monitored by the Wits Ethics Committee. I understand that that my child may approach the researchers to be re-consented when he/she turns 18 if he/she wants to and that he/she may withdraw from the study at any time.

I understand that if I have any questions at any time, they will be answered.

I understand that I can choose to have my child's DNA destroyed at any time.

I understand that confidentiality will be maintained at all times through the use of a code and numbering system.

I have discussed the study with my child, who agrees to the DNA storage.

A signed copy of this consent form will be made available to me if I want one.

The information around the blood and DNA samples taken from my child is clear.

YES ☐ NO ☐

The purpose of this consent form is for me to inform the study what they can or cannot do with these samples.

YES ☐ NO ☐

I acknowledge that all procedure/tests on the stored blood and DNA samples have been or will be approved by the Human Research Ethics Committee of the University of the Witwatersrand.

YES ☐ NO ☐

I am in agreement that my child's DNA may be stored and used for the purposes described above.

YES ☐ NO ☐

I am in agreement that the data generated from my child's DNA may be made available in a public domain without any identifiers.

YES ☐ NO ☐

I agree that a small bit of my child's DNA may be sent out of the country if the research cannot easily be done in South Africa.

YES ☐ NO ☐

I understand that every time a new study is done on my child's DNA, permission will be obtained from the ethics committee for the study to make sure that it is used only for the purposes stated above.

YES ☐ NO ☐

I understand that my child is unlikely to benefit directly from the research done on the DNA. I understand that the tests will reveal information about my child's biological ancestors, about my child's ability to respond to certain drugs and that genetic variants that may affect my child's risk for some diseases may be identified.

YES ☐ NO ☐

I understand that my child withdraw from the study at any time.

YES ☐ NO ☐

CHILD:

Printed Name

RESEARCHER:

Printed Name Signature Date

PARENT OR LEGAL GUARDIAN:

_____

Printed name Signature Date

WITNESS:

Printed Name Signature Date

ASSENT FORM for DNA storage for possible future use

"The purpose of storing my DNA has been explained to me – it may be used for health studies related to developmental disorders. I have been allowed to ask the questions I have, and my questions have been answered to my satisfaction. I have been told whom to contact if I have questions, to discuss problems, concerns, or suggestions related to the research, or to obtain information or offer input about the research. I have read this assent form (or had it read to me) and agree to participate in this research study with the understanding that I may withdraw at any time. I understand that if I want to withdraw from the research my DNA sample will be destroyed. I understand that that I may approach the researchers to be re-consented when I turn 18 if I want to and that I may withdraw from the study at any time. I have been told that I will be given a signed and dated copy of this consent form if I would like one."

_____ _____

Signature of Child/ thumb print Date

_____ _____

Signature of Parent or Legal Guardian Date

_____ _____

Signature of Person Obtaining Consent Date

Appendix D: Informed consent for patients' medical photographs adapted from ACMG

Patient name: _____ Date: _____

☐ check here if minor or unable to provide consent

I consent for medical photographs to be made of me or my child (or person for whom I am legal guardian). I understand that the information may be used in my medical record, for purposes of medical teaching, or for publication in medical textbooks or journals as I have designated below. By consenting to these medical photographs I understand that I will not receive payment from any party. Refusal to consent to photographs will in no way affect the medical care I will receive. If I have any questions or wish to withdraw my consent in the future I may contact:

_____

By signing this form below I confirm that this consent form has been explained to me in terms which I understand.

1) I consent for these photographs to be used in medical publications, including medical journals, textbooks, and electronic publications. I understand that the image may be seen by members of the general public, in addition to scientists and medical researchers that regularly use these publications in their professional education. Although these photographs will be used without identifying information such as my name, I understand that it is possible that someone may recognize me. I also agree for my image to be shown for teaching purposes and to be used for my medical record.

_____ (Signature) _____ (Witness)

2) I agree for my image to be shown for teaching purposes **AND** to be used for my medical record but **NOT FOR** medical publication:

_____ (Signature) _____ (Witness)

3) I agree to use of my image for medical records **ONLY**:

_____ (Signature) _____ (Witness)

For patients between ages 7 and 18 years, a signature below indicates that the information in this consent form has been explained to me, and I assent to use of my images as outlined above:

_____ _____

(Signature of patient) (Witness)

Appendix E: ACMG-AMP Guidelines of Variant Interpretation.

Criteria for Classifying Pathogenic Variants

Very strong evidence of pathogenicity

PVS1 Null variant (nonsense, frameshift, canonical +/−1 or 2 splice sites, initiation codon, single or multi-exon deletion) in a gene where loss of function (LOF) is a known mechanism of disease

Caveats:

- Beware of genes where LOF is not a known disease mechanism (e.g. *GFAP, MYH7*)
- Use caution interpreting LOF variants at the extreme 3' end of a gene
- Use caution with splice variants that are predicted to lead to exon skipping but leave the remainder of the protein intact
- Use caution in the presence of multiple transcripts

Strong evidence of pathogenicity

PS1 Same amino acid change as a previously established pathogenic variant regardless of nucleotide change

Example: Val->Leu caused by either G>C or G>T in the same codon

Caveat: Beware of changes that impact splicing rather than the amino acid/protein level

PS2 *De novo* (both maternity and paternity confirmed) in a patient with the disease and no family history

Note: Confirmation of paternity only is insufficient. Egg donation, surrogate motherhood, errors in embryo transfer, *etc*. can contribute to non-maternity

PS3 Well-established *in vitro* or *in vivo* functional studies supportive of a damaging effect on the gene or gene product

Note: Functional studies that have been validated and shown to be reproducible and robust in a clinical diagnostic laboratory setting are considered

the most well-established

PS4 The prevalence of the variant in affected individuals is significantly increased compared to the prevalence in controls

Note 1: Relative risk (RR) or odds ratio (OR), as obtained from case-control studies, is >5.0 and the confidence interval around the estimate of RR or OR does not include 1.0. See manuscript for detailed guidance.

Note 2: In instances of very rare variants where case-control studies may not reach statistical significance, the prior observation of the variant in multiple unrelated patients with the same phenotype, and its absence in controls, may be used as moderate level of evidence.

Moderate evidence of pathogenicity

PM1 Located in a mutational hot spot and/or critical and well-established functional domain (*e.g*. active site of an enzyme) without benign variation

PM2 Absent from controls (or at extremely low frequency if recessive) (see Table 6) in Exome Sequencing Project, 1000 Genomes or ExAC

Caveat: Population data for indels may be poorly called by next generation sequencing

PM3 For recessive disorders, detected in *trans* with a pathogenic variant

Note: This requires testing of parents (or offspring) to determine phase

PM4 Protein length changes due to in-frame deletions/insertions in a non-repeat region or stop-loss variants

PM5 Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before

Example: Arg156His is pathogenic; now you observe Arg156Cys

Caveat: Beware of changes that impact splicing rather than at the amino acid/protein level

PM6 Assumed *de novo*, but without confirmation of paternity and maternity

Supporting evidence of pathogenicity

PP1 Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease

Note: May be used as stronger evidence with increasing segregation data PP2 Missense variant in a gene that has a low rate of benign missense variation and where missense variants are a common mechanism of disease

PP3 Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc)

Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm should not be counted as an independent criterion. PP3 can be used only once in any evaluation of a variant.

PP4 Patient's phenotype or family history is highly specific for a disease with a single genetic etiology

PP5 Reputable source recently reports variant as pathogenic but the evidence is not available to the laboratory to perform an independent evaluation

Criteria for Classifying Benign Variants

Stand-Alone evidence of benign impact

BA1 Allele frequency is above 5% in Exome Sequencing Project, 1000 Genomes, or ExAC

Strong evidence of benign impact

BS1 Allele frequency is greater than expected for disorder (see table 6)

BS2 Observed in a healthy adult individual for a recessive (homozygous), dominant (heterozygous), or X-linked (hemizygous) disorder with full penetrance expected at an early age

BS3 Well-established *in vitro* or *in vivo* functional studies

shows no damaging effect on protein function or splicing

BS4 Lack of segregation in affected members of a family

Caveat: The presence of phenocopies for common phenotypes (*i.e.* cancer, epilepsy) can mimic lack of segregation among affected individuals. Also, families may have more than one pathogenic variant contributing to an autosomal dominant disorder, further confounding an apparent lack of segregation.

Supporting evidence of benign impact

BP1 Missense variant in a gene for which primarily truncating variants are known to cause disease

BP2 Observed in *trans* with a pathogenic variant for a fully penetrant dominant gene/disorder; or observed in *cis* with a pathogenic variant in any inheritance pattern

BP3 In-frame deletions/insertions in a repetitive region without a known function

BP4 Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc)

Caveat: As many *in silico* algorithms use the same or very similar input for their predictions, each algorithm cannot be counted as an independent criterion. BP4 can be used only once in any evaluation of a variant.

BP5 Variant found in a case with an alternate molecular basis for disease

BP6 Reputable source recently reports variant as benign but the evidence is not available to the laboratory to perform an independent evaluation

BP7 A synonymous (silent) variant for which splicing prediction algorithms predict no impact to the splice consensus sequence nor the creation of a new splice site AND the nucleotide is not highly conserved

Appendix F: Sanger Sequencing Protocol

Sequencing Protocol

POST-PCR STEPS

Post-PCR clean-up

1. Add water to each amplicon, so that the final volume is 100µL.
2. Transfer the contents of each tube to a clean PCR clean-up filter plates then vacuum until dry/10minutes.
3. Blot the bottom of the plate.
4. Add 100µL of ddH_2O to each well, and then vacuum again until dry.
5. Blot the bottom of the plate again.
6. Add 25µL ddH2O to each well, now containing the clean PCR product, and place on a shaker for seven minutes at low speed (about 60 rpm/minute).
7. Then transfer into PCR tubes for cycle sequencing.

Cycle Sequencing

1. Set up cycle sequencing as follows:

Reagent	Volume
Cleaned PCR Product	5µL
Bigdye v3.1	2µL
5× Bigdye buffer	1µL
Primer (forward or reverse)	1µL
ddH2O	1µL
Total	10µL

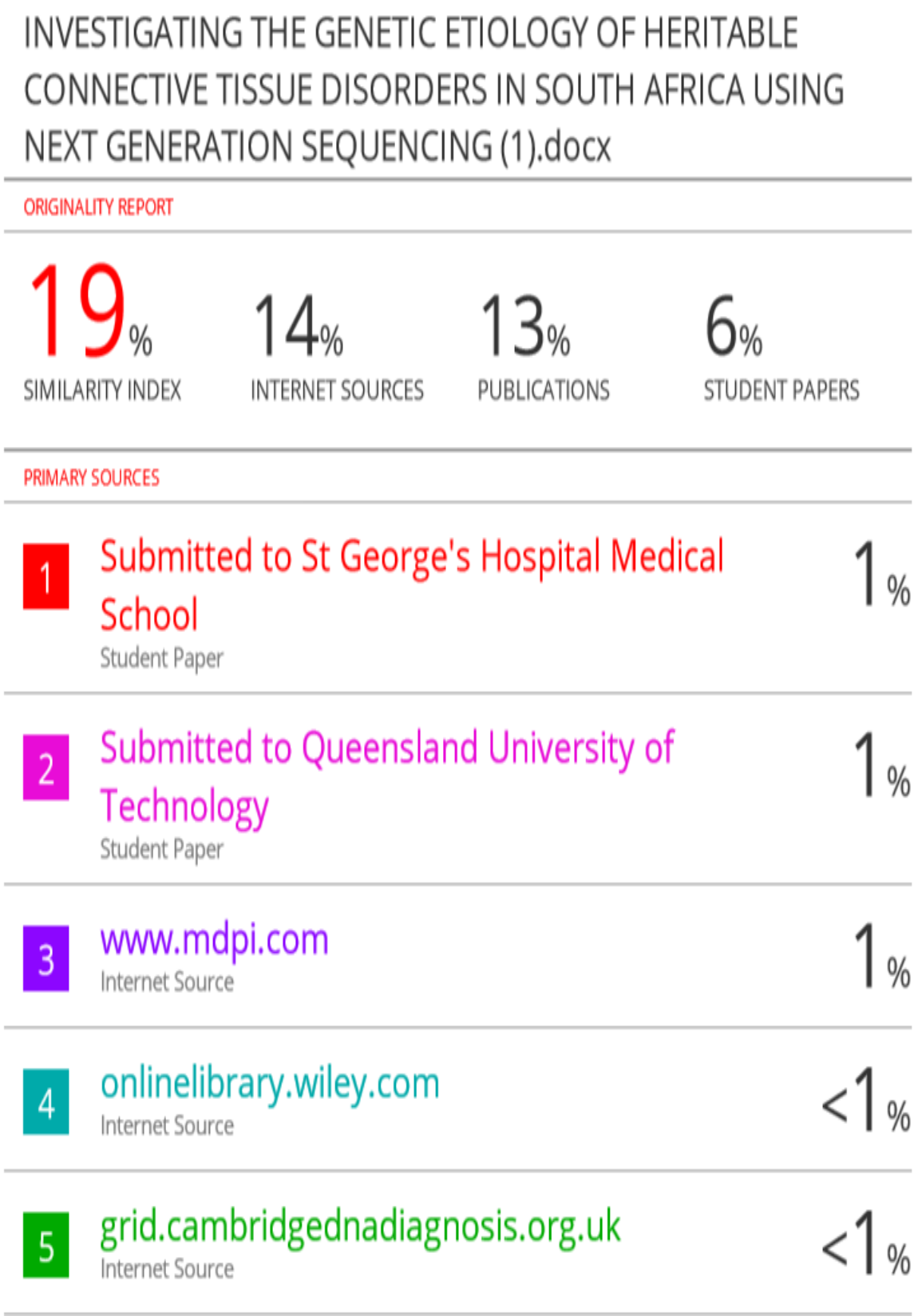
2. Complete cycle sequencing on a thermal cycler using the conditions below:

Temperature	Time	Cycles
96°C	1 minute	1
96°C	30 seconds	25
50°C	15 seconds	
60°C	4 minutes	
4-12°C	Hold	1

Post-cycle sequencing clean up

1. Add 40µL of Injection solution to each tube so that the final volume is 50µL.
2. Transfer the contents of each tube to a clean PCR clean-up filter plate then vacuum until dry.
3. Blot the bottom of the plate.
4. Then, add 25µL of injection solution to each well, and vacuum again until dry.
5. Blot the bottom of the plate again.
6. Add 25µL of injection solution to each well, and place on a shaker for seven minutes at low speed.
7. Transfer the contents of the wells to a plate for sequencing.

Appendix G: Turn-it-in Report.



INVESTIGATING THE GENETIC ETIOLOGY OF HERITABLE CONNECTIVE TISSUE DISORDERS IN SOUTH AFRICA USING NEXT GENERATION SEQUENCING (1).docx

ORIGINALITY REPORT

19%	14%	13%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to St George's Hospital Medical School Student Paper	1%
2	Submitted to Queensland University of Technology Student Paper	1%
3	www.mdpi.com Internet Source	1%
4	onlinelibrary.wiley.com Internet Source	<1%
5	grid.cambridgednadiagnosis.org.uk Internet Source	<1%