UNIVERSITY OF MALTA

The Genetics of Hypermobile Ehlers-Danlos Syndrome: A local study

A dissertation submitted to the Faculty of Health Sciences in partial fulfilment for the degree of Master of Science (Melit.) in Applied Biomedical Science.

Kirsty Scicluna

ID No. 406295M

Supervisor: Professor Isabella Borg MD, PhD
Co-supervisors: Dr. Melissa Formosa PhD
Dr. Rosienne Farrugia PhD

L-Università ta' Malta

FACULTY/INSTITUTE/CENTRE/SCHOOL of Health Sciences

DECLARATIONS BY POSTGRADUATE STUDENTS

(a) Authenticity of Dissertation

I hereby declare that I am the legitimate author of this Dissertation and that it is my original work.

No portion of this work has been submitted in support of an application for another degree or qualification of this or any other university or institution of higher education.

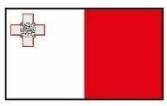
I hold the University of Malta harmless against any third party claims with regard to copyright violation, breach of confidentiality, defamation and any other third party right infringement.

(b) Research Code of Practice and Ethics Review Procedures

I declare that I have abided by the University's Research Ethics Review Procedures. Research Ethics & Data Protection form code 4307_24022020.

As a Master's student, as per Regulation 77 of the General Regulations for University Postgraduate Awards 2021, I accept that should my dissertation be awarded a Grade A, it will be made publicly available on the University of Malta Institutional Repository.

The research work disclosed in this publication is partially funded by the Endeavour Scholarship Scheme (Malta). Scholarships are part-financed by the European Union - European Social Fund (ESF)- Operational Programme II – Cohesion Policy 2014-2020 *"Investing in human capital to create more opportunities and promote the well-being of society*.



Operational Programme II - European Structural and Investment Funds 2014-2020
"Investing in human capital to create more opportunities and promote the well-being of society"
Project part-financed by the European Social Fund
Co-financing rate: 80% European Union; 20% National Funds

Lil Ommi Marthese,

Talli waħdek għamiltni l-mara li jien illum.

Grazzi Ma.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Isabella Borg, for her wise guidance and extensive knowledge, which she has imparted to this research study. Your contribution to the field of medical genetics in Malta and to countless patients and families throughout your career is truly indispensable. Thank you.

Secondly, I would also like to express my deepest and sincerest appreciation to my co-supervisors Dr Melissa Formosa and Dr Rosienne Farrugia for their vast technical insights, constant encouragement, patience and dedication throughout this study.

I would also like to thank:

-Ms Francesca Borg Carabott for her help in the bioinformatic analysis of sequencing data, without whom the interpretation of results would not have been possible.

-Ms Chanelle Cilia, Ms Azra Zejnelagic, Ms Daiva Svobaite Balciuniene and Ms Charlene Portelli as well as the staff at the Laboratory of Molecular Genetics (Ms Ruth Galdies, Ms Maia Lanfranco and Dr Jeanesse Scerri) for their help and guidance during laboratory work.

-Ms Julienne Borg of the Scholarships Unit for her kind help pertaining to the ENDEAVOUR scholarship scheme.

-My principal Ms Julie Haider, colleagues, and friends at the Bacteriology Laboratory for their support.

-A heartfelt thank you also goes to my partner Vanessa for her enduring patience and relentless support throughout this dissertation.

Lastly, I would like to extend my appreciation to the research participants of this study for their time and contribution to expand the molecular understanding of this disorder.

Abstract

Background: The Ehlers-Danlos Syndromes (EDS) are a collection of thirteen rare hereditary connective tissue disorders which present with heterogeneous phenotypes. The commonest and less severe subtype, hypermobile Ehlers-Danlos Syndrome (hEDS), presents a vast spectrum of signs and symptoms, including extensive joint hypermobility, musculoskeletal instability, chronic pain and autonomic dysfunction. To date, various international research efforts failed to uncover its molecular genetic aetiology, and this is further confounded by inter-familial genetic heterogeneity.

Aims: This original local research study sought to elucidate the genetic aetiology of hEDS in a small group of hEDS patients in Malta and to determine the effect of hEDS on bone mass.

Methodology: A multigenerational Maltese family as well as an unrelated participant were recruited. High-throughput sequencing was performed on selected participants using Illumina technology. Genetic analysis was also supplemented with bone mineral density (BMD) scans and testing of serum bone-related analytes.

Results: Two research participants had low bone mass and were classified as osteopenic, while serum levels of bone-related analytes were within the physiological range except in one osteopenic participant. Sequencing data was filtered using an in-house filtering pipeline to obtain a list of genetic variants found segregating with the disease phenotype in the family. This yielded five single nucleotide variants, namely *TNXB* rs61746206 (c.745G>A, p.Glu249Lys), *TNXB* rs140304758 (c.6973G>A, p.Val2325Ile), *SMAD3* rs189286879 (c.206+32287G>A), *FANCA/ZNF276* rs201316239 (n.47G>A) and *PNPLA1* rs45524833 (c.745G>A, p.Glu249Lys), which were not detected in the un-related participant. These variants were further annotated using gene databases and *in-silico* prediction tools to determine their clinical relevance.

Discussion: None of these variants was found to be pathogenic, thus highlighting the need for further research on larger cohorts and families of hEDS patients. The discovery of a potentially causative genetic pathogenic variation will enable the establishment of an early diagnosis and appropriate clinical management of this multisystemic hereditary disorder.

Table of Contents

Chapter 1: Introduction **1**
1.1 The connective tissue and its components 1
1.2 Heritable connective tissue disorders 4
1.3 The Ehlers-Danlos Syndromes 5
1.4 Hypermobile Ehlers-Danlos Syndrome 9
1.4.1 Prevalence across sexes 9
1.4.2 Difficulty in diagnosis 10
1.4.3 The hypermobility spectrum 11
1.4.4 Clinical Presentation 13
1.4.5 Management 22
1.5 Biomarkers 22
1.6 Genetics of hEDS 24
1.6.1 Mode of Inheritance 24
1.6.2 Molecular studies 25
1.6.3 Expression analyses 29
1.7 Statement of the problem 31
1.8 Aims & Objectives 31
Chapter 2: Methodology **33**
2.1 Recruitment of Research Participants 33
2.1.1 Selection of participants 33
2.1.2 Clinical examination 35
2.1.3 Medical History 35
2.1.4 Blood collection and serum measurements 35
2.1.5 Bone Mineral Density Assessment 36
2.2 Sample processing for genetic testing 36
2.2.1 DNA extraction 37
2.2.2 DNA quantification and purity check 38
2.3 High-throughput Sequencing 40
2.3.1 Selection and preparation of samples 42
2.3.2 Library preparation 44
2.3.3 Cluster Generation 46
2.3.4 Sequencing by Synthesis 47
2.3.5 Quality control of data and Bioinformatics 50

2.3.6 Processing of data and filtering of variants 53
2.4 Confirmatory testing 58
2.4.1 Polymerase chain reaction 58
2.4.2 Primer design 59
2.4.3 Optimisation of PCR procedure 60
2.4.4 Agarose gel electrophoresis 61
2.4.5 PCR Amplification 63
2.4.6 Purification of PCR products 66
2.4.7 Sanger sequencing 67
Chapter 3: Results 70
3.1 Recruitment of research participants 70
3.1.1 Clinical examination 70
3.1.2 Medical History 72
3.1.3 Biochemical measurements 73
3.2 High-Throughput Sequencing 76
3.2.1 Quality control of HTS data 76
3.2.2 Processing of data and filtering of variants 78
3.3 Confirmatory testing by Sanger sequencing 82
3.4 Variant analysis 85
3.4.1 *TNXB* rs61746206 (c.517G>A) 85
3.4.2 *TNXB* rs140304758 (c.6973G>A) 91
3.4.3 *SMAD3* rs189286879 (c.206+32287G>A) 94
3.4.4 *FANCA*/*ZNF276* rs201316239 (c.*843C>T) 96
3.4.5 *PNPLA1* rs45524833 (c.745G>A) 98
Chapter 4: Discussion 104
Limitations of the study 125
Recommendations for future research 128
Conclusion 130
References 131
Appendix 155
Appendix A- hEDS diagnostic criteria 156
Appendix B-Medical Questionnaire 157
Appendix C-Protocols and volumes for preparation of buffers 162
Appendix D-Theragen sample QC report 164
Appendix E- TruSeq DNA PCR-free Protocol 167
Appendix F- TruSeq DNA Nano protocol 171
Appendix G-List of analysed genes in gene panels 175

Appendix H-Sample concentration prior to Sanger 183
Appendix I-Theragen WGS analysis report 185
Appendix J-Medical questionnaire answer list 209
Appendix K-Filtered variants in extended coding regions 213
Appendix L-Annotated gene list 237
Appendix M-Publication by author 245

List of Figures

Figure 1.1: Structure of the Extracellular Matrix 2
Figure 1.2: Hierarchal structure of collagen protein materials 3
Figure 1.3: Visual representation of the Beighton score.. 12
Figure 1.4: A diagram depicting the hypermobility spectrum and its associated disorders 13
Figure 1.5: Orthopaedic complaints in patients with hEDS.. 20
Figure 1.6: Immunofluorescence analysis of hEDS and control fibroblasts 30
Figure 2.1: Family pedigree 34
Figure 2.2: Second-generation sequencing technologies 41
Figure 2.3: DNA Library preparation using the Truseq PCR-free and Truseq Nano Illumina kits..... 45
Figure 2.4: Cluster generation by bridge amplification. 47
Figure 2.5: Sequencing by Synthesis (Illumina)..... 49
Figure 2.6: Post-sequencing analysis by Theragen Bio 51
Figure 2.7: Filtering strategy for autosomal dominant variants 53
Figure 2.8: Results of the primer optimisation process 63
Figure 2.9: Thermal profile of the PCR amplification run 64
Figure 2.10: Results of the PCR run prior to Sanger sequencing 65
Figure 2.11: PCR product purification AccuPrep® PCR/Gel Purification Kit..... 67
Figure 2.12: Sanger Sequencing. 68
Figure 3.1: Range of quality values across bases 76
Figure 3.2: Distribution of GC content 77
Figure 3.3: Coverage across reference genome and reference GC content..... 78
Figure 3.4: Detected genotypes by Sanger sequencing 82
Figure 3.5: Genotyping results of recruited family members..... 84
Figure 3.6: Variant *TNXB* rs61746206 in IV.8, as visualised using Integrative Genomics Viewer (IGV) 85
Figure 3.7: Conservation of the cytosine reference allele and coded amino acid across species..... 86
Figure 3.8: A 3D model of the human Tenascin X protein spanning amino acids 1-842 89
Figure 3.9: Predicted impact of *TNXB* rs61746206 on the protein structure of TN-X.. 90

Figure 3.10: Output from IGV, showing variant *TNXB* rs140304758 in IV.8 91
Figure 3.11: Conservation scores of the *TNXB* rs140304758 locus across several species alignments.....92
Figure 3.12: Variant *SMAD3* rs189286879) in IV.8.....94
Figure 3.13: Conservation tracks for *SMAD* rs189286879 (highlighted with a blue line) on UCSC genome browser 96
Figure 3.14: Variant c.*843C>T in IV.8. 97
Figure 3.15: Conservation tracks for the region surrounding *FANCA/ZNF276* rs201316239 (locus highlighted with a blue line) as viewed on the UCSC genome browser 98
Figure 3.16: Variant *PNPLA1* rs45524833 in IV.8, as visualised by IGV.....99
Figure 3.17: Conservation estimates of the reference allele (top of figure, position denoted by a blue line) and reference amino acid (bottom) for the variant locus of *PNPLA1* rs45524833, as displayed on the UCSC genome browser. 100
Figure 3.18: Results of the MutPred2 web server for Glu249Lys in *PNPLA1*, showing a pathogenicity score of 0.548, putative altered motifs and perturbed molecular mechanisms.....100
Figure 3.19: A structural model of Omega-hydroxyceramide transacylase, translated from the *PNPLA1* transcript ENST00000394571.2.. 102
Figure 3.20: A 3D model of the Omega-hydroxyceramide transacylase protein.....103
Figure 4.1: HaploReg output for regulatory motif perturbations by *TNXB* rs61746206. 110
Figure 4.2: Position of filtered variants in the linear structure of Tenascin X (TN-X).111
Figure 4.3 Linear structure of *SMAD3* and affected transcripts.....114
Figure 4.4: Canonical TGFβ signalling.....115
Figure 4.5: Genomic locations of *FANCA* and *ZNF276* within chromosome 16 118
Figure 4.6: List of transcription factor binding motifs altered by the FANCA/ZNF276 variant. 119
Figure 4.7: Position of *PNPLA1* rs45524833 in *PNPLA1* and Omega-hydroxyceramide transacylase 121
Figure 4.8: *PNPLA1* rs45524833 alters transcription factor binding motifs of Ets_known9, Nr2f2 and Pou1f1_2 on the forward strand. 122

List of Tables

Table 1.1: 2017 Classification of the Ehlers-Danlos Syndromes 7

Table 2.1: Measurement of serum analytes in hEDS participants 36

Table 2.2: Sample concentration and purity after DNA extraction 40

Table 2.3: An overview of the type of sequencing, platforms and companies used for high-throughput sequencing 43

Table 2.4: Preparation of samples to be sent for HTS 44

Table 2.5: Variant databases and in-silico prediction tools used to assess filtered variants. 55

Table 2.6: 3D protein structure repositories and prediction tools used to assess altered protein properties. 57

Table 2.7: Primer sequences used to confirm filtered variants 60

Table 2.8: Volumes of reagents used for PCR optimisation in each vial. 61

Table 3.1: Diagnostic outcomes of recruited individuals using the 2017 Diagnostic Criteria for hEDS 71

Table 3.2: Results of analyte testing from peripheral blood. 73

Table 3.3: Results of Bone Mineral Density measurements 75

Table 3.4: HTS analysis statistics 77

Table 3.5: Number of variants remaining after each filtering step for every gene panel/ all coding regions. 79

Table 3.6: List of shortlisted variants from filtering pipelines. 81

Table 3.7: Variant genotypes of recruited participants 83

Table 3.8: Predicted effect of *TNXB* rs61746206 on *TNXB* transcripts by various *in-silico* prediction tools, including pathogenicity scores. 88

Table 3.9: A table showing the *in-silico* putative effect of *TNXB* r140304758 on gene transcripts, using variant effect prediction tools and deleteriousness scores.....93

Table 3.10: Putative effect and deleteriousness scores of *PNPLA1* rs45524833 on *PNPLA1* transcripts, using several *in-silico* prediction tools......101

List of Abbreviations

A	AAF	Alternative allele frequency
	ACMG/AMP	American College of Medical Genetics and Genomics and Association for Molecular Pathology
	AD	Autosomal dominant
	ALK1	Activin A Receptor Like Type 1
	ALT	Alanine Aminotransferase
	ALP	Alkaline phosphatase
	ANA	Anti-nuclear antibodies
	ANS	Autonomic nervous system
	APOB	Apolipoprotein B-100
	AR	Autosomal recessive
	AST	Aspartate Aminotransferase
	AVM	Arteriovenous malformations
B	.bsl	Base call file
	BAP	Bone-specific alkaline phosphatase
	BMD	Bone mineral density
	bp	Base pair
C	.cif	Cluster intensity file
	C4BPA	C4b-binding protein α-chain
	CAH	Congenital adrenal hyperplasia
	CDNA	Complementary DNA
	CDS	Cording sequence
	cEDS	Classical Ehlers-Danlos Syndrome
	clEDS	Classical-like Ehlers-Danlos Syndrome
	COL1A1	Collagen Type I Alpha 1 Chain
	COL1A2	Collagen Type I Alpha 2 Chain
	COL3A1	Collagen Type III Alpha 1 Chain
	COL5A1	Collagen Type V Alpha 1 Chain
	COL5A2	Collagen Type V Alpha 2 Chain
	COL5A3	Collagen Type V Alpha 3 Chain
	COL6A3	Collagen Type VI Alpha 3 Chain
	CNV	Copy number variation
	CREs	Cis-regulatory elements
	CRP	C-reactive protein
	C9	Complement component C9
	C1R	Complement C1r subcomponent
	CT	Connective tissue
	CYP21A2	Cytochrome P450 Family 21 Subfamily A Member 2
D	DNA	Deoxyribonucleic acid
	ddNTP	Di-deoxyribonucleotide triphosphate

	dNTP	Deoxyribonucleotide triphosphate
	DXA	Dual-energy X-ray absorptiometry
E	E2F	E2 Transcription Factor
	EBF2	Early B-cell Factor 2
	EBF3	Early B-cell Factor 3
	ECM	Extracellular matrix
	EDS	Ehlers-Danlos Syndromes
	EDTA	Ethylenediaminetetraacetic acid
	EGF	Epidermal growth factor
	ELB	Erythrocyte lysing buffer
	ENG	Endoglin
	ESR	Erythrocyte sedimentation rate
F	*FANCA*	Fanconi anemia complementation group A
	FN	Femoral neck
G	GABP	GA-binding protein
	GC	guanine-cytosine
	GGT	Gamma-glutamyl transferase
	GI	Gastrointestinal
	G-HSD	Generalised Hypermobility Spectrum Disorder
	GOF	Gain-of-function
	GRCh37	Human genome build 37
	GRCh38	Human genome build 38
H	HCTDs	Heritable connective tissue disorders
	HEDGE	Hypermobile Ehlers Danlos Genetic Evaluation
	HHT	Hereditary Haemorrhagic Telangiectasia
	HR-pQCT	High-resolution peripheral quantitative computed tomography
	HSD	Hypermobility Spectrum Disorder
	HTS	High-throughput sequencing
I	ICTP	Carboxyterminal cross-linked telopeptide of type I collagen
	indel	Insertion/deletion
	IP	Inheritance pattern
J	JAK-STAT	Janus kinase-signal transducer and activator of transcription protein
	JH	Joint hypermobility
L	LC/MS/MS	nano-liquid chromatography tandem mass spectrometry
	LS	Lumbar spine
	LOD	Logarithm of the odds
	LOF	Loss-of-function
	LZTS1	Leucine Zipper Tumour Suppressor 1
M	MB	Megabytes
	MCAD	Mast cell activation disorder

	MH1	MAD homology 1
	MH2	MAD homology 2
	MMP-9	Matrix metallopeptidase 9
	mRNA	Messenger RNA
N	n	Number of participants
	NGS	Next generation sequencing
	NTC	No-template control
O	OC	Osteocalcin
	OH	hydroxy
	OMIM	Online Mendelian Inheritance in Man
P	PCR	Polymerase chain reaction
	pEDS	Periodontal Ehlers-Danlos Syndrome
	PI3K-AKT	Phosphoinositide-3-kinase–protein kinase B/Akt
	PICP	Carboxyterminal peptide of type I collagen
	PIIINP	Aminoterminal propeptide of type III procollagen
	PNPLA1	Patatin Like Phospholipase Domain Containing 1
	POTS	Postural orthostatic tachycardia syndrome
	PTH	Parathyroid hormone
Q	Q	Phred quality
R	RF	Rheumathoid factor
	RNA	Ribonucleic acid
	RP58	Transcriptional repressor 58
S	SARA	Smad anchor for receptor activation
	SDS	Sodium dodecyl sulphate
	SMAD2	SMAD family member 2
	SMAD3	SMAD family member 3
	SMAD4	SMAD family member 4
	SNV	Single nucleotide variant
T	.tiff	Tagged image file format
	TAE	Tris-acetate-EDTA
	TADs	Topologically-associated domains
	TE	Tris-EDTA
	TGFβ	Transforming Growth Factor-Beta
	TH	Total Hip
	T_m	Melting temperature
	TN-X	Tenascin X
	TNXB	Tenascin XB
	TTR	Transthyretin
V	.vcf	Variant call format
	vEDS	Vascular Ehlers-Danlos Syndrome
	VTN	Vitronectin
	VUS	Variant of uncertain significance
W	WES	Whole exome sequencing

	WGS	Whole genome sequencing
Z	ZNF135	Zinc finger protein 135
	ZNF276	Zinc finger protein 276
	ZNF460	Zinc finger protein 460

Chapter 1: Introduction

The Ehlers-Danlos syndromes (EDS) are a group of thirteen rare hereditary connective tissue disorders which mainly affect the structure, production and interaction of collagen and other proteins in the extracellular matrix (ECM). These perturbations in the connective tissue (CT) result in a myriad of clinical presentations specific to each subtype, characterised by the combination of musculoskeletal instability and fragility of cutaneous tissue, as will be discussed below.

1.1 The connective tissue and its components

Amongst the four basic types of human tissue: epithelial, connective, muscular and neural tissue, the connective tissue (CT) network is the most abundant and widely distributed in human anatomy. It can be broadly classified into two categories: connective tissue proper and specialised connective tissue such as reticular tissue, adipose tissue, blood, bone and cartilage (Ogawa, LaRue, et al. 2010). The CT acts as a physical three-dimensional scaffold; binding, supporting and connecting other tissue types while also being involved in metabolic processes such as cell maturation, differentiation and signalling cascades (Theocharis, Manou, et al. 2019, Gjaltema and Bank 2017, Rozario and DeSimone 2010). CT is composed of fibroblasts which are suspended in and secrete the ECM, as well as proteins such as collagen and elastin (Alberts, Johnson, et al. 2002). The chemical composition of this matrix is determined by the structural requirements of the connective tissue subtype (Gelse 2003): cartilage and bone require a strong, rigid, mineralised matrix composition in comparison to tendons and ligaments (McKee, Perlman, et al. 2019), which primarily require more elasticity for movement. Fibrillar proteins like collagens exert their anchoring properties by interacting with cells via integrins (**Figure 1.1**). Integrins are cell-membrane proteins that connect the ECM to the intracellular cytoskeleton (Campbell and Humphries 2011).

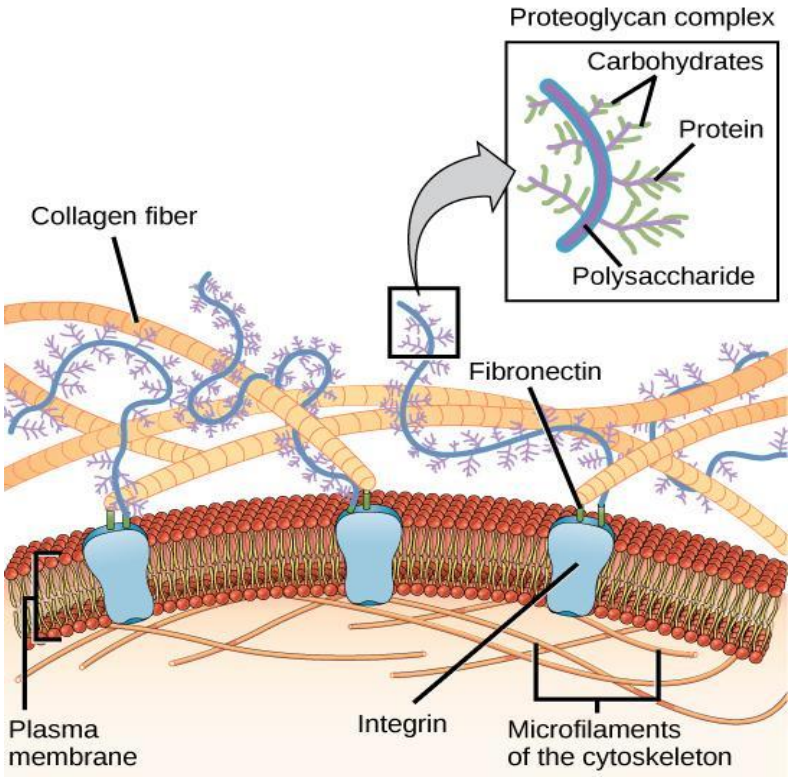


Figure 1.1: Structure of the Extracellular Matrix
A figure showing the interaction of the ECM with the intracellular cytoskeleton. Collagen and proteoglycan complexes are attached to cell membrane-bound integrins, which are in turn bound to cytoplasmic microfilaments. Image by OpenStax Biology, obtained from Wikimedia Commons at https://commons.wikimedia.org/wiki/File:Figure_04_06_01.jpg. Permission to reuse under the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/deed.en)**.**

Intracellular events are also mediated by ECM-associated proteins such as proteases, growth factors and cytokines which interact with cell membrane proteins (Hynes 2009, Chiarelli, Carini, et al. 2019). The ECM is also involved in the mediation of immune cells to damaged tissue, enhancing migration by chemotaxis and immune cell proliferation (Frevert, Felgenhauer, et al. 2018). Members of the collagen family are the most abundant proteins in the ECM (Cortini, Villa, et al. 2019) and are a major contributor to structural stability in interstitial tissue and organs (Chiarelli, Carini, et al. 2019, Cortini, Villa, et al. 2019, De Paepe and Malfait 2004). Collagens are especially prevalent in the musculoskeletal system, where they provide tensile and structural strength in bone, cartilage, ligaments and tendons (Gelse 2003, De Paepe and Malfait 2004). Apart from their physical function, collagens are also implicated in a multitude of metabolic processes including cell adhesion, signalling by

interaction with membrane receptors, mediation of growth factors, wound healing and tissue repair (Hay 1981, Myllyharju and Kivirikko 2001, De Paepe and Malfait 2004).

Recent molecular studies identified 29 isotypes of collagen (Cortini, Villa, et al. 2019) with Type I, III and V collagen being the predominant types in bone, blood vessels and muscle respectively (Chiarelli, Carini, et al. 2019, De Paepe and Malfait 2004). Despite considerable differences in their structure, size and distribution across tissues, all collagen subtypes are configured from three polypeptide chains arranged in a right-handed triple helix (Fraser, MacRae, et al. 1979). Each of these polypeptide chains is an α-chain composed of a repetition of glycine, hydroxyproline and proline residues (**Figure 1.2**; Gelse 2003).

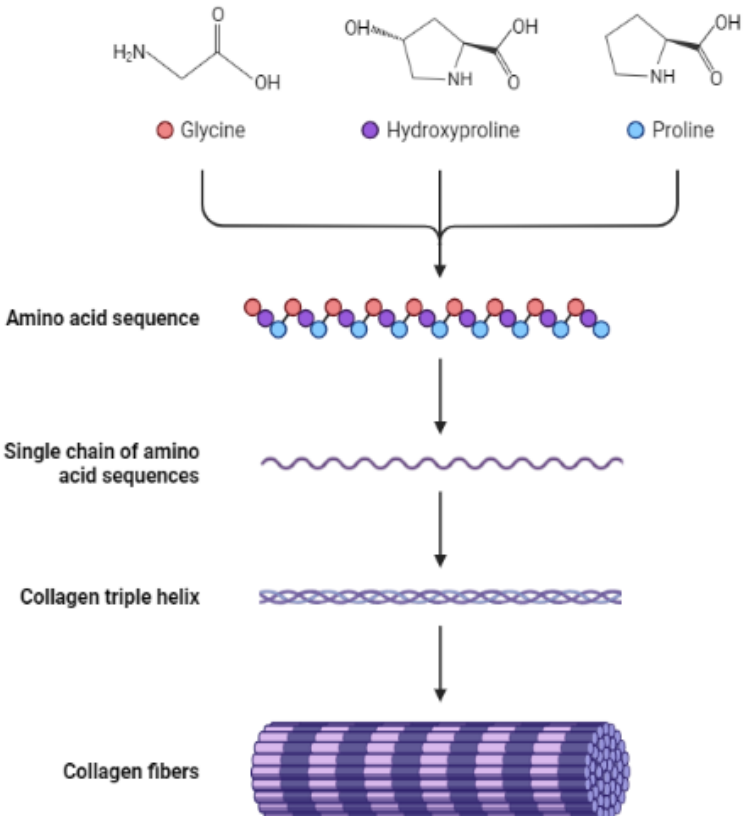


Figure 1.2: Hierarchal structure of collagen protein materials
A figure showing the formation of amino acid chains composed of Glycine, Hydroxyproline and Proline residues. These chains form collagen triple helixes by covalent bonding, which are then assembled into collagen fibrils. Reprinted from "Collagen amino acid chain", by BioRender.com (2023). Retrieved from https://app.biorender.com/biorender-templates, with permission to use under BioRender.com licensing and usage rights.

In each chain, the presence of glycine, the smallest amino acid, at every third position ensures a tight assembly of the helix, while hydroxyproline and proline residues are essential for covalent cross-linkage of multiple helixes (Myllyharju and Kivirikko 2001).

Differences in the order of these amino acid residues result in different α-chains. Collagen types containing three identical α-chains include collagens II, III, VII, VIII and X while helixes assembled from two or three different polypeptide chains form types I, IV, V, VI, I X and XI (Gelse 2003). Each helix may be coded by genes on several chromosomes: for example, collagen type 1 is composed of two α1-chains and one α2-chain, coded by the genes collagen type I, alpha 1 (*COL1A1*) and collagen type I, alpha 2 (*COL1A2*) respectively (De Paepe and Malfait 2004). Genes coding for collagen α-chains are renowned for their susceptibility to glycine substitutions, defective Ribonucleic acid (RNA) splicing and partial gene deletions (Kuivaniemi, Tromp, et al. 1997) causing quantitative and qualitative disturbances with aberrant collagen deposition. Molecular modifications in genes encoding collagen, as well as enzymes catalysing, processing and regulating ECM components are the basis for a variety of pathological mechanisms (Lamandé and Bateman 2019) including fibrosis (Gjaltema and Bank 2017), osteoarthritis, osteoporosis, and most importantly, hereditary connective tissue disorders (Gelse 2003).

1.2 Heritable connective tissue disorders

Heritable connective tissue disorders (HCTDs) are a group of rare congenital disorders originating from disturbances in the production, function and regulation of CT (Bascom, Schubart, et al. 2019). Due to the extensive presence of CT throughout all tissues, most HCTDs are characterised by a heterogeneous presentation spanning multiple organs, which also presents diagnostic challenges due to substantial phenotypic overlap (Bascom, Schubart, et al. 2019). Musculoskeletal complications are a characteristic feature of most HCTDs, including

generalised tissue laxity and joint hypermobility (JH) (Grahame 2000) that is, a range of joint motion that extends beyond normal physiological limits (Beighton, Solomon, et al. 1973).

The diagnostic criteria for HCTDs evolved from the initial broad Berlin nosology (Beighton, de Paepe, et al. 1988) to specific ones such as the Ghent criteria for Marfan syndrome (De Paepe, Devereux, et al. 1996) and the Villefranche criteria for the Ehlers-Danlos Syndromes, EDS (Beighton, De Paepe, et al. 1998) which are revised episodically. In recent years, the diagnosis of these disorders shifted from a clinical to a molecular basis *via* gene panels and/or genome sequencing due to advances in deoxyribonucleic acid (DNA) sequencing technology (Murphy-Ryan, Psychogios, et al. 2010, Rossi, Lee, et al. 2019). This is also evidenced by the revision in the Nosology of genetic skeletal disorders, which has been repeatedly updated over time to reflect novel gene associations and pathways implicated in skeletal dysplasias (Unger, Ferreira, et al. 2023). The genetic bases of well-described HCTDs like Marfan syndrome, Loeys-Dietz syndrome and Osteogenesis Imperfecta have already been identified (Basel and Steiner 2009, Beyens, Albuisson, et al. 2018, Dietz, Cutting, et al. 1991, Francomano, Liberfarb, et al. 1987, Malfait, Francomano, et al. 2017). However, the aetiology and pathological mechanism of some HCTD subtypes, including those of the Ehlers-Danlos syndromes (EDS), remain to be elucidated (Malfait, Francomano, et al. 2017).

1.3 The Ehlers-Danlos Syndromes

As previously stated, EDS are a protean group of rare congenital disorders distinguished by substantial phenotypic and genetic heterogeneity between subtypes (Castori 2016a, Malfait, Francomano, et al. 2017). The incidence of all EDS subtypes combined has been estimated at around 1 in 5,000 individuals (Steinmann, Royce, et al. 2003). The observed phenotypic and molecular heterogeneity as well as symptomatic overlap with other HCTDs has complicated greatly the diagnosis of EDS, which remains a contentious issue to this day (Castori, Camerota, et al. 2010, Malfait, Francomano, et al. 2017, Martin 2019). The syndromes have been reported

since antiquity (Byers and Murray 2014). However, it was not until 1901-1908 that the condition was separately described by dermatologists Edvard Ehlers (Ehlers 1901) and Henri-Alexandre Danlos (Danlos 1908), after whom the condition is eponymously named. Clinical classification of the disorders was undertaken in the late 20th century by the establishment of the Berlin Nosology (Beighton, de Paepe, et al. 1988), later expanded into the Villefranche criteria by the discovery of disease-causing variants using novel biochemical and molecular genetic methods (Beighton, De Paepe, et al. 1998).

Due to the increasing use of high-throughput sequencing (HTS) and the discovery of additional variants in genes correlated with distinct phenotypes, the 2017 International Classification of the Ehlers-Danlos Syndromes was established (Malfait, Francomano, et al. 2017). This classification describes 13 EDS subtypes (**Table 1.1**), each associated with causative genetic modifications, excluding hypermobile Ehlers-Danlos Syndrome (hEDS). This classification broadly expanded and facilitated the diagnoses of EDS, increasing the number of cases having a definite molecular diagnosis (Sulli, Talarico, et al. 2018). As is evident from this classification, the aetiology of EDS transcends defective collagen structure and function, also encompassing other pathological pathways involving intracellular processing (Syx, Symoens, et al. 2015) and most interestingly, activation of complement. This latter subtype, periodontal EDS (pEDS) links inflammatory response with connective tissue dysfunction (Cortini, Villa, et al. 2019) and is an intriguing area of study in the elucidation of the genetics behind the unresolved hEDS subtype.

Table 1.1: 2017 Classification of the Ehlers-Danlos Syndromes

AD: Autosomal dominant, AR: Autosomal recessive, ECM: Extracellular matrix, EDS: Ehlers-Danlos Syndrome, GI: Gastrointestinal, IP: Inheritance pattern.

EDS Subtype	Abbreviation	Defining Phenotype	IP	Molecular basis	Affected Protein	Protein Function
Classical EDS	cEDS	Elastic, fragile skin, atrophic scarring, molluscoid pseudo tumours, joint hypermobility	AD	*COL5A1, COL5A1* Rare: *COL1A1*	Type V collagen Type I collagen	Fibrillation of types I and III collagen, modulation of fibril diameter and assembly Fibril-forming collagen in most connective tissues
Classical-like EDS	clEDS	Classic EDS symptoms without atrophic scarring	AR	*TNXB*	Tenascin XB	Maintenance of the collagen scaffold and regulation of the stability of elastic fibres
Cardiac-valvular EDS	cvEDS	Cardiac valvular defects due to mitral and/or aortic valve insufficiency	AR	*COL1A2*	Type I collagen	Fibril-forming collagen in most connective tissues
Vascular EDS	vEDS	Translucent skin, arterial & uterine rupture, colonic perforation	AD	*COL3A1* Rare: *COL1A1*	Type III collagen Type I collagen	Fibril-forming collagen frequently in association with type I collagen Fibril-forming collagen in most connective tissues
Hypermobile EDS	hEDS	Severe joint hypermobility chronic musculoskeletal pain, dysautonomia, GI involvement	AD	Unknown	Unknown	Unknown
Arthrochalasia EDS	aEDS	Congenital bilateral hip dislocation	AD	*COL1A1, COL1A2*	Type I collagen	Fibril-forming collagen in most connective tissues
Dermatosparaxis EDS	dEDS	Hyperextensible, redundant skin, ecchymosis	AR	*ADAMTS2*	ADAMTS-2	Processing of procollagen into mature collagen

EDS Subtype	Abbreviation	Defining Phenotype	IP	Molecular basis	Affected Protein	Protein Function
Kyphoscoliotic EDS	kEDS	Kyphoscoliosis, congenital muscle hypotonia	AR	*PLOD1* *FKBP14*	LH1 FKBP22	Hydroxylation of lysyl residues in collagen peptides, forming cross-links between collagen chains Acceleration of protein folding during protein synthesis
Brittle Cornea Syndrome	BCS	Progressive keratoconus & keratoglobus, blue sclerae	AR	*ZNF469* *PRDM5*	ZNF469 PRDM5	Regulating collagen fibre synthesis and organisation Regulating expression of proteins involved in ECM development and maintenance
Spondylodysplastic EDS	spEDS	Short stature, muscle hypotonia, bowing of limbs	AR	*B4GALT7* *B3GALT6* *SLC39A13*	ß4GalT7 ß3GalT6 ZIP13	Synthesis of several glycosylated and saccharide structures Transport of zinc into the cell, essential to the function of connective tissues
Musculocontractural EDS	mcEDS	Multiple congenital contractures, craniofacial deformities	AR	*CHST14* *DSE*	D4ST1 DSE	Transfers sulphate groups between molecules Production of a glycosaminoglycan which fills in gaps in connective tissue
Myopathic EDS	mEDS	Congenital muscle hypotonia, joint contractures	AD/ AR	*COL12A1*	Type XII collagen	Mediates interaction between collagen I fibrils and ECM
Periodontal EDS	pEDS	Severe periodontitis, gingivitis	AD	*C1R, C1S*	C1r, C1s	Complement subcomponents, activation of the complement cascade

1.4 Hypermobile Ehlers-Danlos Syndrome

Hypermobile EDS (hEDS), formerly known as EDS type III, joint hypermobility syndrome, benign joint hypermobility syndrome and EDS hypermobility type, is considered to be the commonest HCTD (Colombi, Dordoni, et al. 2015) and despite numerous research efforts, is also the least understood (Copetti, Morlino, et al. 2019, Syx, Symoens, et al. 2015). Contemporary estimations on the prevalence of hEDS are incomplete since hEDS is thought to comprise 80-90% of all EDS cases (Tinkle, Castori, et al. 2017). Prevalence estimations vary from 1 in 5000 (Steinmann, Royce, et al. 2003) to 20 in 1000 (Hakim and Sahota 2006). However, it is important to note that since these estimations are based on diagnostic criteria that are now obsolete, the actual prevalence of this condition is yet undetermined.

hEDS shares multiple features with classical EDS (cEDS), most notably generalised joint laxity and skin extensibility, however, the degree of JH, joint dislocations, chronic musculoskeletal pain and fatigue is far more pronounced in hEDS (Challal, Minichiello, et al. 2015, Tinkle, Castori, et al. 2017). From a medical perspective, hEDS is seemingly milder than classical and vascular EDS due to the absence of severe cardiovascular events and attenuated cutaneous features (Tinkle, Castori, et al. 2017). Nonetheless, hEDS is associated with a high disability potential (Castori, Sperduti, et al. 2011) and significant psychosocial impact (Challal, Minichiello, et al. 2015).

1.4.1 Prevalence across sexes

hEDS is inherited in an autosomal dominant (AD) manner (Levy 2018, Tinkle, Castori, et al. 2017). Thus, theoretically, an equal number of males and females is expected (Murray, Yashar, et al. 2013). However, clinical practice as well as research studies recorded a female majority of cases (90%) (Bendik, Tinkle, et al. 2011, Murray, Yashar, et al. 2013, Castori, Camerota, et al. 2010). Furthermore, upon clustering hEDS patients into two subgroups (mild or severe)

based on their clinical presentation, males were more likely to reside in the mild subgroup compared to females (Copetti, Morlino, et al. 2019). To explain this discrepancy, Castori et al. suggested that males have a lower pain perception than females, also arguing that an increased muscle mass and stiffer tendons in males supposedly counteract the effects of joint instability (Castori, Camerota, et al. 2010). Recently, it was suggested that this skewed sex ratio is due to the high concentration of oestrogen in females, which influences pain perception (Forghani 2019, Hugon-Rodin, Lebègue, et al. 2016, Martin 2019, Tinkle, Castori, et al. 2017). This proposal is substantiated by the aggravation of joint laxity in females after puberty and around menstruation (Hugon-Rodin, Lebègue, et al. 2016, Quatman, Ford, et al. 2008) as well as new insights on the deleterious effect of oestrogen on CT (Leblanc, Schneider, et al. 2017). Nonetheless, all of these hypotheses remain purely speculative until the molecular basis of hEDS is elucidated (Rodgers, Gui, et al. 2017).

1.4.2 Difficulty in diagnosis

As the molecular mechanism behind hEDS has not been described, the diagnosis of hEDS relies exclusively on clinical examination (Malfait, Francomano, et al. 2017). The diagnosis is often by exclusion (Castori, Camerota, et al. 2010). A thorough physical examination of the patient and a detailed family history are paramount prior to performing an extended gene panel to exclude other HCTDs (Castori 2012). Physical and functional manifestations as well as the presence of a family history are then applied to the recently updated diagnostic criteria (International Consortium on EDS & Related Disorders 2017). Aside from the lack of molecular confirmation, clinical overlap with other disorders and symptom heterogeneity between patients, a significant problem in the diagnosis of hEDS remains the lack of awareness of this condition among physicians (Ross and Grahame 2011). Patients are, either misdiagnosed and given inadequate treatment (Castori 2016a), or diagnosed with obsessive

disorders (Bulbena, Baeza-Velasco, et al. 2017) or remain undiagnosed for many years (Demmler, Atkinson, et al. 2019, Ross and Grahame 2011).

1.4.3 The hypermobility spectrum

JH, or 'double jointedness' is defined as the physical ability of a joint to extend past its normal range of motion (Castori, Tinkle, et al. 2017). Hypermobile individuals are able to contort their bodies according to the hypermobile properties of their individual joints. JH can either be localised to a particular area, restricted to the joints of the hands and feet (peripheral JH) or present in multiple areas (generalised JH; Castori and Hakim 2017). JH is commonly observed in the general population with a wide prevalence of 2-57% (Remvig, Jensen, et al. 2007). However, this estimate varies according to age, gender and ethnicity (Castori 2012). JH is observed to decrease with age; thus, it is more common in children (Remvig, Jensen, et al. 2007).

For the most part, JH manifests as an asymptomatic often familial finding, unaccompanied by secondary complications (Castori, Tinkle, et al. 2017). However, JH may also be symptomatic, causing spontaneous joint dislocations, chronic pain and muscle weakness (Castori 2016a). The distinction between asymptomatic JH, JH is confined solely to musculoskeletal complaints and 'syndromic' JH with multisystem involvement and underlying genetic aetiology, is paramount for treatment and management (Castori 2012). A thorough physical examination and extensive family history are required in order to differentiate between the two hundred and sixteen genetic disorders associated with JH (Castori 2016).

The assessment of JH is commonly performed using the semi-qualitative Beighton score (Jacobs, Cornelissens, et al. 2018). The Beighton score tests the ability of an individual to perform a set of contortions using the hands, the limbs and the hip joints (**Figure 1.3**; Beighton, Solomon, et al. 1973). This assessment has a maximum score of 9, with the cut-off for

generalised JH being ≥6 for children, ≥5 for adolescents and adults up to 50 and ≥4 for adults over 50 years of age (Malfait, Francomano, et al. 2017). Both the Beighton score, as well as a five-part questionnaire for historical hypermobility, are included as mandatory criteria for the diagnosis of hEDS (Malfait, Francomano, et al. 2017).

Specific joint laxity	YES	NO
1. Passive apposition of thumb to forearm	☐ Left ☐ Right	☐
2. Passive hyperextension of V-MCP > 90°	☐ Left ☐ Right	☐
3. Active hyperextension of elbow >10°	☐ Left ☐ Right	☐
4. Active hyperextension of knee >10°	☐ Left ☐ Right	☐
5. Ability to flex spine placing palms to floor without bending knees	☐	☐

*Each "YES" is 1 point. A score ≥ 4 out 9 is generally considered an indication of JH. (MCP: metacarpophalangeal).

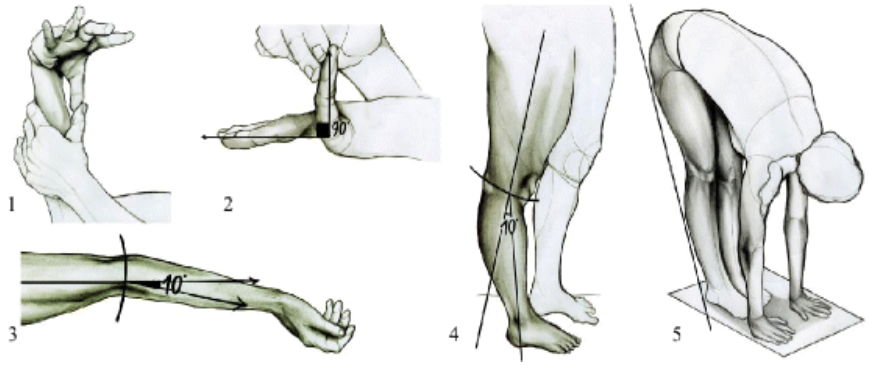


Figure 1.3: Visual representation of the Beighton score
Image obtained from Folci and Capsoni (2016), an article published by BMC Musculoskeletal Disorders. Permission to reuse under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

The 2017 international EDS classification sought to delineate the hEDS criteria in order to improve clinical diagnoses as well as to refine and homogenise research cohorts (Malfait, Francomano, et al. 2017), which could facilitate the discovery of a genetic mechanism (Tinkle, Castori, et al. 2017). These criteria also clarified the nosological name of the disorder by merging hEDS with clinically indistinguishable conditions such as Joint Hypermobility Syndrome (Castori, Dordoni, et al. 2014, Castori, Tinkle, et al. 2017, Malfait, Francomano, et al. 2017). For these reasons, the new criteria for hEDS are limited only to individuals showing widespread manifestations or clear Mendelian inheritance (Castori and Hakim 2017). This results in the exclusion of patients previously diagnosed with hEDS using outdated criteria (McGillis, Mittal, et al. 2020, Castori, Tinkle, et al. 2017).

In order to resolve this conundrum, the authors behind the 2017 international classification (Malfait, Francomano, et al. 2017) proposed a continuous spectrum of hypermobility (**Figure 1.4**), with asymptomatic JH at the mild end and diagnosed hEDS at the more severe end.

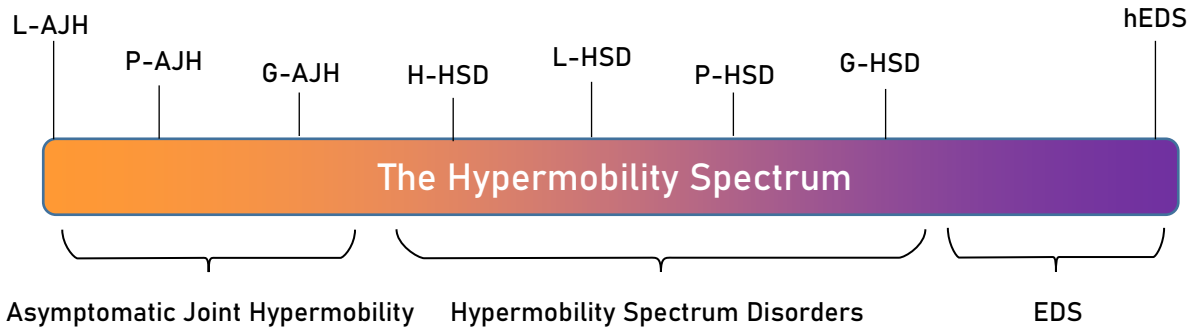


Figure 1.4: A diagram depicting the hypermobility spectrum and its associated disorders
L-AJH: Localised Asymptomatic Joint Hypermobility, P-AJH: Peripheral Asymptomatic Joint Hypermobility, G-AJH: Generalised Asymptomatic Joint Hypermobility, H-HSD: Historical Hypermobility Spectrum Disorder, L-HSD: Localised Hypermobility Spectrum Disorder, P-HSD: Peripheral Hypermobility Spectrum Disorder, G-HSD: Generalised Hypermobility Spectrum Disorder, hEDS: Hypermobile Ehlers-Danlos Syndrome, EDS: Ehlers-Danlos Syndrome. Adapted from text (Castori et al., 2017) and image found on https://www.hypermobilityconnect.com/the-spectrum-of-hypermobility/ (site accessed on 5/5/2020).

Patients who present with only some of the mandatory features associated with hEDS were classified as having 'Generalised Hypermobility Spectrum Disorder' (G-HSD; Castori et al., 2017), one of the 'Hypermobility Spectrum Disorders' (HSD) which reside in the middle of the said spectrum. In hEDS, intrafamilial phenotypic variation has been reported with individuals from the same family presenting with hEDS and others presenting with isolated JH and HSD (Castori, Tinkle, et al. 2017).

1.4.4 Clinical Presentation

Besides the well-known musculoskeletal and dermal manifestations, hEDS includes various other features in its clinical phenotypic spectrum (Baeza-Velasco, Bourdon, et al. 2018, Rombaut, Malfait, et al. 2010). As the knowledge acquired so far on this disorder remains mostly clinical, evaluation of the major symptoms described in the literature is key for the recruitment of adequate patient cohorts, in addition to deducting a causative molecular mechanism. For this purpose, the documented major effects of hEDS will be reviewed below.

However, it is important to note that since most research studies were conducted before the updated 2017 classification, symptoms described in these cohorts may be applicable to hEDS patients and to other individuals who do not fit the new criteria. Therefore, studies conducted prior to 2017 must be interpreted with a degree of caution.

1.4.4.1 Muscular tone in hEDS

Patients diagnosed with hEDS present with muscle weakness, decreased limb muscle strength and low tolerance to stress when compared to controls in whom there are no significant differences in muscle mass (Rombaut, Malfait, De Wandele, Taes, et al. 2012). A prolonged period of muscle pain and fatigue after exertion is also observed which consequently, impacts patients' functional abilities (Rombaut, Malfait, et al. 2012, Scheper, Rombaut, et al. 2017). Rombaut et al. suggest that a qualitative defect in the muscular ECM, which affects the transfer of force through the muscle, is the probable underlying aetiological factor for these findings (Rombaut, Malfait, et al. 2012).

1.4.4.2 Effect on skin and oral mucosa

Most hEDS patients have smooth, mildly hyperextensible velvety skin. (Castori, Dordoni, et al. 2015). There is impaired and prolonged wound healing (Castori, 2012), with the formation of atrophic, non-papyraceous scars (Castori, Dordoni, et al. 2015). An increased prevalence of stretch marks (*striae distensae*) which may be red in appearance, is also documented (Castori 2012, Castori, Dordoni, et al. 2015). Although commonly found in the general population, an increased rate of piezogenic papules (i.e., subcutaneous lumps caused by herniation of fat upon pressure) is found in 60-63% of hEDS patients (Castori, Camerota, et al. 2010, Castori, Dordoni, et al. 2015).

These clinical presentations are reinforced by morphological abnormalities observed in skin biopsies from hEDS individuals (Hermanns-Lê and Piérard 2007, Hermanns-Lê, Reginster, et al. 2012). Using transmission electron microscopy, an irregular assembly of collagen fibrils

with an inconsistent fibril diameter which occasionally forms a ‘flower-like’ cross-sectional conformation, was observed (Hermanns-Lê and Piérard 2007, Hermanns-Lê, Piérard, et al. 2015). Moreover, the authors note an accumulation of hyaluronic acid, a major ECM component involved in inflammation and tissue repair (Hermanns-Lê and Piérard 2007).

1.4.4.3 Pain and anaesthetic resistance

Despite the fact that pain is a widely documented symptom in all EDS subtypes, hEDS patients report a high severity and frequency of painful stimuli (Rombaut, Malfait, De Wandele, Mahieu, et al. 2012, Scheper, Rombaut, et al. 2017), which can be attributed to the high degree of hypermobility and joint trauma (Voermans, Knoop, et al. 2010). Most studies on hEDS note that a high, if not absolute percentage of their cohorts, suffer from chronic pain (Bénistan and Martinez 2019, Palmer, Manns, et al. 2017). The commonest pain areas are the spine (90% of participants), knees (86%), shoulders (84%), hips (83%), neck (80%), hands (79%), and wrists (78%) (Palmer 2014) although other sources state that the weight-bearing lower limbs are the most affected (Bénistan and Martinez 2019, Kumar and Lenert 2017, Tinkle and Levy 2019). Additionally, abdominal pain and headaches were also documented in a large percentage of individuals (Bénistan and Martinez 2019, Castori, Camerota, et al. 2010). The severity and frequency of pain in hEDS patients follow an accumulative pattern, with localised occasional pain becoming more frequent, widespread and severe over time (Bénistan and Martinez 2019, Castori, Morlino, et al. 2013, Voermans, Knoop, et al. 2010). hEDS patients also show partial or complete failure of response to local anaesthetics (Arendt-Nielsen, Kaalund, et al. 1991, Bénistan and Martinez 2019, Castori, Camerota, et al. 2010, Hakim, Grahame, et al. 2005) which response is short-lived when compared to controls (Arendt-Nielsen, Kaalund, et al. 1990). This may result from the drainage of the anaesthetic agent from the dermis due to aberrant connective tissue structure (Cesare, Rafer, et al. 2019).

1.4.4.4 Fatigue

Chronic fatigue is reported in 80-88% of hEDS patients (Celletti, Galli, et al. 2012, Krahe, Adams, et al. 2018, Voermans, Knoop, et al. 2010), reaching a peak in older patients (Castori, Sperduti, et al. 2011). Chronic fatigue is regarded as experiencing tiredness or lack of energy for more than 6 months (Hakim, De Wandele, et al. 2017), resulting in a diminished ability to perform daily activities (Celletti, Castori, et al. 2013, Tinkle, Castori, et al. 2017). Fatigue from minimal physical exertion affects gait, posture, and balance, which in turn amplifies muscle weakness and the risk of accidents (Tinkle, Castori, et al. 2017). Interestingly, De Wandele et al. alluded to the hypothesis that fatigue is a consequence of elevated cytokine levels from inflammatory processes related to recurrent joint trauma (De Wandele, Rombaut, et al. 2013), which overstimulate the central nervous system. Similar fatigue and pain patterns as well as structural abnormalities in connective tissue (Hermanns-Lê, Piérard, et al. 2013) have been noted in hEDS and fibromyalgia (Castori, Celletti, et al. 2011, Hermanns-Lê, Piérard, et al. 2013, Ofluoglu, Gunduz, et al. 2006). However, due to the clinical overlap between the two conditions as well as the lack of confirmatory diagnostic tests, this similarity can also be explained by the frequent misdiagnosis of hEDS individuals as having fibromyalgia (Chopra, Tinkle, et al. 2017), leading to the inclusion of these patients in fibromyalgia cohorts.

1.4.4.5 Gastrointestinal dysfunction

Amongst the EDS subtypes, gastrointestinal (GI) symptoms are predominantly associated with hEDS, due to the high attendance of hEDS patients at GI clinics (Brockway 2016). In a study of 170 hEDS female patients, GI involvement emerged as a significant non-musculoskeletal complaint (Hakim and Grahame 2004). This observation was later corroborated by other studies (Castori, Camerota, et al. 2010, Castori, Morlino, et al. 2015, Farmer, Fikree, et al. 2014, Fikree, Grahame, et al. 2014, Zarate, Farmer, et al. 2010).

Castori et al. described functional GI symptoms reported by hEDS patients as chronic GI discomfort, gastroesophageal reflux, bloating & nausea, abdominal pain, constipation, and diarrhoea (Castori, Camerota, et al. 2010). Similar presentations were also reported in larger studies (Fikree, Grahame, et al. 2014, Fikree, Aktar, et al. 2015, Fikree, Aktar, et al. 2017). Most of these features can be attributed to dysmotility, i.e., the altered movement of contents through the GI tract (Menys, Keszthelyi, et al. 2017). Structural GI abnormalities are observed as abdominal or hiatal hernias and rectal prolapse (Castori, Morlino, et al. 2015, Tinkle, Castori, et al. 2017). These manifestations are hypothesised to be due to an abnormal connective tissue structure (Brockway 2016), which allows distensibility of the digestive organs thus reducing the efficacy of sphincters and supporting ligaments (Castori, Morlino, et al. 2015). Additionally, Fikree et al. (2017) noted that collagen is involved in the differentiation of nervous tissue in the gut, while CT abnormalities might contribute to nervous dysregulation of the GI tract (Fikree, Aktar, et al. 2017).

1.4.4.6 Dysautonomia

The autonomic nervous system (ANS) connects the central nervous system to the visceral organs as well as smooth and cardiac muscle (Wehrwein, Orer, et al. 2016). These motor neurons regulate physiological processes (such as regulation of blood pressure and volume, breathing, digestion, immune response and hormone secretion) independent from voluntary control (Gibbons 2019, Wehrwein, Orer, et al. 2016). Autonomic dysregulation and/or failure is associated with a plethora of clinical disorders (Wehrwein, Orer, et al. 2016) which is also inclusive of EDS and hEDS (Hakim, O'Callaghan, et al. 2017).

Dysautonomia in hEDS mainly manifests as postural orthostatic tachycardia syndrome (POTS) which is present in around 34-50% of individuals (Celletti, Camerota, et al. 2017, De Wandele, Rombaut, et al. 2014, Gazit, Nahir, et al. 2003, Hakim and Grahame 2004). POTS is defined by an increase in heart rate of ≥30 bpm (≥40 bpm in paediatric patients) without a decrease in

blood pressure upon minutes of standing upright, causing dizziness, intense headaches, palpitations, nausea, 'brain fog' and fatigue (Kohn and Chang 2019, Miller, Stiles, et al. 2020).

1.4.4.7 hEDS and immune function

In recent years, the high rate of rheumatic disorders in hEDS patients is gaining more interest, (Casanova, Sharp, et al. 2018, Cheung and Vadas 2015, Chiarelli, Ritelli, et al. 2019, Daens, Grossin, et al. 2018, Lee and Mueller 2017, Lyons, Yu, et al. 2016, Seneviratne, Maitland, et al. 2017). Of particular note is the link between hEDS and mast cell activation disorder (MCAD), which share dysautonomic symptoms such as fatigue, gastrointestinal dysfunction, postural intolerance, headaches and palpitations (Kohn and Chang 2019).

MCAD is described as an increased activity of mast cells which are a type of white blood cells found in connective tissues. Mast cells degranulate upon injury, initiating the inflammatory cascade by releasing immune mediators (Seneviratne, Maitland, et al. 2017). Mast cells are usually associated with allergic disorders such as asthma and food allergies (Seneviratne, Maitland, et al. 2017). However, mast cell dysfunction could affect nearby nerve fibres and blood vessels, leading to dysautonomia (Jesudas, Chaudhury, et al. 2019). The effect of these mediators is consistent with widespread symptomatic presentation (similar to that of hEDS), and thus may affect the development of connective tissue (Afrin 2016). Further studies pertaining to the rate of MCAD in hEDS individuals as well as the effect of mast cell mediators on connective tissue are required in order to test these hypotheses (Kohn and Chang 2019).

1.4.4.8 Psychological impact

In addition to debilitating physical manifestations, a high rate of psychological and behavioural conditions are also observed in hEDS individuals, most commonly described as anxiety, depression and panic/fear disorders (Baeza-Velasco, Bourdon, et al. 2018, Berglund, Pettersson, et al. 2015, Bulbena, Baeza-Velasco, et al. 2017, Murray, Yashar, et al. 2013, Sinibaldi, Ursini, et al. 2015). A meta-analysis of the psychiatric literature pertaining to hEDS

confirmed this association (Smith, Easton, et al. 2014). Fear related to pain from movement and potential injury (kinesiophobia) is also recurrent in hEDS individuals (Celletti, Castori, et al. 2013, Rombaut, Malfait, et al. 2011, Scheper, Rombaut, et al. 2017).

1.4.4.9 Cardiovascular effects

An increased susceptibility to bruising upon minimal trauma is manifested by hEDS patients (Castori, Camerota, et al. 2010). Since coagulation laboratory results are usually within normal ranges (De Paepe and Malfait 2004, Mast, Nunes, et al. 2009), cardiovascular effects in hEDS can be attributed to capillary fragility from a CT defect rather than a clotting factor deficiency (Gazit, Jacob, et al. 2016). Contrary to vascular EDS (vEDS) and other HCTDs, arterial dissections are not reported in hEDS (Gazit, Jacob, et al. 2016), although mitral valve prolapse and mild dilation of the aortic root have been described in a minority of patients (Asher, Chen, et al. 2018, Atzinger, Meyer, et al. 2011). These anomalies were shown to be non-progressive and clinically insignificant (Atzinger, Meyer, et al. 2011, Ritter, Atzinger, et al. 2017).

1.4.4.10 Orthopaedic manifestations

Besides the manifestations described above, the most evident musculoskeletal symptom in hEDS remains generalised joint hypermobility, which causes joint trauma and recurrent dislocations including dislocation of the temporomandibular joints. Similarly, hEDS patients are able to open their jaw wider, with consequent dislocation and ensuing pain, difficulties in mastication and jaw locking (Ancillao, Galli, et al. 2012, Castori 2012). It is hypothesised that joint laxity in hEDS impacts the formation of the skeleton *in utero* and even after birth (Castori 2012), resulting in minor, structural orthopaedic manifestations (**Figure 1.5**). These are observed mainly in the upper and lower limbs (such as fixed subluxations, hind-foot pronation and pes planus, but may also include scoliosis, a high-arched/narrow palate and facial asymmetry, amongst others (Castori 2012, Cimolin, Galli, et al. 2014).

Figure 1.5: Orthopaedic complaints in patients with hEDS
Images show characteristic joint hypermobility at the fingers (A), toes (B), elbow (C) and knees (D). Passive hyperextension of the large toe (E) and ankle (F) is also observed. Consequent joint instability may result in fixed subluxations of the distal ulna (G), elbow (H) and first metacarpal (I). Hindfoot pronation (J) and hallux valgus (K) may also be present. Image obtained from Castori, 2013.

The degenerative effect of hypermobility and associated joint trauma on bone structure in EDS patients are cumulative throughout life (Castori, Tinkle, et al. 2017). This is however supplemented by the effect of congenital deficits in collagen, which is a major component of bone (Guarnieri and Castori 2018). Type I collagen constitutes up to 90% of the bone organic ECM (Guarnieri and Castori 2018) thus the quantitative and qualitative integrity of collagen fibrils ensures adequate regulation of bone metabolism and structure (Mazziotti, Dordoni, et al. 2016). Consequently, defects in type I and type V collagen in EDS disturb these processes and may impact bone mass thereby increasing susceptibility to fractures (Guarnieri and Castori 2018).

EDS patients demonstrated lower bone mineral density (BMD) at the lumbar spine (LS) (Annapureddy, Block, et al. 2014, Dolan, Arden, et al. 1998, Eller-Vainicher, Bassotti, et al. 2016, Theodorou, Theodorou, et al. 2012), femoral neck (FN) (Annapureddy, Block, et al. 2014, Dolan, Arden, et al. 1998, Eller-Vainicher, Bassotti, et al. 2016, Gulbahar, Sahin, et al.

2006), trochanter (Dolan, Arden, et al. 1998, Gulbahar, Şahin, et al. 2006) and total body (Yen, Lin, et al. 2006) when compared to controls. In two studies, this was accompanied by a higher fracture risk (Dolan, Arden, et al. 1998, Eller-Vainicher, Bassotti, et al. 2016) with no significant discrepancy in 25-hydroxyvitamin D (25-OH vitamin D), calcium and phosphorus levels between EDS patients and controls (Eller-Vainicher, Bassotti, et al. 2016). Conversely, Carbone et al. found no significant difference in BMD and bone-specific alkaline phosphatase (BAP), osteocalcin (OC), the carboxyterminal peptide of type I collagen (PICP), aminoterminal propeptide of type III procollagen (PIIINP) and serum carboxyterminal cross-linked telopeptide of type I collagen (ICTP) between 23 hEDS patients and non-hEDS controls after adjusting for height, weight and physical activity (Carbone, Tylavsky, et al. 2000). Additionally, Mazziotti et al. (2016) found no significant difference in BMD at the LS and FN. However, hEDS participants showed lower FN BMD compared to cEDS and controls. Fractures were also more prevalent in individuals having various EDS subtypes when compared to controls despite having a normal BMD, indicating a qualitative deficit rather than a quantitative one (Mazziotti, Dordoni, et al. 2016).

Literature on the prevalence of BMD and fracture risk in EDS and hEDS is scant (Levy 2018) and in most cases, predates the updated 2017 diagnostic criteria. Most study cohorts are small and do not focus on hEDS only but include multiple EDS subtypes, which in turn are caused by different causal mechanisms. Thus, it is speculative to consider these research findings as canon (Tinkle, Castori, et al. 2017). Ideally, the association between hEDS and its orthopaedic effects should be enhanced by recruiting cohorts using the contemporary diagnostic criteria for hEDS with accompanying biochemical and molecular studies to determine the effect of pleiotropic variants on these measurements. In addition to the gold-standard Dual-energy X-ray absorptiometry (DXA) scan, techniques such as high-resolution peripheral quantitative

computed tomography (HR-pQCT) may be utilised in order to assess the microarchitecture of cancellous and cortical bone in hEDS.

1.4.5 Management

The fact that the underlying genetic aetiology of hEDS remains unknown further contributes to the lack of standardised treatment availability for these patients. However, lack of standardised treatment should not be a deterrent for diagnosis. Early diagnosis may be crucial in reducing pain and increasing the patient's quality of life (Fikree, Aktar, et al. 2017, Scheper, de Vries, et al. 2015). Management of the disorder should be multidisciplinary and tailored to each individual according to his/her clinical phenotype (Castori 2012) which may include acute symptoms such as joint dislocations as well as chronic complaints (Tinkle, Castori, et al. 2017). Pain may never be completely resolved (Tinkle, Castori, et al. 2017). Treatment should focus on ameliorating symptoms and preventing further injury, usually by referring patients for physiotherapy, psychological counselling and by prescribing analgesics (Bénistan and Martinez 2019, Castori 2012, Cohen and Markham 2017).

1.5 Biomarkers

Research on serum biochemical markers in hEDS patients is scarce and features serum determination of Tenascin X (TN-X; Zweers, Bristow, et al. 2003, Yamada, Watanabe, et al. 2019). TN-X, coded by the *TNXB* gene on chromosome 6, is a 450-kilodalton extracellular matrix glycoprotein which is ubiquitously expressed at high concentrations in cardiac, dermal and musculoskeletal tissue as well as in the kidney, lung, adrenals, reproductive organs, vascular tissue and the gastrointestinal tract (Valcourt, Alcaraz, et al. 2015). Due to its interaction with several members of the collagen family as well as other matrix proteins such as Transforming Growth Factor-beta (TGFβ) and Decorin, TN-X is crucial in the assembly of collagen fibrils, regulation of collagen deposition by fibroblasts and maintenance of ECM

architecture by preserving the integrity of elastic fibres. Consequently, this protein also plays a vital part in embryonic organogenesis, wound healing, cell migration, cell differentiation and tumour suppression (Petersen and Douglas 2013, Valcourt, Alcaraz, et al. 2015). Zweers et al. (2003) reported significantly reduced TN-X levels in the sera of 20 hEDS research participants whilst identifying insertion and deletions in the *TNXB* gene in 17 of these individuals (Zweers, Bristow, et al. 2003, refer to section 1.6.2). Additionally, upon measuring serum TN-X concentrations in 17 hEDS patients using nano-liquid chromatography tandem mass spectrometry (LC/MS/MS), 9 individuals had significantly low TN-X levels when compared to unaffected controls (Yamada, Watanabe, et al. 2019).

A study by Watanabe et al. (2016), which included 18 hEDS patients and 13 controls, also reported significant differences between hEDS individuals and controls, in six serum proteins. The latter included complement C1r subcomponent (C1R), vitronectin (VTN), complement component C9 (C9), and C4b-binding protein α-chain (C4BPA) with the addition of apolipoprotein B-100 (APOB), and transthyretin (TTR). The majority of these proteins are associated with the complement system, suggesting the presence of an inflammatory process in hEDS (Watanabe, Satoh, et al. 2016).

In a study of 50 hEDS patients, 84% of the cohort had an elevated erythrocyte sedimentation rate (ESR) but no elevation was reported in C-reactive protein (CRP) or Anti-nuclear antibodies (ANA) (Tamilselvam and Malarvizhi 2017) levels. Fasting plasma glucose, lipid profile and rheumatoid factor (RF) were elevated in 22%-50% of patients. However, these results must be interpreted with caution due to the lack of a control group as well as the lack of information on the cut-off value used to determine altered levels. In a case study of a female hEDS patient, lab results of markers of haematological, hepatic, immune and thyroid function, lipid profile and basic metabolic panel were all reported as normal (Zhang, Tai, et al. 2005). In another study, low levels of 25-OH vitamin D were reported in 81% of hEDS participants, but this was not

correlated with fracture risk or low BMD (Mazziotti, Dordoni, et al. 2016). Since hypovitaminosis D has been linked to a higher risk of vertebral fractures especially when combined with secondary hyperparathyroidism (Hernández, Olmos, et al. 2013), the absence of parathyroid hormone (PTH) testing was cited as the cause for the lack of association between 25-OH vitamin D levels and fracture risk in this cohort.

1.6 Genetics of hEDS

The emergence of novel genetic techniques such as HTS has transformed the diagnosis of EDS, facilitating the discovery of new variants in genes not directly associated with collagen fibrillogenesis. Elucidating the molecular basis of these subtypes is crucial for an accurate early diagnosis and adequate management as well as for the recruitment of research cohorts, and future treatment strategies (Malfait, Francomano, et al. 2017). Thus, the next challenge for the EDS research community is the determination of the underlying genetic aetiology of hEDS, the most elusive and complex EDS subtype.

1.6.1 Mode of Inheritance

hEDS is observed to be inherited in an AD fashion, with few instances of generation-skipping and alternative inheritance patterns reported in some pedigrees (Castori, Dordoni, et al. 2014, Tinkle, Castori, et al. 2017). However, this may be due to the wide heterogeneity in clinical presentation including intrafamilial variation (Castori, Dordoni, et al. 2014). Additionally, in a recent cohort, 22% of hEDS participants did not have a family history of hEDS (Bénistan and Martinez 2019).

The clinical heterogeneity observed in hEDS has been attributed to incomplete penetrance, with variable symptom presentation and severity (or expressivity; Byers and Murray 2014); Castori, 2012; Castori, Camerota, et al., 2010; Castori, Dordoni, et al. 2014). Castori et al. proposed a threshold model, in which a variant contributes a genetic disposition to

hypermobility that may be supplemented by other genetic/environmental modifier factors such as gender, nutrition, physical activity and joint trauma, thus 'reaching the threshold' for the full manifestation of the disorder (Castori 2012, Castori, Sperduti, et al. 2011, Castori, Celletti, et al. 2013).

This variability in disease presentation is regarded as a major reason why a clear molecular mechanism for hEDS has not been found, despite numerous research efforts (see below; Byers and Murray 2012, Schubart, Schaefer, et al. 2019; Tinkle et al., 2017). The number of large, multi-generational families is restricted (Syx, Symoens, et al. 2015) thus limiting the feasibility of family studies by which the penetrance and transmission of any causative variant may be followed (Castori, Dordoni, et al. 2014). Apart from clinical heterogeneity, some authors have also alluded to the presence of genetic heterogeneity (Byers and Murray 2012, Chiarelli, Carini, et al. 2016, Cortini, Villa, et al. 2019, De Wandele, Rombaut, et al. 2013, Tinkle, Castori, et al. 2017, Zweers, Hakim, et al. 2004), which would explain why certain genes implicated in some studies are not replicated by others (Syx, Symoens, et al. 2015).

1.6.2 Molecular studies

Early literature on the molecular basis of hEDS was directly focused on the genes encoding collagen, as these defects were already known to cause other EDS subtypes. An early segregation study excluded the involvement of genes encoding collagen type III (*COL3A1*), type V (*COL5A2*) and type VI (*COL6A3*; (Henney, Brotherton, et al. 1992), with the author also stating a lack of correlation with *COL1A1* and *COL1A2*. Conversely, a glycine to serine substitution in the type III collagen, α-chain 1 gene (*COL3A1*) was found in one family with hEDS features (Narcisi, J.Richards, et al. 1994). However, since variants in type III collagen are usually found in the vascular subtype (Mizuno, Boudko, et al. 2013), these findings are instead thought to reflect a family with vEDS with mild or undeveloped cardiac complaints (Malfait, Hakim, et al. 2006). Furthermore, the gene coding for type V collagen (*COL5A3*) was

deemed unlikely to be associated with hEDS due to the absence of obvious pathogenic variants in 13 individuals (Hoffman, Dodson, et al. 2008).

Apart from the collagens, the ECM protein TN-X has also been associated with the pathogenesis of some EDS subtypes such as hEDS and the more defined classical-like EDS (clEDS). In clEDS, homozygous or compound heterozygous autosomal recessive (AR) variants in the gene coding for the TN-X protein (*TNXB*) result in either biallelic gene deletion or defective messenger RNA (mRNA) transcripts which undergo nonsense-mediated decay (Schalkwijk, Zweers, et al. 2001). Consequently, this leads to a complete deficiency of the TN-X protein. Sequencing and molecular analysis of *TNXB* are thus crucial in the diagnosis of clEDS (Malfait, Francomano, et al. 2017), which can be supplemented by quantification of the TN-X protein in serum (Demirdas, Dulfer, et al. 2017). Individuals with clEDS present with a hypermobile phenotype along with skin hyperextensibility, cardiac-valvular abnormalities and chronic pain, amongst others (Demirdas, Dulfer, et al. 2017, Malfait, Francomano, et al. 2017). The chromosomal loci for *TNXB* and *CYP21A2* overlap. Thus, deletions at these loci give rise to continuous *CYP21A2* and *TNXB* chimeric genes (Merke, Chen, et al. 2013, Schalkwijk, Zweers, et al. 2001). *CYP21A2* codes for 21-hydroxylase, an essential catalytic enzyme in the glucocorticoid and mineralocorticoid pathways. Perturbations in the synthesis of this protein lead to 21-hydroxylase deficiency, resulting in insufficient production of cortisol and/or aldosterone, also known as congenital adrenal hyperplasia (CAH; El-Maouche, Arlt, et al. 2017). The presence of this chimeric gene affects the level and function of both TN-X and 21-hydroxylase, resulting in a contiguous gene deletion syndrome termed CAH-X, which presents with symptoms seen in both CAH and hEDS (Merke, Chen, et al. 2013). In addition to clinical evidence supporting the link between TN-X and connective tissue disorders, the general functional implications of Tenascin X on the musculoskeletal system have also been confirmed in a mouse model (Voermans, Knoop, et al. 2010). TN-X knockout mice showed muscle

weakness, reduced collagen fibril diameter and elevated ECM turnover in skeletal muscle (Voermans, Knoop, et al. 2010).

Schalkwijk et al. (2001) first reported an association between TN-X deficiency caused by variants in *TNXB*, and EDS (Schalkwijk, Zweers, et al. 2001). These findings were substantiated by a later study, which demonstrated the presence of a hEDS-like phenotype in individuals with heterozygous variants affecting *TNXB* (Zweers, Bristow, et al. 2003). In contrast to the previous study, these novel variants in *TNXB* were inherited in an AD manner, resulting in haploinsufficiency of the gene product. Additionally, whilst members within the same families had the same *TNXB* variants, no two families in this study shared a common variant, demonstrating a high level of inter-familial locus heterogeneity. In a later study, Zweers et al. (2003) demonstrated the presence of three additional variants in *TNXB* which did not directly lead to haploinsufficiency but instead altered the morphology of elastic fibres. In this group, Zweers reported reduced serum TN-X levels in 7.5% and *TNXB* variants in only 2.5% of the cohort (Zweers, Bristow, et al. 2003). These variants were not replicated in another unrelated cohort (Zweers, Kucharekova, et al. 2005). However, the method used for genetic analysis was not well-defined in the latter cohort, with serum samples being collected instead of peripheral blood as required for DNA testing.

The lack of correlation between variants in *TNXB* and serum TN-X levels was also shown by Yamada et al. (2019) in which only one research participant had a heterozygous nonsense variant g.32039800G>A (p.Arg1653*; rs201184519) but had normal serum TN-X levels. In contrast to other studies on TN-X (Zweers, Dean, et al. 2005), this study did not assess the effect of this seemingly loss-of-function variant on elastic fibre morphology. A significant reduction in serum TN-X levels was also found in research participants without variants in *TNXB*, the cause of which was attributed to epigenetic changes, thus indicating the lack of correlation between serum TN-X levels and *TNXB* variants in hEDS. Since whole exome

sequencing (WES) rather than whole genome analysis was performed, this observation can also be explained by the presence of undetected variants in deep intronic or intergenic regions which could alter splicing and/or transcription regulation. This study, despite having comparable results, is distinguished from those conducted by Zweers et al. (Zweers, Dean, et al. 2005, Zweers, Bristow, et al. 2003, Zweers, Kucharekova, et al. 2005) as individuals were assessed using the 2017 diagnostic criteria and analysed using HTS. These studies indicate that a small subset of hEDS patients has variants in *TNXB*, yet the genetic aetiology of the remaining majority remains unclear (Watanabe, Satoh, et al. 2016).

In a comprehensive study by Syx et al., a large 3-generation Belgian family with hEDS was investigated (Syx, Symoens, et al. 2015), which consisted of affected and seemingly unaffected members. After excluding the involvement of genes encoding type I, III, V collagen and *TNXB*, a genome-wide linkage scan identified a linked region between 8p22 and 8p21.1. After Sanger sequencing did not determine any variants in candidate genes in this region, WES was performed in two family members. Filtering for variants in this locus and in-silico prediction tools narrowed the search to one missense variant: a histidine to glutamine substitution g.20110809G>T (p.His211Gln, rs755849215) in the leucine zipper, putative tumour suppressor-1 (*LZTS1*) gene. Upon sequencing of *LZTS1* in an additional cohort of 230 probands with hEDS, 3 more variants were identified: g.20112644G>C (p.His17Asp, rs368057069), g.20107439G>A (p.Arg529Trp, rs773191392) and g.20110693C>A (p.Ser250*). *LZTS1* is an unlikely candidate for hEDS as it is normally associated with the prevention of oncogenesis; however, it may be implicated in neuronal growth and migration (Kropp and Wilson 2012). The genetic heterogeneity in hEDS was highlighted by the lack of linkage of the identified locus in four other families, thus further studies are required to test the role of *LZTS1* and its significance in hEDS (Syx, Symoens, et al. 2015).

A recent study which assessed the general diagnostic efficacy of HTS panels in EDS found that most variants of unknown significance (VUS) identified in this study were found in hEDS participants (n=76; (Weerakkody, Vandrovcova, et al. 2016). However, these VUS include variants in *COL1A1*, *COL1A2*, *COL3A1* and *COL5A2*, which are usually associated with other subtypes rather than hEDS. Due to the fact that this study had a selection bias for vascular phenotypes, that it predates the 2017 criteria, as well as the lack of association between collagen genes and hEDS in other studies, these findings may be attributed to misclassification of the cohort.

1.6.3 Expression analyses

Transcription-wide expression analysis on hEDS fibroblasts was able to identify 46 up-regulated genes as well as 162 down-regulated genes when compared to controls (Chiarelli, Carini, et al. 2016). Genes with increased expression included genes involved in the immune response and redox process while reduced expression was observed in proteins regulating kinase activity, cell-cell adhesion, inflammation, and differentiation of osteoclasts. Perturbations of signalling transduction pathways associated with CT integrity were also noted, including abnormalities in TGFβ and Janus kinase-signal transducer and activator of transcription protein (JAK-STAT) signalling as well as the phosphoinositide-3-kinase–protein kinase B/Akt (PI3K-AKT) pathway. These abnormalities in gene expression were demonstrated by immunofluorescence analysis in fibroblasts, which exhibited intracellular retention of elastin, of type I, III and V collagen as well as a lack of structural fibres and receptors (**Figure 1.6**).

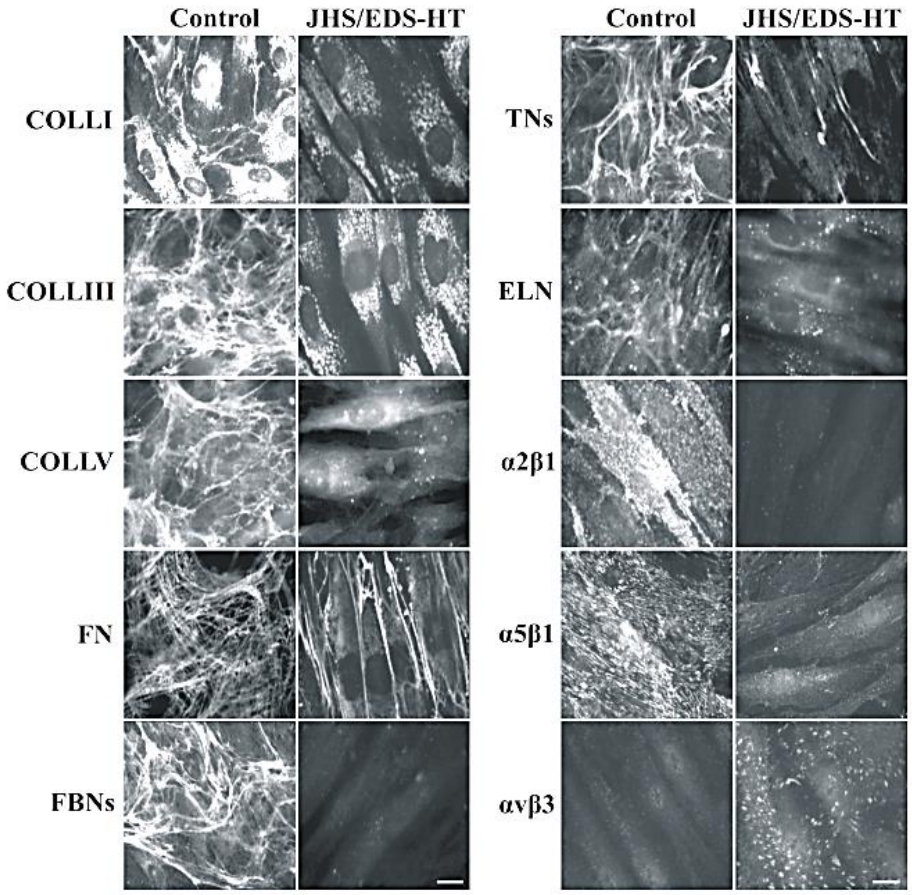


Figure 1.6: Immunofluorescence analysis of hEDS and control fibroblasts
Note the intracellular retention of collagens, the decreased quantity of fibrillins, elastin and the significant decrease of integrin proteins in hEDS patients when compared to the left panel. COLL1: collagen type I, COLLIII: collagen type III, COLLV: collagen type V, FN: fibronectin, FBNs: fibrillins, TNs: tenascins, ELN: elastin, α2β1, α5β1, αvβ3: integrin proteins. Image obtained from Chiarelli et al.,2016.

This altered assembly of ECM proteins lays the groundwork for the CT dysfunction observed in hEDS, which contributes to its phenotype (Chiarelli, Carini, et al. 2016). Interestingly, these perturbations are also observed in vEDS and cEDS fibroblasts (Zoppi, Gardella, et al. 2004), indicating that these abnormalities are not exclusive to a single molecular mechanism.

The cohort in Chiarelli, Carini, et al. (2016) was further investigated in a consequent study, which reported high levels of matrix metallopeptidase 9 (MMP-9), an ECM-degrading enzyme (Zoppi, Chiarelli, et al. 2018). The decreased concentration of fibronectin previously reported may be a result of enhanced MMP-9 expression which, in turn, generates fibronectin fragments and activates an immune response. Zoppi et al. (2018) also noted that hEDS fibroblasts undergo

a conformational change into myofibroblasts, adapting smooth muscle-like characteristics which potentially aid wound healing as well as generating an inflammatory response (Bagalad, Mohan Kumar, et al. 2017). After deducing that this transition is caused by the αvβ3 integrin-ILK-Snail1/Slug transduction pathway, Zoppi et al. conclude that hEDS stems from a chronic, inflammatory state (Zoppi, Chiarelli, et al. 2018). The presence and effect of upregulated immune mediators and reduced expression of inflammatory regulators in hEDS are further described in the following study (Chiarelli, Carini, et al. 2019). This chronic, inflammatory state is in line with clinical manifestations described in hEDS, notably chronic musculoskeletal pain, neurologic dysfunction, and GI disturbances (Chiarelli, Carini, et al. 2019).

1.7 Statement of the problem

As is evident in this review, the pathogenesis of hEDS is a complex phenomenon which is further aggravated by both clinical and genetic heterogeneity, nosological variations and multiple, unsubstantiated hypotheses. The diagnosis and assessment of hypermobile Ehlers-Danlos Syndrome (hEDS) patients remain a challenge due to the lack of a definite diagnostic biomarker. This is in contrast to the diagnosis of other EDS subtypes and hereditary connective tissue disorders, all of which utilise a genetic test for diagnosis. hEDS patients are often dismissed and misdiagnosed, leading to psychological distress and impeding referral to physiotherapy. Lastly, the phenotypic manifestations as well as the causative variants of this disorder have never been investigated in the Maltese population to date.

1.8 Aims & Objectives

This research study is based on the hypothesis that singular or multiple gene variants which perturb the integrity of connective tissue result in the protean clinical manifestations of hEDS. Thus, we sought to explore and identify gene variants which might be associated with the

aetiology of hEDS in local individuals, which would significantly improve the molecular understanding of hEDS. This was attempted by:

- Recruiting a multi-generational hEDS Maltese family to augment knowledge on the clinical manifestations and functional limitations of the condition in the local population.
- Identifying gene variants that might be associated with hEDS by analysing high-throughput sequencing data of recruited participants and assessing their clinical significance to the pathophysiology of the disorder.
- Evaluating the debilitating effect of hypermobility on bone-related phenotypes by BMD measurements and analysing biochemical markers of bone metabolism.

Chapter 2: Methodology

2.1 Recruitment of Research Participants

2.1.1 Selection of participants

A multigenerational family visiting the Medical Genetics Outpatients clinic at Mater Dei Hospital was recruited for this study between 2020-2021 (**Figure 2.1**). The proband was initially being investigated for Hereditary Haemorrhagic Telangiectasia (HHT). A physical examination at the time of her visit to the genetics clinic revealed that she was hypermobile and that she had other features of hEDS. Given the AD of inheritance of HHT and hEDS, family members at risk of having the same diagnoses were offered appointments to attend the genetics clinic. Following informed consent, six first-degree relatives were eventually recruited for this study. Relatives at risk, residing abroad and no longer in contact with the rest of the family, could not be recruited. During the recruitment process, it was important to include an unaffected family member in order to facilitate the variant filtering process (see Section 2.3.6). An unrelated hEDS patient attending the genetics clinic (SDT-005) was also recruited for this study. Ethical approval for this research project was granted by the Faculty Research Ethics Committee (FREC) prior to commencing the recruitment of participants (Application no. 4307-24022020).

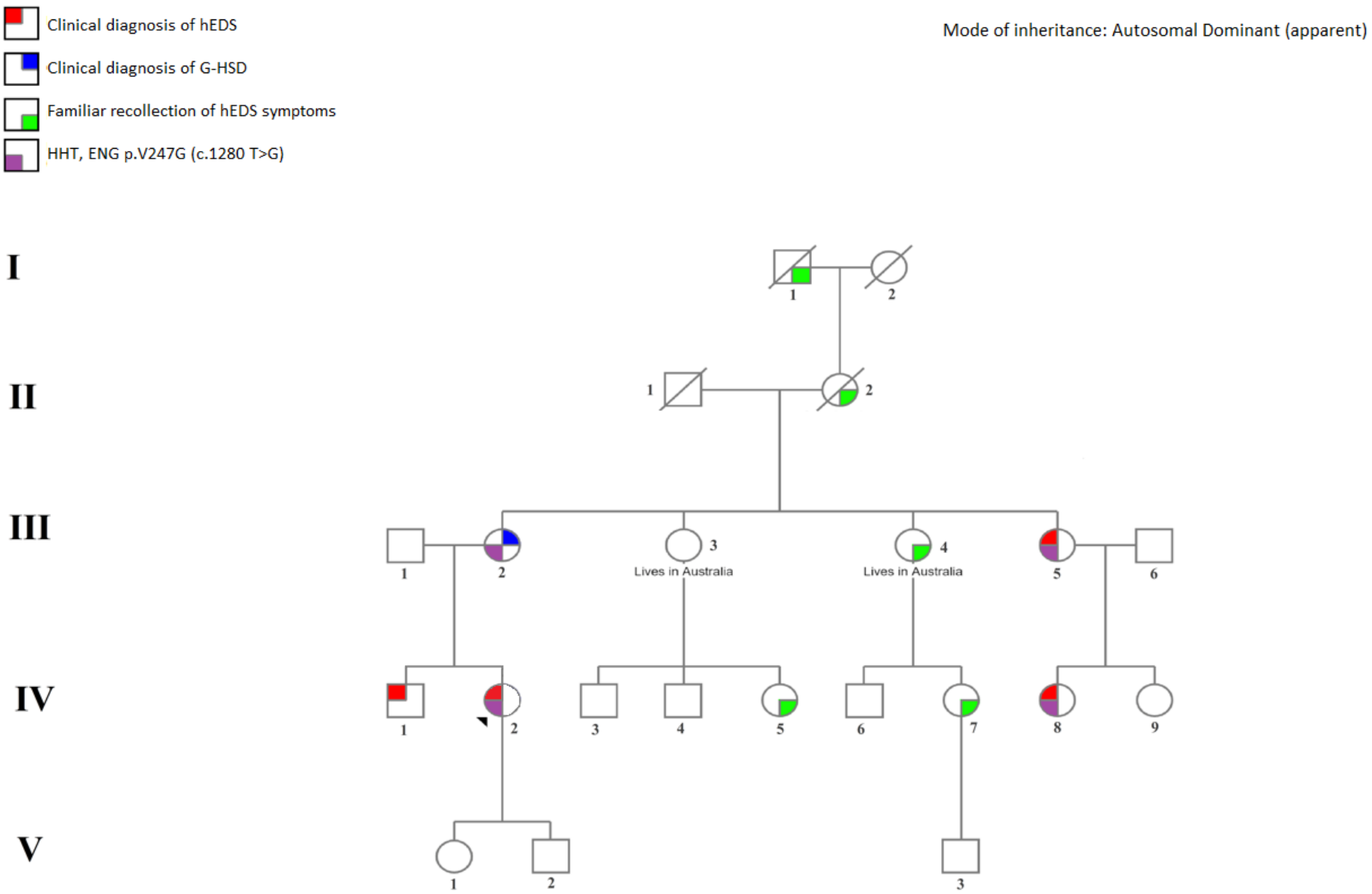


Figure 2.1: Family pedigree

A figure showing the family tree of recruited participants, their phenotypes and HHT genotype, if available. Proband (IV.2) is indicated by a black arrow. Individuals III.2, III.5, IV,1, IV.2, IV.8 and IV.9 were recruited to participate in this study. G-HSD= Generalised Hypermobility Spectrum Disorder, hEDS= hypermobile Ehlers-Danlos Syndrome, HHT= Hereditary Haemorrhagic Telangiectasia.

2.1.2 Clinical examination

The seven research participants attended the genetics clinic for a full consultation with their consultant medical geneticist. At the time of the consultation, the 2017 Diagnostic Criteria checklist provided by the International EDS Society was used in order to assess the research participants regarding their hypermobility (Beighton score, Beighton, Solomon, et al. 1973), history of joint dislocations, presence of pain as well as other cutaneous, skeletal and cardiac features amongst others (Appendix A). A full physical examination was also carried out.

2.1.3 Medical History

Recruited participants were also subjected to an interviewer-led questionnaire (Appendix B) which included detailed questions on systemic complications which have been described in the literature, but were not included in the hEDS Diagnostic Criteria checklist, as well as the functional limitations resulting from these symptoms.

2.1.4 Blood collection and serum measurements

Peripheral blood samples from all research participants were collected in vacutainers containing sodium ethylenediaminetetraacetic acid (EDTA) and later used for DNA extraction and genetic analysis (refer to 2.2.1). Blood samples were also collected for the following tests: 25-OH vitamin D, ionised calcium, serum albumin, serum protein, serum phosphorous, serum magnesium, liver profile including total bilirubin, alanine aminotransferase (ALT), aspartate aminotransferase (AST) , gamma-glutamyl transferase (GGT) and alkaline phosphatase (ALP) as well as plasma PTH. These tests were selected in consideration of the skeletal involvement in hEDS, as well as reports in the literature indicating low bone density values in this condition (Eller-Vainicher et al., 2016; Gulbahar et al., 2006; Mazziotti et al., 2016). These samples were immediately sent to the Pathology laboratories at Mater Dei Hospital for analysis (**Table 2.1**), whilst the EDTA samples were stored at -80°C at the Faculty of Health Sciences, University of Malta.

Table 2.1: Measurement of serum analytes in hEDS participants
A table showing the type of vacutainer in which the blood was collected, the analyser used to measure the analytes as well as the technique employed for each test. EDTA= Ethylenediaminetetraacetic acid, SST= Serum-separating tube.

Analyte	Collection tube	Analyser	Technique
25-OH vitamin D	SST	Siemens Atellica IM1600	Chemiluminescence
Albumin	SST	Cobas C501	Immunoturbidimetric assay
Alkaline phosphatase	SST	Cobas C501	Colourimetric assay
Ionised Calcium	SST	Diestro 103AP	Ion-selective electrode
Magnesium	SST	Cobas C501	Colourimetric-endpoint method
Parathyroid hormone	EDTA	Siemens Immulite 2000	Chemiluminescence
Phosphorous	SST	Cobas C501	Ammonium molybdate UV spectrophotometry

2.1.5 Bone Mineral Density Assessment

An appointment for a BMD scan was given to all recruited participants in order to assess if any deterioration of the bone microarchitecture is present and if it can be correlated with hEDS. The BMD (g/cm^2), T-scores and Z-scores were measured at the LS, consisting of measurements at L2, L3 and L4, as well as at the FN and total hip (TH). Acquisition of BMD data was performed using the Horizon DXA densitometer (Hologic, USA) for participants IV.1,IV.2,III.5,IV.8 and SDT-005, while a Discovery DXA densitometer (Hologic, USA) was used for III.2 as the BMD assessment was performed at the Gozo General Hospital rather than at Mater Dei Hospital.

2.2 Sample processing for genetic testing

In order to obtain DNA for HTS and further variant confirmatory tests, whole blood samples need to be treated to extract genetic material from peripheral blood leucocytes. A total of 7

samples were processed to obtain high-quality DNA using a modified salting-out method (Miller, Dykes, et al. 1988).

2.2.1 DNA extraction

Several methods can be employed to extract DNA from peripheral blood leucocytes, each with their respective advantages and limitations pertaining to yield, cost, safety considerations and time for processing. The method employed in this study is a modified salting-out technique originally proposed by Miller et al. (Miller, Dykes, et al. 1988). The main principle of this technique is the separation of aqueous DNA from more-hydrophobic cellular contaminants, which precipitate in high-salt concentrations (Chacon-Cortes and Griffiths 2014). An erythrocyte-lysing buffer (ELB) is first used to eliminate non-nucleated red blood cells, which is followed by lysis of leucocytes and protein degradation using proteinase K and sodium dodecyl sulphate (SDS). DNA is then separated from the cell lysate using a high salt solution and later precipitated using an ethanol solution. The method described below uses an increased volume of proteinase K for better digestion of the cell lysate. Despite being more time-consuming than its alternatives, this technique was chosen as it is efficient, gives a good yield, is cost-effective and does not involve the use of hazardous reagents.

In order to minimise contamination, DNA extraction was carried out in an area designated for handling of DNA specimens. Protocols and volumes used in the preparation of the buffers mentioned below can be found in Appendix C. Approximately 2ml of EDTA-preserved blood from each sample was transferred to a sterile 15ml conical bottom tube, to which 9ml of ELB was added. The solution was kept at 4°C for 15 minutes with occasional mixing before centrifugation at the same temperature for 5 minutes at 1048*g* (gravitational force equivalent). After centrifugation, the supernatant was discarded, and the leucocyte cell pellet was washed with 9 ml of ELB. The re-suspended cell pellet was incubated for a further 15 minutes at 4°C and centrifugation was repeated under the same conditions used previously. This second wash

ensures that the cell pellet is devoid of any erythrocytes and residual cellular debris. Additional washes with ELB followed by centrifugation were required for some samples in order to obtain a clean cell concentrate. The cell pellets were then re-suspended in 3ml pH-adjusted SE buffer containing Proteinase K and 1% SDS and incubated overnight in a water bath at 37°C.

The following day, 1 ml of 6M sodium chloride solution was added to each sample prior to agitating the solution on a rotator for 30 minutes, with brief vortexing every 10 minutes. Degraded proteins and cellular debris were then precipitated by centrifuging solutions at 1048*g* for 15 minutes at 18°C. The aqueous portion containing dissolved DNA was aspirated using a sterile Pasteur pipette and transferred to a sterile conical bottom screw-cap tube, taking care not to aspirate the precipitated protein layer. A volume of 100% ethanol was added to the aqueous solution in a 2:1 ratio to precipitate the DNA upon inversion of the tube. The DNA threads were then transferred to a sterile 1.5ml micro-centrifuge tube and washed in 1minute with 0% ethanol. The mixture was then centrifuged for 2 minutes before discarding the supernatant. Another 1 ml of cold 70% ethanol was added to the residual DNA and centrifuged again before decanting the supernatant. The DNA pellet was then air-dried for two hours, preventing the possibility of contamination by inverting the micro-centrifuge tubes on absorbent tissue paper. After drying, 100 µl of autoclaved Tris-EDTA (TE) buffer was added to the pellet, with the resulting mixture left to dissolve overnight on constant rotation. The DNA solution was then stored at 4°C until the quantity and quality of the extracted DNA could be confirmed by spectrophotometry.

2.2.2 DNA quantification and purity check

In order to verify that the DNA extraction protocol was successful and to ensure that the genetic material obtained was adequate to be used for consequent downstream applications such as HTS and variant confirmatory testing, the concentration and purity of extracted DNA was determined. Sample purity was checked using the Nanodrop 2000 UV Spectrophotometer

(Thermo Fisher Scientific Inc., Massachusetts, USA) while sample quantification was assessed using Qubit 3.0 fluorometer (Thermo Fisher Scientific Inc., USA).

DNA purity assessed by Nanodrop measures the absorbance of the tested solution at different wavelengths. Nucleic acids characteristically absorb light at a wavelength of 260nm, while proteins and contaminants from DNA extraction absorb light at 280nm and 230nm respectively. By measuring the ratio of sample absorbance at 230nm, 260nm and that at 280nm, one can determine the relative purity of the DNA solution. A ratio of approximately 1.8:1 for 260/280nm is considered as a pure solution, while a ratio of 2-2.2:1 is considered satisfactory for 260/230nm. Lower 260/280nm and 260/230nm ratios indicate the presence of proteins and/or other contaminants retained from the DNA extraction process (Matlock 2015). To assess DNA purity, equipment was first blanked using TE buffer, identical to the TE buffer used to preserve the DNA for storage. This ensures that any readings are solely attributed to the purity of the DNA material. The pedestal of the spectrophotometer was wiped with a clean laboratory wipe before pipetting 1.2 µl of TE buffer, taking care not to aspirate any air bubbles and not to touch the pedestal with the pipette tip. A blank reading of 0-1ng/µl is considered adequate. After the equipment was blanked, the absorbance readings of the samples were measured consecutively using the same technique, thoroughly wiping the pedestal as previously described between samples to minimise cross-contamination.

The DNA samples were then quantified using the Qubit 3.0 fluorometer, which measures DNA concentration by measuring fluorescent signals emitted upon the selective binding of fluorescent dyes to nucleic acids. A working solution containing the fluorophore mixture and instrument buffer was prepared by pipetting 7µl of Qubit reagent to 1393µl of Qubit buffer, using a 1:200 ratio as stipulated by the manufacturer. Two 190µl volumes of the working solution were added separately to 10µl each of low-concentration and high-concentration calibration standards. Precisely 199 µl of this working solution was also added to 1µl of each

DNA sample. The standard and sample solutions were quickly vortexed, incubated for 2 minutes at room temperature and instantly quantified using the analyser.

All DNA samples were relatively cleansed of most impurities retained from the extraction process, at a concentration above the 30g/µl minimum stipulated by the sequencing company (**Table 2.2**). Samples were then sent for HTS.

Table 2.2: Sample concentration and purity after DNA extraction
Concentration was determined using the Qubit 3.0 fluorometer while sample purity, calculated as ratios of values obtained at different wavelengths, was obtained by Nanodrop spectrophotometer.

Sample	Concentration (ng/µl)	Absorbance 260:280nm	Absorbance 260:230nm
III.2	121.6	1.76	1.98
IV.1	42.7	1.73	1.32
IV.2	84.8	1.36	1.88
III.5	89.6	1.74	1.72
IV.8	52.8	1.68	1.80
IV.9	30.4	1.77	1.60
SDT-005	239.0	1.66	2.11

2.3 High-throughput Sequencing

Evolved from dideoxy synthesis described by Sanger et al. (Sanger, Nicklen, et al. 1977), HTS refers to the application of various second and third-generation sequencing technologies which determine the nucleotide sequence of millions of DNA fragments simultaneously (Slatko, Gardner, et al. 2018). The advent of HTS and its ability to sequence large genetic regions and/or the whole genome is considered a paradigm shift in multiple scientific fields, especially in the field of molecular genetics, leading to a more thorough understanding of evolution, health and disease (Ambardar, Gupta, et al. 2016). HTS can be applied to sequence either specific, targeted

genes of interest, the protein-coding regions (WES) or the entire genome (whole genome sequencing, WGS) which encompasses both intronic and exonic regions.

Several commercial sequencing platforms have since been developed (**Figure 2.2**) which differ in their methods of fragment preparation, cluster amplification and arrangement as well as in the detection of the nucleotide base sequence (Ambardar, Gupta, et al. 2016, Reuter, Spacek, et al. 2015, Slatko, Gardner, et al. 2018).

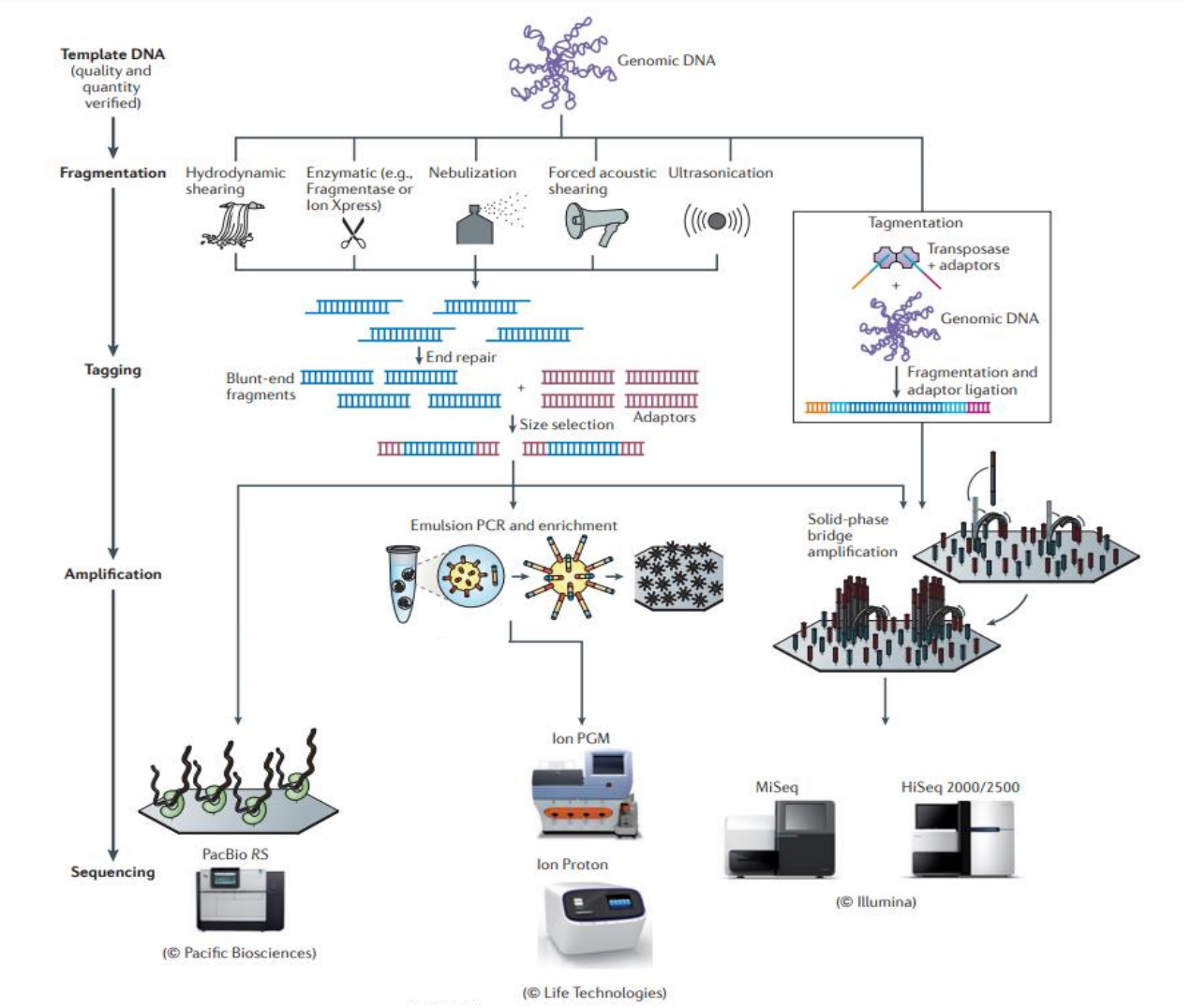


Figure 2.2: Second-generation sequencing technologies
An overview of the major sequencing platforms, showing the variation in techniques developed to sequence genomic DNA. Obtained from Loman, Constantinidou, et al. (2012) with permission from Springer Nature to reuse figure under license number 5579020523357.

Despite these differences, these sequencing platforms feature a similar workflow: 1. Library preparation by fragmentation of sample DNA, 2. clonal amplification of the DNA fragments,

3. cyclic parallel sequencing of fragments and detection to determine base sequence followed by 4. data analysis to check and process the massive amount of data obtained. In contrast with other established sequencing platforms namely those developed by Pacific Biosciences, Applied Biosystems' Ion torrent and Oxford Nanopore Technologies, the sequencing by synthesis approach developed by Illumina (Illumina Inc, California, USA) is considered to be more cost-efficient, yields lower error rates and lower false single nucleotide variant (SNV) calls, can be performed using lower quantities of DNA and also provides a higher number of reads per base (Quail, Smith, et al. 2012). In consideration of these advantages over other platforms, Illumina's sequencing by synthesis platform was chosen for this study and will be discussed in more detail further on.

2.3.1 Selection and preparation of samples

Out of the six recruited family members, four individuals (III.5, IV.1, IV.8 and IV.9) were selected for HTS. WGS was performed for III.5,IV.1 and IV.9 to avoid WES-associated limitations such as inefficient exon capture, incomplete sequencing, and limited number of genes on enrichment platforms (Syx, Symoens, et al. 2015). Selection of participants was based on choosing the ideal candidates according to the devised filtering strategy, as will be discussed below. Unrelated participant SDT-005 was also selected for WGS to test whether any discovered variants in the recruited family are present in other non-related individuals with a similar clinical presentation. Two of the four family members selected for HTS (III.5 and IV.8) had already been sequenced as part of their clinical investigation prior to this study, therefore this data was used, and their DNA samples were kept in storage for confirmatory testing.

HTS of these two individuals (III.5 and IV.8) was performed on an Illumina platform by Centogene (Rostock, Germany) as part of their medical investigations. Both samples were analysed on an Illumina platform, using the Twist Human Core Exome Plus kit for exomic analysis in one sample (IV.8). Data was processed with an in-house bioinformatics pipeline, in

which reads were aligned to the human genome build 37 (GRCh37) for reference and variants with a minimum 20x coverage depth were filtered (Prof. I. Borg, pers. comm. 9th January 2021). Since a rare HCTD was suspected to be present, variants having an alternative allele frequency (AAF) of ≤1% only were selected as it is improbable that a variant which is commonly occurring in the general population would be the causative factor in a rare disease. In consideration of the family history and clinical phenotype, all variants showing a strong genotype-phenotype correlation which are referenced in ClinVar, CentoMD® and the Online Mendelian Inheritance in Man (OMIM) databases are reported back to the requesting consultant medical geneticist. However, in III.5 and IV.8, the sequencing company did not report any relevant variants correlating to the individuals' phenotype. Thus, the raw sequencing data of these participants, as well as the Variant Call Format (.vcf) files, were obtained from Centogene with the research participants' permission in order to be used in this project.

An Illumina library preparation and sequencing platform was chosen for IV.1, IV.9 and SDT-005 to reduce any discrepancies in the preparation and capture methods between samples, thus facilitating the filtering strategy and reducing the risk of bias. WGS was performed at Theragen Bio (Gyeonggi-do, South Korea) to sequence these individuals, for whom no genetic data was available (**Table 2.3**).

Table 2.3: An overview of the type of sequencing, platforms and companies used for high-throughput sequencing
WGS: Whole genome sequencing, WES: Whole exome sequencing.

Sample ID	Sequencing company	Type of sequencing	Sequencing platform used
III.5	Centogene	WGS	Illumina
IV.8	Centogene	WES	
IV.1	Theragen Bio	WGS	
IV.9	Theragen Bio	WGS	
SDT-005	Theragen Bio	WGS	

The DNA samples were prepared for transportation to the sequencing company by pipetting a volume of DNA solution in a 1.5ml snap-lock tube, which was then topped up with a volume of TE buffer in order to obtain an adequate volume and concentration for sequencing (**Table 2.4).** As per the sequencing company's sample requirements, DNA samples were required to have a concentration of ≥30ng/µl in 100µl volume. The concentration of the samples as previously measured by Qubit was used rather than the values obtained by Nanodrop as fluorometric quantification is more specific and more accurate at low DNA concentrations than spectrophotometric measurements (Masago, Fujita, et al. 2021).

Table 2.4: Preparation of samples to be sent for HTS
A table showing the DNA concentration of each sample as well as the volume of DNA and the volume of TE buffer added for transportation.

Sample ID	DNA concentration by Qubit (ng/µl)	DNA solution (µl)	TE buffer (µl)
IV.1	42.7	35	65
IV.9	30.4	49	51
SDT-005	239	6	94

2.3.2 Library preparation

Library preparation is initiated by shearing sample DNA using Covaris' Adaptive Focused Acoustics® Technology, which uses ultrasonic acoustic energy to denature and cleave the genetic material. This process creates randomly sheared 350 or 550 base pair (bp) DNA fragments with 5' and 3' overhangs at each side of the fragment. These overhangs are then removed by 3' to 5' exonucleases to remove 3' overhangs, while 5' overhangs are removed by 5' to 3' polymerases. An adenine base is then added to blunt ends to prevent self-ligation of the DNA strands to each other rather than binding to specific short oligonucleotide adaptor sequence in the following step. Accordingly, a thymine base overhang on the adaptor end

ensures adequate binding of the DNA fragment. Before and following adaptor ligation, sample purification beads are used to clean ligated fragments (Illumina 2017).

In instances where the template DNA concentration is minute, Illumina also offers Nano DNA library preparation (**Figure 2.3**), which amplifies the ligated DNA fragments by Polymerase chain reaction (PCR, refer to section 2.4.1). A reduced number of 8 cycles are performed during PCR to avoid over-amplification of certain fragments, thus reducing fragment bias. If the initial DNA sample concentration is sufficient, PCR-free library preparation is usually preferred due to the reduction of guanine-cytosine (GC) bias of DNA polymerases, as well as lower rates of duplication and false positives (Broad Institute Genomics, 2018; Illumina, 2017).

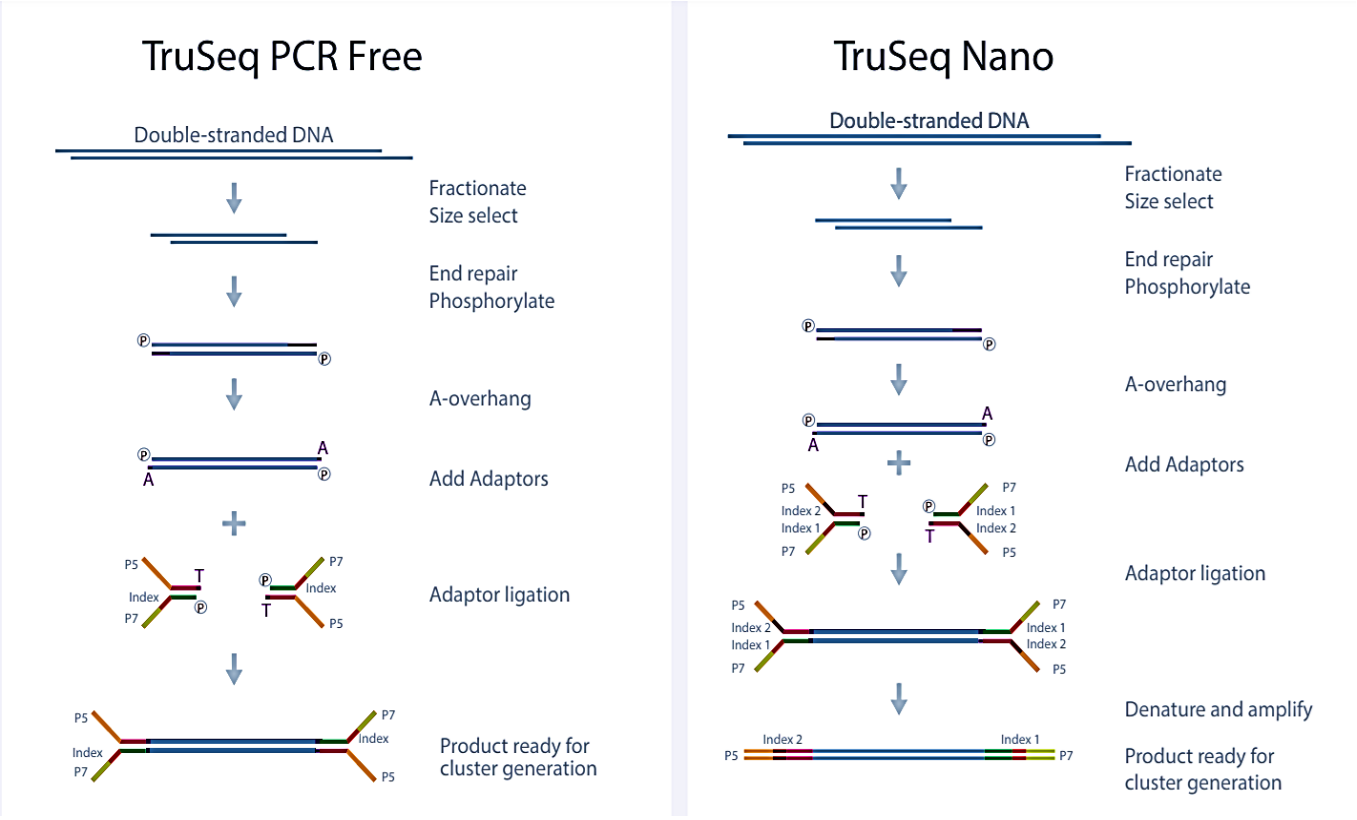


Figure 2.3: DNA Library preparation using the Truseq PCR-free and Truseq Nano Illumina kits
A schematic showing the common processes involved in library preparation, in which DNA is first fragmented, adenylated at the ends and ligated to P5 and P7 adaptors. Additionally, the Truseq Nano library preparation involves the amplification of fragments using PCR. Note the straightening of the Y-shaped adaptor sequences after PCR. Image obtained from https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-dna.pdf.

Quality checks are then performed to determine the quantity of the DNA library as well as to check the fragment size using automated DNA electrophoresis. The protocols used for both Nano DNA and PCR-free library preparation can be found in Appendices E & F. Initially, library preparation was requested to be performed by the sequencing company using the Illumina TruSeq DNA PCR-free Library Preparation kit, followed by sequencing using the Illumina NovaSeq platform. Upon receipt of samples and internal quality control checks performed by the sequencing company, two of the three samples sent for sequencing had insufficient concentration (<30ng/µl) to proceed with the PCR-free library preparation (Refer to QC report in Appendix D) and thus, the Illumina Truseq Nano library preparation kit was performed by Theragen Bio. for all three samples.

2.3.3 Cluster Generation

Once the DNA fragments are ligated to adaptors, these are then applied to a glass flow cell with lanes, which serves as the solid support for the following cluster generation and sequencing by synthesis process. The flow cell is coated with an acrylamide coating, containing immobilised oligonucleotide sequences complementary to the adaptor sequences bound to the DNA fragments. Each fragment is covalently hybridised to the flow cell surface by this interaction and isothermally amplified by DNA polymerase, creating a copy of the strand. The now-double stranded molecule is denatured, and the original DNA template is washed away, with the copy retained on the flow cell. This strand bends and attaches to an adjacent free tethered sequence on the flow cell, forming a ‘bridge’. DNA polymerase generates a complimentary copy of this strand via this bridge, thus generating two copies of the fragment: the forward and the reverse strand. The double-stranded molecule is then denatured, and the process is repeated multiple times to generate clusters. DNA clusters are thousands of clonal copies of the same DNA fragment near the flow cell, to generate a signal that can be detected by the sequencer’s optical system. The reverse strands are then cleaved and washed away to

enable sequencing and base-calling in one direction only. Once these clusters are generated simultaneously for all fragments, further bridge amplification is halted by blocking the 3' end of the strands (**Figure 2.4**).

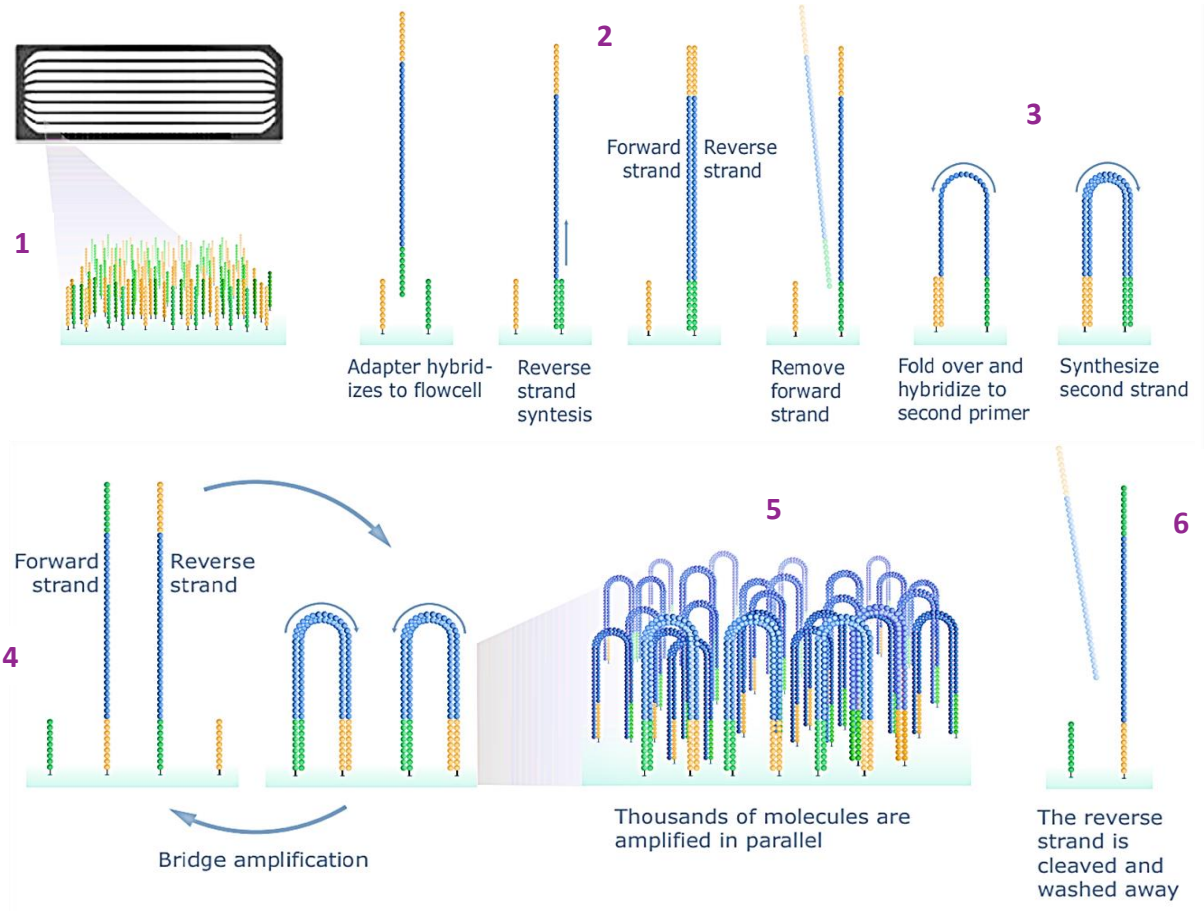


Figure 2.4: Cluster generation by bridge amplification
A diagram showing 1. The flow cell coated with immobilised adaptor molecules, 2. The hybridisation of the DNA fragment strand to the flow cell adaptors and synthesis of new strand by DNA polymerase. The original template is removed, 3. Bridge formation to another adaptor molecule to synthesisfatee complimentary strand, 4. Bridge formation of these two strands to produce clonal copies, 5. Generation of the DNA fragment cluster and 6. Removal of the reverse strands from these clusters in preparation for sequencing. Image obtained from https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-dna.pdf.

2.3.4 Sequencing by Synthesis

To initiate the sequencing process, a sequencing primer attaches to the adaptor region of the forward strands retained from cluster generation. The first read is initiated by the extension of the sequencing primer upon introducing deoxynucleotide triphosphates (dNTPs) of the four

bases: adenine, thymine, guanine and cytosine. These dNTPs are attached to reversible terminators at the 3'OH end which blocks further polymerisation after each base so that only one nucleotide is incorporated into the sequence per cycle. A unique fluorescent label for each base is also bound to the dNTPs. The now fluorescently labelled dNTPs are added and compete to bind to the complementary nucleotide base of the template strand, each base at a time, toward the flow cell surface. After the addition of each dNTP, a light source excites the fluorescent label to emit a signal which is read by an optical sensor. As each of the four dNTPs is assigned a specific dye colour, the wavelength and the signal intensity of the emitted signal determines which base is added and thus, determines the 'base call' of the original forward strand (Hagemann 2015). The 3'OH terminator on the dNTPs is then chemically removed, and the process is repeated to determine the next base. For each cluster containing identical clones, all strands are sequenced and read simultaneously.

Once the first read is completed, the cluster template undergoes paired-end turnaround to initiate read 2, to sequence the reverse strand. After the strand is successfully sequenced base-by-base, the first read product containing dNTPs is washed away. The 3' ends of the template strand are unblocked (from the end of the cluster generation process), which allows the strand to bend, form a bridge and hybridise to an adjacent oligonucleotide on the flow cell. A sequencing primer is added to synthesise the complementary strand along the bridge. The now double-stranded bridge is denatured and linearised before removing the initial forward strand (generated by library preparation) to leave the reverse strand attached to the flow cell surface. The 3' ends of the strand are blocked again to prevent further bridge formation with other nucleotides. A read 2 primer is attached to the synthesised reverse strand template. Sequencing using fluorescently labelled dNTPs is then initiated, as previously described for the forward strand.

In the case that multiple samples are pooled to the flow cell simultaneously, also known as multiplexing, small sequences of DNA called 'indexes' in the adaptor region are attached to correlate the DNA fragment, thus the read, to a particular sample. If indexes are present in the adaptor region, these are sequenced after each read using index sequencing primers for identification (**Figure 2.5**).

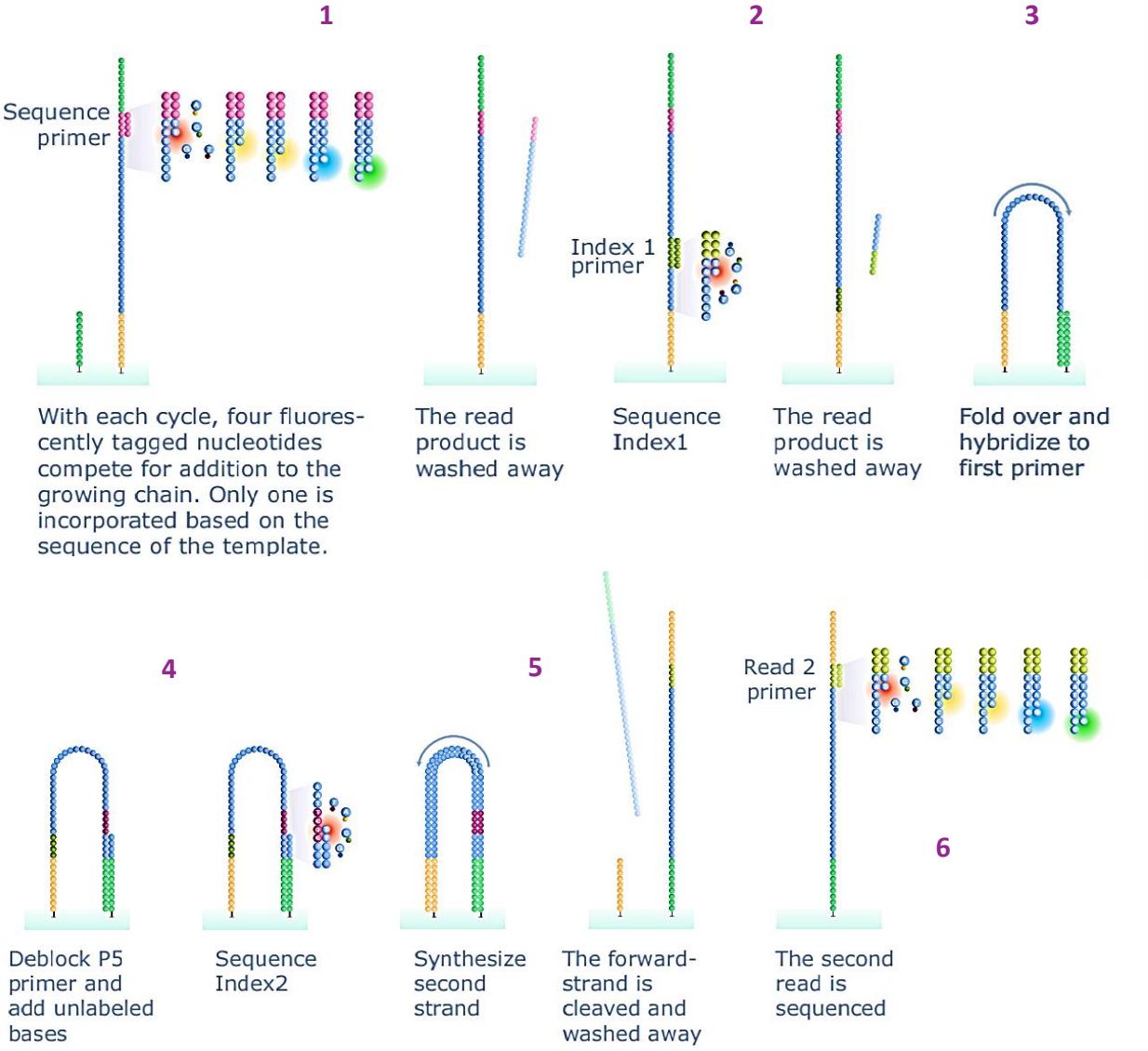


Figure 2.5: Sequencing by Synthesis (Illumina)
A diagram of the sequencing process, depicting 1. Attachment of the sequencing primer and generation of the first read by fluorescently labeled blocked nucleotides (dNTPs). Read product is then washed away, 2. If more than one sample is to be sequenced, attachment of the index 1 primer and sequencing of the first index read, 3. Bridge formation of the forward strand to other nucleotides, 4. Generation of the second index read, 5. Formation of the double stranded bridge to produce the reverse strand, washing away the forward strand and 6. Generation of read 2 by sequencing the reverse strand using the read 2 primer and dNTPs. Image obtained from https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/for-all-you-seq-dna.pdf.

2.3.5 Quality control of data and Bioinformatics

The fluorescent signals of each attached base during sequencing are detected by a coupled-charge device camera which captures an image in Tagged Image File Format (.tiff file) of each cluster showing the fluorescent signal for each added base (Reuter et al., 2015). The instrument control software then extracts the data from these images and the intensity of the signal emitted (in cluster intensity file or.cif format). This data is used by base-calling software to determine which base was added as well as generating a Phred quality (Q) score for each base. The Q-score calculates the probability that the assigned base call is incorrect, thus reflecting the instrument's certainty that the base call is accurate. According to the Phred quality score equation, a Q-score of 30 would indicate that the base call accuracy is 99.9%, alternatively meaning that only 1 in 1000 bases is erroneous. This score is generated after several reads are assessed, and only generated for clusters which pass a quality filter after a number of reads are generated. For the entire run, the percentage of bases having a Q-score higher or equal to 30 is calculated. The base call and the Q-score are saved in base call files (.bcl files). If multiple samples were analysed at once, these reads are separated and identified using index sequences specific for each sample. The .bcl files are then converted to FASTQ files, which contain instrument information, the read sequences as well as quality (Q-score) data. Trimming of adaptor and low-quality sequences was performed with Trimmomatic (Bolger, Lohse, et al. 2014, version 0.4.4) whilst FastQC (Andrews 2010, version 0.11.7) software was used to obtain quality control statistics of the data.

Clusters are read simultaneously in a parallel process, generating an immense number of reads in the FASTQ file for both the forward and reverse strands of all the DNA fragments in the sample. Producing reads for both the forward and the reverse strand, termed as the paired-end approach, is beneficial to determine base calls which were not as successfully called for by the calling software and is also beneficial in accurate mapping of the read to the reference genome

(Ambardar, Gupta, et al. 2016). The forward and reverse strand reads are then paired, in addition to the clustering of several reads which feature the same base calls. This process aids the organisation and generation of contiguous sequences, which are then aligned to a reference genome (**Figure 2.6**) using the Burrows-Wheeler Aligner (Li and Durbin 2009, version 0.7.17) using the Burrows-Wheeler algorithm (Burrows and Wheeler 1994). Paired-end 150-bp reads were aligned to the genome reference consortium human build 37 (GRCh37 or hg19) assembled by the University of California Santa Cruz (Church, Schneider, et al. 2011), generating Binary Alignment Map (BAM) files. Samtools's software set (Li, Handsaker, et al. 2009, version 1.8) was used to sort, filter and index the .BAM files, before running the data through Qualimap (Okonechnikov, Conesa, et al. 2016, version 2.2.1) to review quality control of sequence alignment.

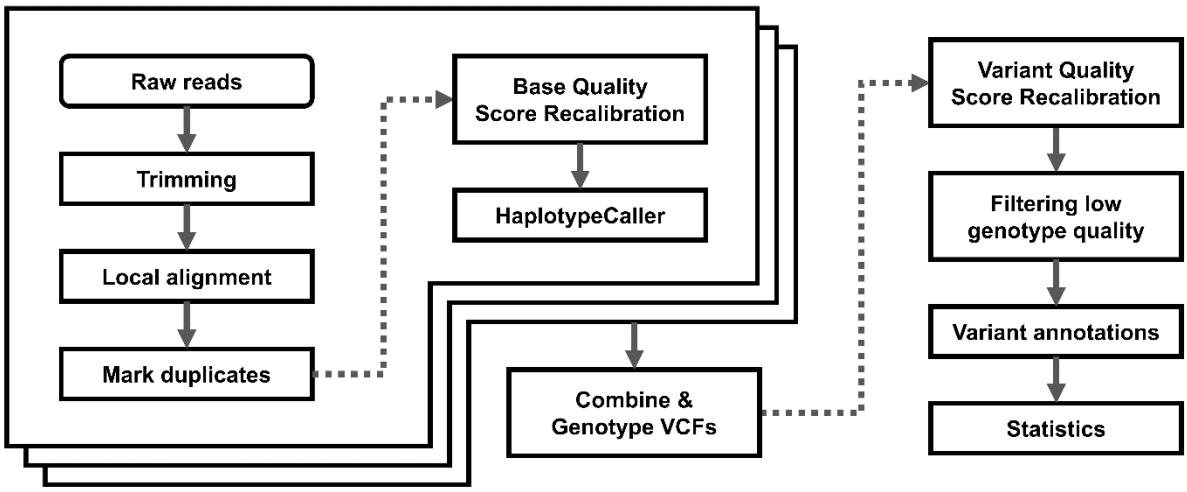


Figure 2.6: Post-sequencing analysis by Theragen Bio
A flowchart showing the sequencing company's sequential in-house bioinformatic pipeline in order to process the obtained sequencing data.

SNVs and insertions/deletions (indels) were identified and assembled using the Gene Analysis Toolkit program (GATK; McKenna, Hanna, et al. 2010, version 4.0.11.0) whilst CNVkit (Talevich, Shain, et al. 2016, version 0.9.8) and LUMPY software (Layer, Chiang, et al. 2014, version 0.3.0) were used to detect copy number variations (CNVs) and structural variants respectively. After variant calling, SnpEff (Cingolani, Platts, et al. 2012, version 4.3) was used

to annotate the detected variants, thus providing the following information for each variant in a .vcf file format to be used for variant filtering (section 2.3.6):

- Genomic coordinates and identification (rs number for known variants): Chromosome number, genomic position, location on forward/reverse strand and name of the affected gene.
- Reference allele, documented alternative alleles and the allele discovered during sequencing.
- Variant consequence (missense/3'UTR/in-frame deletion/etc.) and variant impact (High/modifier/moderate/low).
- Variant frequency in several populations from the gnomAD database (Chen, Francioli, et al. 2022).
- Affected transcript information- Transcript Ensembl ID, Biotype (protein coding, processed transcript, retained intron, etc.), identification of canonical transcript and position in exon/intron.
- Variant position in complementary DNA (cDNA) and coding sequence (CDS), amino acid position in the resultant protein and amino acid substitution, protein domain name.
- Clinvar ID number (Landrum, Lee, et al. 2018) and PubMed ID (Sayers, Beck, et al. 2021)if a variant is referenced in the literature.
- Deleteriousness scores and in-silico predictions from SIFT (Vaser, Adusumalli, et al. 2016), Polyphen-2 (Adzhubei, Schmidt, et al. 2010), CADD (Rentzsch, Schubach, et al. 2021), MutPred2 (Pejaver, Urresti, et al. 2020), and Mutation taster (Steinhaus, Proft, et al. 2021).
- Protein database IDs from Ensembl and Uniprot Knowledgebase (The UniProt Consortium 2023).
- Conservation scores from PhyloP (17-way,30-way and 100-way;Pollard, Hubisz, et al. 2010).
- Phenotypic correlation from several databases: Mouse Genome Database (MGD; (Blake, Baldarelli, et al. 2021),Online Mendelian Inheritance in Man (OMIM; (McKusick-Nathans Institute of Genetic Medicine n.d.) and The Human Phenotype Ontology (HPO; (Köhler, Gargano, et al. 2021).
- Zygosity and coverage of the variant in each research participant sample.

2.3.6 Processing of data and filtering of variants

In order to narrow down the significant number of variants expected to be present in each individual and identify variants which segregate with the disease's phenotype, a filtering pipeline for AD variants was devised (**Figure 2.7**). An AD method of inheritance was chosen as the condition has been described by multiple sources to be inherited in an AD manner (Levy, 2018; Malfait et al., 2017; Syx et al., 2015), a pattern which is also evident from this family's pedigree.

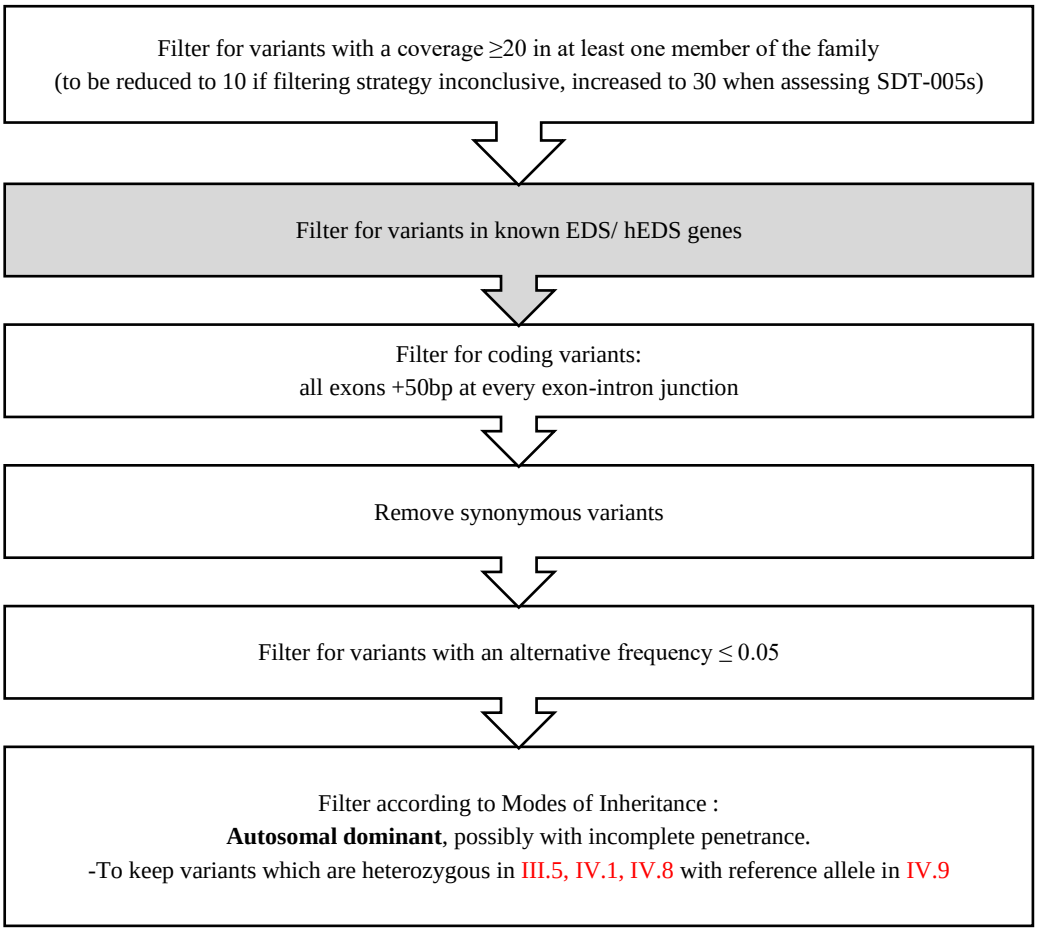


Figure 2.7: Filtering strategy for autosomal dominant variants
A flow chart depicting the process by which variants were filtered according to selected criteria. The second filter (shaded in grey), limiting the location of variants in a pre-defined associated gene list, was omitted when expanding the search to variants in all sequenced genes. bp=base pair.

Using the VCF file containing a list of identified variants for each of the five sequenced individuals as discussed above, a two-pronged approach was implemented. Initially, only

variants in genes related to hEDS/EDS as featured in the published literature as well as those included in gene panels for skin conditions and skeletal dysplasias were selected. The list of the analysed genes can be found in Appendix G. Once this has been completed and a list of segregating variants in these genes was obtained, the process was repeated using the same quality control filters, however expanding the search for variants in all coding regions. In both approaches, a minimum coverage depth of 20 was selected, to exclude variants found in regions of the genome which were not adequately sequenced. If sequencing reads are mapped correctly to the reference genome, having a high depth of coverage ensures that the presence of any identified variant can be called with confidence if several sequencing cycles are yielding the same result (ThermoFisher Scientific n.d.).

Once all variants with sufficient coverage were filtered, only variants that influence protein-coding regions and 5'/3' UTRs were filtered as the majority of disease-causing alterations are considered to reside within/proximal to the protein-coding parts of the genome (Majewski, Schwartzentruber, et al. 2011, Botstein and Risch 2003). Variants residing at least 50 base pairs before or after an intron-exon junction were also retained in order not to skip any potential splice-site variant. In the case that this filtering pipeline yielded no shortlisted variants, WGS data from the remaining participants would have been available to expand the filtering strategy to the whole genome. From the remaining number, variants which do not alter the amino acid sequence, also known as synonymous variants, were also excluded. Since hEDS, despite being regarded as the commonest of the EDS subtypes, is still considered as a rare disease, variants having an AAF of more than 0.05 in the general population were filtered out. Ultimately, to retain variants segregating with the disease phenotype, variants were filtered according to their mode of inheritance. By keeping variants present in the heterozygous state in affected individuals and unaffected participants having the reference allele, one can distinguish the variants that are causing a phenotypic effect, which might in turn be the causative alteration/s.

After the filtering process, two lists of variants were obtained from each filtering pipeline: 1). AD variants in gene panels and 2). AD variants in whole genome/exome. Despite these filters, the AD in the whole genome/exome pipeline still yielded a significant number of variants. To minimise further the number of possible causative variants, functional and pathological annotation of the gene associated with each variant was performed, using Ensembl (Cunningham, Allen, et al. 2022), ClinVar (Landrum, Lee, et al. 2018), Varsome (Kopanos, Tsiolkas, et al. 2019), Genecards (Stelzer, Rosen, et al. 2016), Malacards (Rappaport, Twik, et al. 2017) and OMIM (McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University 2023) databases. A small number of these variants resided in genes whose function has not as yet been elucidated or which are not associated with any medical condition. Any variants in genes implicated in the development or arrangement of connective tissue or the ECM, in addition to variants in genes associated with connective tissue disorders, were selected using annotations from these gene databases as well as the Nosology of genetic skeletal disorders (Unger, Ferreira, et al. 2023). Once the genotype of each recruited individual was established and confirmed by Sanger analysis (refer to section 2.4 below), the location, allele frequency and *in-silico* predicted effect of each variant was elucidated using a variety of online tools and software, as described in **Table 2.5** :

Table 2.5: Variant databases and in-silico prediction tools used to assess filtered variants.

Database/ prediction tool	Function or method of analysis
Integrative Genome viewer (IGV), version 2.16.1. Robinson, Thorvaldsdóttir, et al. (2011)	To manually check the presence and location of the variant in question in sequencing data of participants
University of California Santa Cruz (UCSC) Genome browser Kent, Sugnet, et al. (2002)	To visualise the variant in its genomic position and mining of variant information from various databases in allocated tracks
gnomAD browser, version v3.1.2. Chen, Francioli, et al. (2022)	A vast database of allele frequencies, widely used to obtain alternative allele frequencies (AAF) across populations (Chen, Francioli, et al. 2022). Also provides a measurement of genetic constraint, provided as missense Z-score for the 1 kilobyte (kb) region surrounding the variant of interest. Areas which have fewer variants than expected are assigned a Z-score exceeding 0, with higher values indicating a higher level of genomic constraint.

Database/ prediction tool	Function or method of analysis
Ensembl, release 109, using build GRCh37.p13, Cunningham, Allen, et al. (2022)	A genomic database used to obtain information on variants and the affected transcripts from other data banks
Variant Effect predictor (VEP) McLaren, Gil, et al. (2016)	To assess the functional consequence of variants on the respective transcripts as well as on the coding sequence.
Varsome, version 11.7. Kopanos, Tsiolkas, et al. (2019)	A variant search engine, sequence annotation and impact analysis tool which aggregates data from thirty external databases.
SNPeff, version 4.3t. Cingolani, Platts, et al. (2012)	Used by sequencing company (TheragenBio) in their bioinformatic pipeline to annotate SNVs detected from raw sequencing data and predict their effect on the gene product.
dbNSFP, version 0.9. Liu, Jian, et al. (2011)	Compiles scores of multiple predictive and conservation algorithms, in addition to allele frequencies from various population databases.
Clinvar Landrum, Lee, et al. (2018)	A public archive of genetic variants and their association with human disorders or phenotypes, as publicly submitted by researchers and clinical entities.
Intervar Li and Wang (2017)	For the clinical interpretation of variants according to a set of 28 classification criteria established by the American College of Medical Genetics and Genomics and Association for Molecular Pathology (ACMG/AMP).
Sorting Intolerant From Tolerant (SIFT), version 6.2.1. Vaser, Adusumalli, et al. (2016)	To predict whether the amino acid change resulting from a missense variation is tolerated by the resulting protein based on the biochemical properties of amino acids. Scores exceeding 0.05 imply that the SNV is tolerated, with values less than 0.05 considered as deleterious.
Polymorphism Phenotyping v2 (PolyPhen-2), version 2.2.3, release 405c. Adzhubei, Schmidt, et al. (2010)	Calculates the probability that an amino acid substitution is damaging to the structure and function of the protein using sequence homology, protein model structures and annotation of active or binding sites. In contrast to SIFT, prediction scores exceeding 0.908 are considered as probably damaging, between 0.907 and 0.447 as probably damaging while values less than or equal to 0.446 are assumed to be benign.
MutationTaster Steinhaus, Proft, et al. (2021)	Compiles prediction and conservation scores with allele frequencies and clinical phenotypes from multiple sources in order to predict the effect of variants on sequence conservation, splice-sites and protein structures. This data and the resulting predicted effect can be obtained for each coding transcript.
Mutpred2, version 2.0. Pejaver, Urresti, et al. (2020)	Determines if amino acid substitutions are pathogenic or benign using mean neural network scores, taking into consideration the alteration's effect on various protein properties such as secondary structure and binding site availability. Variants with scores exceeding 0.5 are implied to be pathogenic.
Functional Analysis through Hidden Markov Models with eXtended Features (FATHMM-XF), version 2.3. Rogers, Shihab, et al. (2018)	Predicts the functional consequence of both coding non-synonymous SNVs and non-coding variants using machine-learning models based on single-kernel datasets. Predictions are expressed as a confidence value (p-score) from 0 to 1, with a deleteriousness cut-off of >0.5.
Genomic Evolutionary Rate Profiling (GERP++) Davydov, Goode, et al. (2010)	Identifies evolutionarily constrained loci or elements across multiple species by comparing the degree of substitution to that expected without functional constraint. A substitution deficit, implying that the base or region is retained across species due to evolutionary importance, is denoted by a positive score, whilst a negative score infers that the base or element is highly variable and is not thoroughly phylogenetically conserved.
PhyloP Siepel, Bejerano, et al. (2005), Pollard, Hubisz, et al. (2010)	a command line program within the Phylogenetic Analysis with Space/Time models (PHAST) package. Identifies bases across different species' alignments which are either conserved or are undergoing accelerated evolution under neutral drift. Similarly to GERP++, bases which are predicted to be conserved across species are assigned positive scores, with rapidly changing loci assigned negative scores.

Database/ prediction tool	Function or method of analysis
Combined Annotation Dependent Depletion (**CADD)**, version v1.6. (Rentzsch, Schubach, et al. 2021)	Provides pathogenicity rankings for SNVs and insertions/deletions based on the integration of conservation, functional and splice-site data from genomic sources. Variants having high scores (Ensembl cutoff score of 30) are increasingly likely to be deleterious.
The Rare Exome Variant Ensemble Learner (**REVEL**) Ioannidis, Rothstein, et al. (2016)	Integrates scores from thirteen individual variant prediction tools to assess the pathogenicity of variants. A score higher than 0.5 is considered as likely disease-causing, while scores lower than this threshold are considered as likely benign
MetaLR Dong, Wei, et al. (2015)	Similarly to REVEL, this prediction tool 0 compiles prediction scores from nine independent tools and allele frequency data to predict the deleteriousness of missense variants using logistical regression. Pathogenicity score spans from 0.1, with higher scores are likely to be damaging.
JASPAR (Castro-Mondragon, Riudavets-Puig, et al. 2022)	An open-access database containing manually curated, annotated information on DNA-binding patterns of transcription factors for several eukaryotic species. This data is available as position frequency matrices (PFMs) and TF flexible models (TFFMs).
Haploreg, version 4.2 (Ward and Kellis 2012)	An online tool which predicts the impact of variants in haplotype blocks on features of the non-coding regulatory genome, such as transcription binding motifs and chromatin states. It also provides information on sequence conservation and effects of variants on gene expression. These annotations are based on data from the GENCODE project, which was obtained using chromatin immunoprecipitation with sequencing (ChIP-Seq) technique.

Despite the fact that most prediction tools assess the effect of the variant on the canonical transcript only, information was retrieved for all affected coding transcripts when available. Consequently, the putative effect of variants on the protein structure (when available) was elucidated using a number of protein databases and tools (**Table 2.6**)

Table 2.6: 3D protein structure repositories and prediction tools used to assess altered protein properties

Database/ prediction tool	Function or method of analysis
Alphafold Jumper, Evans, et al. (2021)	An artificial intelligence program which predicts the three-dimensional (3D) structure of proteins according to their amino acid sequence, thus generating a vast database of protein models and confidence scores (pLDDT) per-residue between 0 and 100.
Uniprot Knowledgebase Release 2023_02, The UniProt Consortium (2023)	A protein database providing protein annotation, structural models, protein sequences and involvement in clinical diseases as well as direct links to other sister protein databases to assess involvement in enzyme function, interactions with other proteins and molecular pathways.
Cupsat Parthiban, Gromiha, et al. (2006)	An online tool which predicts the effect of point mutations on protein stability by utilising atom and torsion angle potentials to calculate the difference in free energy of unfolding ($\Delta\Delta G$, expressed as kcal/mol) between reference and altered amino acids.
Missense 3D web server Ittisoponpisan, Islam, et al. (2019)	A web resource which calculates the effect of missense mutations on the biochemical interactions of amino acid residues and the specific structural alterations introduced by amino acid changes.

2.4 Confirmatory testing

Once a list of possible causative variants has been obtained, Sanger sequencing was required to verify the results obtained by HTS as it is regarded as the gold standard for variant confirmation (Refer to Section 2.4.5 below). For this procedure, a sufficient concentration of high-quality DNA covering the gene locus of interest was required. To achieve these requirements, PCR was performed to amplify the areas in which these variants reside using designated primers.

2.4.1 Polymerase chain reaction

PCR is a widespread key technique in molecular biology which replicates in vitro the natural process of DNA replication and elongation during somatic cell division. The technique, first described and applied for DNA elongation by Mullis et al. (Mullis, Faloona, et al. 1986), allows the exponential replication of identical DNA segments by thermally stable DNA polymerase. The two strands of the DNA fragment are first disassociated by subjecting them to a high temperature of 95°C, which denatures the weak hydrogen bonds between the strands. A pair of two short synthetic oligonucleotide sequences complementary to the region of interest called a primer set, bind separately to each strand. Optimum primer annealing is dependent on the melting temperature (Tm) of the short oligonucleotide sequences, calculated using the Wallace-Ikatura formula [4(G+C) + 2(A+T), at which temperature this phase must be carried out (Wallace, Shaffer, et al. 1979). Once the primers anneal to the strands, a DNA polymerase (*Taq* polymerase) simultaneously reads the DNA template and recreates the complementary strand by attaching deoxynucleotides (dNTPs). This elongation step is carried out at a temperature of 72°C, after which four DNA strands are obtained from the original two strands. The process is then repeated on each of these strands for around 25-35 cycles, which results in exponential amplification of the original singular DNA fragment (Waters and Shapter 2014).

2.4.2 Primer design

For each variant, the flanking sequence of the alternative allele was found on Ensembl using the Genome Reference Consortium Human Build 37 (GRCh37, Cunningham et al., 2022; http://grch37.ensembl.org/index.html). This sequence was then inputted into the Primer 3 online tool (Untergasser, Cutcutache, et al. 2012; https://primer3.ut.ee/), which yields ideal primer pairs to amplify the area of interest. Primer 3 suggests primer pairs in consideration of ideal primer conditions for amplification such as primer length, PCR product size, melting temperature (T_m) and %GC content. These selected primer pairs were then checked using the BiSearch tool to detect other sequences in the DNA which might allow non-specific binding (Tusnády, Simon, et al. 2005; http://bisearch.enzim.hu/). Due to the fact that variants might be present in the flanking sequence which might affect predicted primer binding sites, primer sequences were checked using SNPcheck (EMQN and Certus Technology n.d.; https://genetools.org/SNPCheck/snpcheck.htm). While primer sets which do not have any variants in the binding regions were preferred, any variants found in these areas were assessed to determine their AAF. Forward and reverse primer sets were also checked for the possibility of hairpin loop formation using the OligoCALC oligonucleotide property calculator (Kibbe 2007; http://biotools.nubic.northwestern.edu/OligoCalc.html). Primer sets which do not show possible self-complementarity were preferred over self-annealing sequences. In addition, it was ensured that the GC content of the primer sequences was approximately 40-60% and that the difference in melting temperature between the reverse and forward primers did not exceed 5°C. Primer sets for each variant were run through the *in-silico* PCR tool by the University of California Santa Cruz (UCSC; https://genome-euro.ucsc.edu/) to obtain the melting temperature of each primer as well as the size of the amplified product. Once selected, primers were ordered and supplied by Eurogentec at a concentration of 100µM with SePOP desalting

to increase oligonucleotide purification (Kaneka Eurogentec, Seraing, Belgium). The selected primer pairs for the obtained filtered variants can be found in **Table 2.7**:

Table 2.7: Primer sequences used to confirm filtered variants
A table showing the variants of interest after filtering HTS data, the corresponding affected gene and the reverse and forward primer sequences ordered for PCR.

Gene	Variant	Forward Primer (5'- 3')	Reverse Primer (5'-3')	Amplicon Size
PNPLA1	rs45524833	CGCAGCCTTGGTAATTCTCC	CATTTCCAGCTCTCCCAGGTG	236bp
TNXB	rs61746206	CGACCCTGATCATTGCAGTC	CCTCTGCAGTCTCCATGGTG	165bp
TNXB	rs140304758	GATGGTGACCCTGTCCTCATG	GGATGAAGAAATGGCCCCAG	223bp
FANCA	rs201316239	GAAATGTCTTCCCAGCTGTG	CCGCTGCTTGAAGGTTTGTC	150bp
SMAD3	rs189286879	GCAGCACAGAGGCAAACCAC	CACCCTCCCATCCATGTATG	305bp

2.4.3 Optimisation of PCR procedure

To test the ideal PCR conditions, an optimisation run using the selected primers was carried out. As different primers have different ideal annealing temperatures, a gradient PCR approach was employed to determine the optimum annealing temperature to be used. Gradient PCR, in contrast to the conventional single-temperature PCR method described above, amplifies DNA across an inputted range of temperatures using a temperature gradient block setting. The thermal cycler used for this purpose, Biometra PCR Thermal Cycler (analytikjena GmbH, Germany), divides this temperature range across 12 different graduations across the universal block, with each PCR vial corresponding to each graduation.

One sample containing a large volume of good-quality DNA was used as the genetic template for this trial, from which a 50ng solution was prepared. Each forward and reverse primer was reconstituted using 20µl of 50µM oligonucleotide solution and 20µl of distilled, nuclease-free water. The One-Taq 2x Master Mix with standard buffer (New England Biolabs, Ipswich, UK) was used, containing dNTPs, Magnesium chloride, pH buffers and stabilisers, *Taq* DNA

polymerase enhanced with Deep Vent® DNA Polymerase as well as Xylene Cyanol FF and Tartrazine tracking dyes. For each primer set, twelve PCR reaction vials were prepared, each containing a 10µl solution containing 5µl of OneTaq 2xMaster Mix solution, 0.1 µl each of the prepared forward and reverse primer solutions, 1 µl of DNA solution and 3.8 µl of distilled water (**Table 2.8**).

Table 2.8: Volumes of reagents used for PCR optimisation in each vial

Reagent	x1 Vial (µl)
OneTaq Master Mix 2X (NEB, UK)	5.0
Forward Primer (50uM)	0.1
Reverse primer (50uM)	0.1
DNA (50ng)	1.0
Distilled H_2O	3.8

An additional vial containing the solution described above without the addition of DNA was also prepared for each primer set and used as a no-template control (NTC). Reaction vials were tightly capped with domed sealing caps to prevent any possibility of evaporation of reagents. Gradient PCR was then performed using the Biometra PCR Thermal Cycler (Biometra GmbH, Göttingen, Germany) using a temperature gradient ranging from 52℃ to 62℃ distributed across 12 reaction vials by the thermal cycler.

2.4.4 Agarose gel electrophoresis

Following amplification, PCR products were visualised by gel electrophoresis to determine if amplification was successful across the specified temperature range. Upon application of an electrical current, DNA fragments amplified by PCR traverse a porous medium such as an agarose or polyacrylamide gel, at a distance inversely proportional to its molecular weight and size.

A 1% agarose gel was prepared by dissolving 1g of Agarose powder (Sigma-Aldrich, Missouri, USA) and 100ml of 1x Tris-acetate-Ethylenediaminetetraacetic acid (TAE) buffer (consisting of 40mM Tris, 20mM acetate and 2mM EDTA, adjusted to pH8.1), after which the solution was heated in a microwave until all the agarose powder dissolved. Precisely 10µl of Ethidium bromide (10ug/l concentration; Sigma-Aldrich, USA) was added to the solution for ultraviolet visualisation of the PCR products after electrophoresis. The solution was then poured into a medium casting tray (Biometra, Germany) and combs were put in place before letting the gel set. The gel was then put in an electrophoresis tank filled with a solution of 1xTAE buffer and Ethidium bromide. After removal of the combs, 5µl of the PCR products were pipetted into the wells, also including a 100 bp ladder (Solis Biodyne, Tartu, Estonia) as a size marker at the end of the lane. The tank was covered and a constant voltage of 120 volts with variable current was applied for approximately 30 minutes. After electrophoresis, PCR products were then visualised using UVP Biospectrum Imaging System (Fisher Scientific, USA) to assess if the amplification reaction was successful using the designated primers as well as the level of amplification at each temperature graduation (**Figure 2.8**).

All five primer sets successfully amplified the sample DNA, visualised by the presence of thick bands each corresponding to a specific temperature. It was noted however that some primers sets were less successful at the higher end of the temperature range Despite mostly satisfactory bands across the temperature range, the fourth band of each primer set (corresponding to a temperature of 54°C) was identified as the common optimum annealing temperature of all five primer sets.

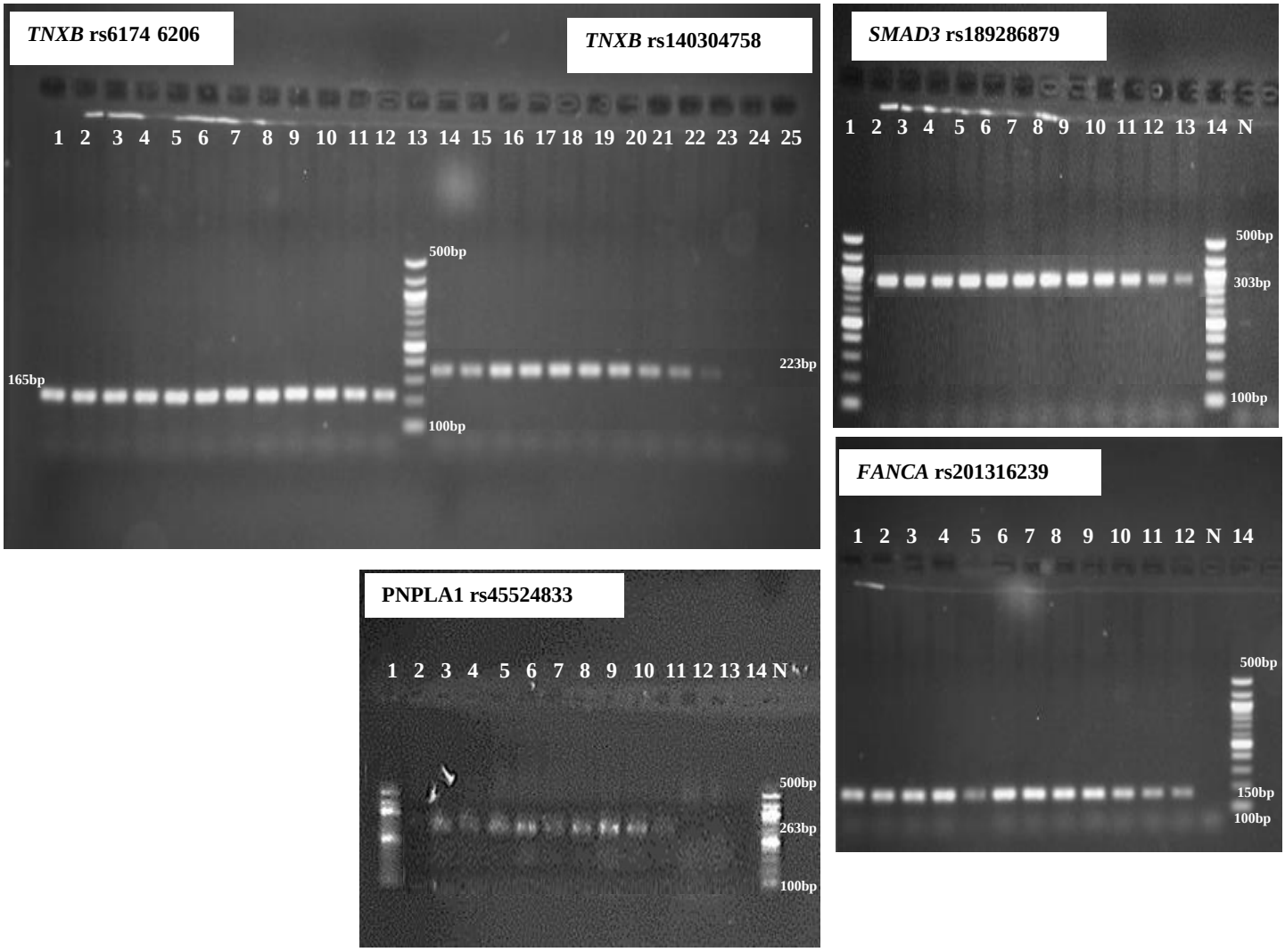


Figure 2.8: Results of the primer optimisation process
Electrophoretogram images of the PCR products obtained by Gradient PCR for each variant. Note the decreasing visibility of bands, signifying reduced concentration of PCR products upon increasing temperature in *TNXB* rs140304758 (well 23) and *PNPLA1* rs45524833 (well 11). Satisfactory bands were obtained across all temperatures for *TNXB* rs6174 6206, *SMAD3* rs189286879 and *FANCA* rs201316239. N denoted the presence of a no-template control. Sizes of ladder fragments (100-500bp) and PCR products indicated accordingly. Outputs obtained using UVP Biospectrum Imaging System (Fisher Scientific, USA).

2.4.5 PCR Amplification

After an optimum annealing temperature of 54℃ was identified, patient samples were separately amplified using each primer set, thus obtaining an amplified DNA segment for the five filtered variants, for each individual. The thermal profile used by the thermal cycler for the PCR reaction is depicted in (**Figure 2.9**).

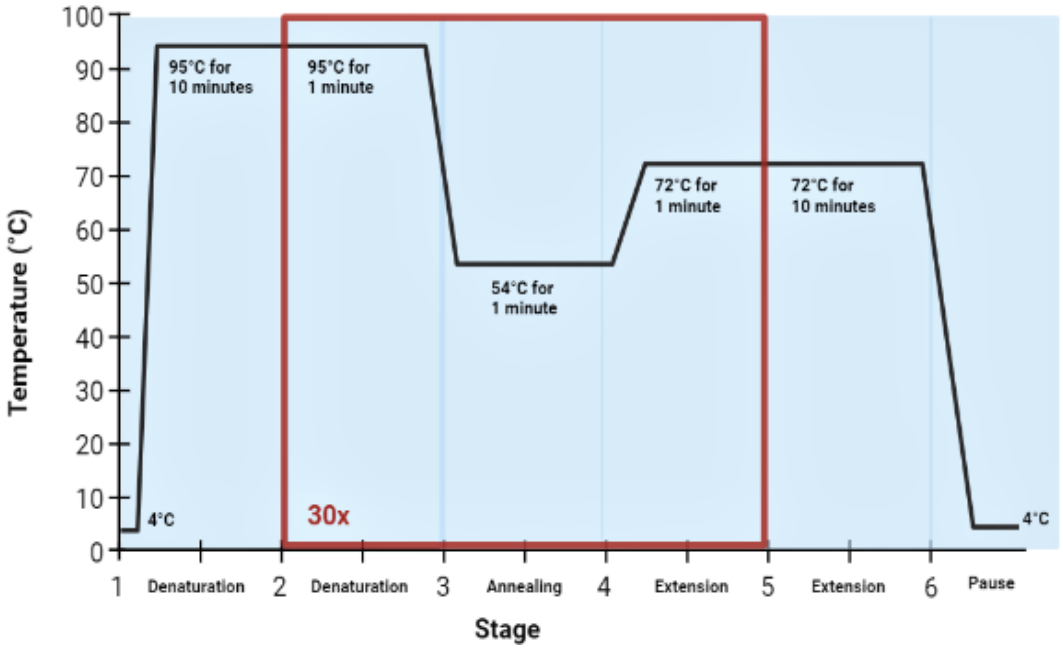


Figure 2.9: Thermal profile of the PCR amplification run
A graph showing the stages of the PCR reaction, including the temperatures used and the duration of each process. Area highlighted with a red box denotes the primer amplification cycle, which was repeated thirty times to yield an adequate concentration of amplified DNA. Image created with BioRender online tool (https://www.biorender.com/).

DNA samples of two recruited family members who were not selected for HTS were also included for amplification and sequencing, to test if the filtered variants are present in these participants as well. The DNA concentration of these samples was re-assessed as previously described (section 2.2.2). Samples with low DNA concentration were supplemented by additional DNA samples previously collected for these individuals during their clinical appointments, which were stored at the cytogenetics laboratory at Mater Dei Hospital. A 50ng DNA solution was first prepared for each sample according to DNA concentration readings, using the formula C1V1=C2V. As the PCR products were prepared for Sanger sequencing, a 30µl solution was prepared instead of 10µl used for optimisation, from which 2µl were used for visualisation of the PCR products by gel electrophoresis, as described above. All samples were adequately amplified for the variants in *TNXB*-617,*TNXB*-140, *SMAD 3*, *FANCA* and *PNPLA1* using the optimised protocol (**Figure 2.10**).

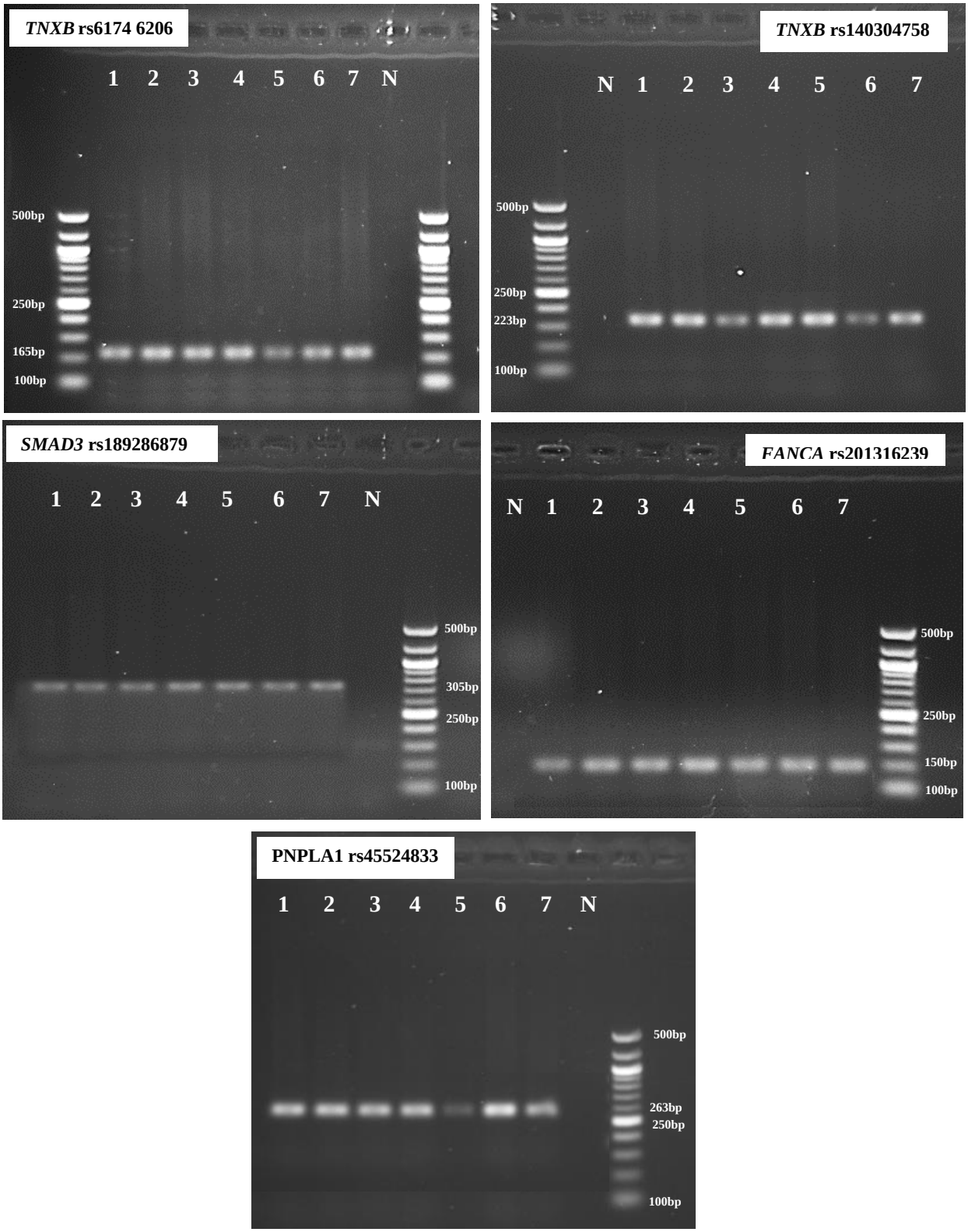


Figure 2.10: Results of the PCR run prior to Sanger sequencing
Gel images illuminated by UV light showing the amplified PCR products for each variant, with each band in wells 1-7 corresponding to each individual. Blank wells adjacent to the 100-500bp ladders contained no template controls (NTCs), which did not contain sample DNA and thus were not amplified. Bands can be seen in all patient samples for variants in *TNXB*-617,*TNXB*-140,*SMAD3*, *FANCA* and *PNPLA1*. Sizes of ladder fragments (100-500bp) and PCR products indicated accordingly. Outputs obtained using UVP Biospectrum Imaging System (Fisher Scientific, USA).

2.4.6 Purification of PCR products

As the PCR products obtained contained residual reagents from the PCR procedure such as unused dNTPs and primers, purification of the amplified DNA from these impurities is required before performing Sanger Sequencing. The PCR products were thus purified using a column-based method (AccuPrep® PCR/Gel Purification Kit, Bioneer, Daejeon, South Korea), which proposes a maximum of 95% recovery of fragment DNA (Bioneer n.d.).

Approximately 5 volumes (137.5µl) of Fragment DNA Binding Buffer were added to each PCR product solution in a micro-centrifuge tube and vortexed. This mixture was then transferred into a DNA binding column in a collection tube, as supplied by the manufacturer, prior to centrifugation for one minute at 28341*g*. The DNA fragments bind to the filter in the column while the residual PCR reagents were eluted in the collection tube and discarded. Buffer 2 was reconstituted by adding 100ml of absolute ethanol, after which 500µl of this solution was added to the bound DNA fragments column. The column with the collection tube was then centrifuged for another minute at 28341*g*, before discarding the eluted solution containing salts and soluble impurities in the collection tube. This process was repeated by re-adding another volume of 500µl and centrifuging the tube again. Once the flow-through was discarded, additional centrifugation was performed using the same conditions to 'dry' the bound DNA from residual ethanol in the buffer. The filter column containing the bound DNA fragments was then assembled onto a 1.5ml snap-lock micro-centrifuge tube, to which 30µl of Elution Buffer were added (**Figure 2.11**).

During pipetting of buffers onto the filter column, care was taken not to probe the filter with the pipette tip, so as not to displace and remove any DNA fragments bound to the filter. The buffer in the filter tube assembly was then left at room temperature for one minute to be absorbed by the filter and centrifuged for another minute at 28341*g* for elution.

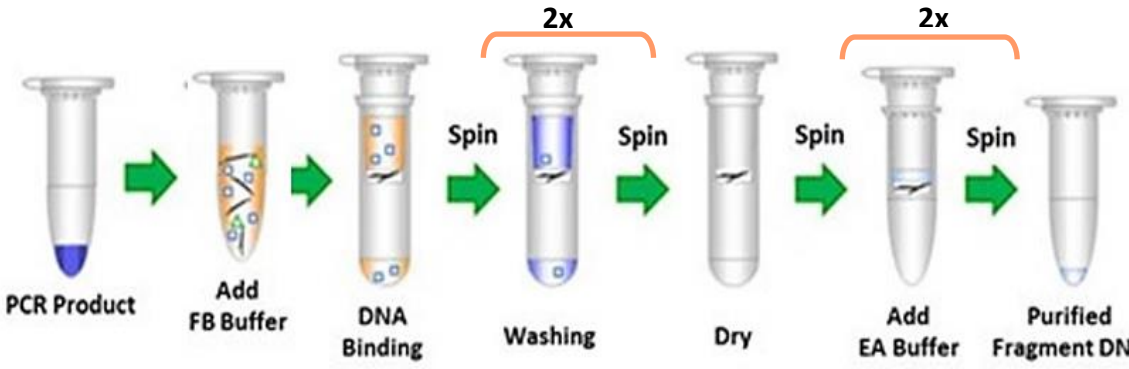


Figure 2.11: PCR product purification AccuPrep® PCR/Gel Purification Kit
A schematic showing the steps required to purify the obtained PCR products. Steps delineated by an orange bracket were repeated twice to obtain a high-purity high-yield DNA solution. Image obtained and modified from https://us.bioneer.com/pagecat1/accuprep/pcr-gel-purification-kit?tab=technical.

This elution process was performed twice to maximise the DNA yield from the filter. A solution of amplified DNA fragments free from PCR contaminants was thus obtained in the micro-centrifuge tube. The concentration and quality of purified DNA solutions were then checked using Nanodrop, as described in section 2.2.2. All samples had a concentration equal to or higher than 5ng/ml (Refer to Appendix H) and were thus adequate for Sanger analysis.

2.4.7 Sanger sequencing

Sanger's chain-termination method is a technique developed in 1977 by Sanger et al. . It is regarded as the first-generation sequencing process from which most modern sequencing techniques developed. Despite this advent of automated massively parallel high-throughput technologies, Sanger sequencing remains the benchmark, gold standard technique for confirmation of variants identified by NGS due to its reliability and lower error rate, especially in the deduction of false positive results obtained by HTS in GC-rich areas (De Cario, Kura, et al. 2020). For this application, the DNA fragment containing the variant of interest is added to a PCR mixture containing DNA polymerase, a single primer, dNTPs and a lower concentration of di-deoxynucleotides (ddNTPs) of the four bases each containing a unique fluorescent label (**Figure 2.12**). The ddNTPs lack the OH- hydroxyl group on the 3' end of sugar residue, blocking the binding of further nucleotides thus 'terminating' the amplification chain.

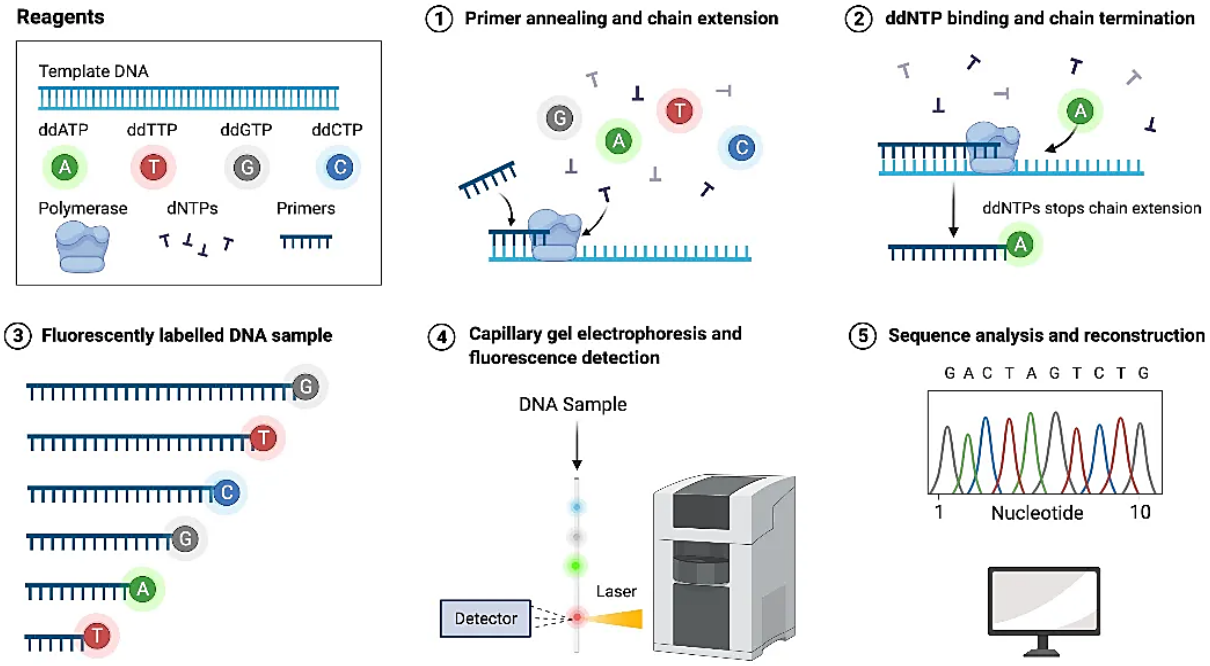


Figure 2.12: Sanger Sequencing
Figure showing the reagents required and processes during analysis (1-5), resulting in the chromatogram (5) and obtained DNA sequence. Image obtained from https://microbeonline.com/dna-sequencing-sanger-sequencing-method/.

During the PCR reaction, the DNA strand is replicated by the binding of routine dNTPs, with the random incorporation of the corresponding ddNTP which stops further extension (Hagemann 2015). This results in many copies terminated at random lengths with different molecular weights. These copies are then separated by capillary gel electrophoresis, which separates the fragments according to the speed by which they traverse the gel matrix. The shortest fragments, having a lower molecular weight, are the first to migrate through the gel.

At the end of the process, a laser excites the fluorescent label on the terminal ddNTP, which emits a different colour for the incorporated base (A, T, G or C). Thus, the sequence of the DNA fragment is read and called by the analyser in the 5' to 3' direction by following the colour of the emitted light (ThermoFisher Scientific n.d.). The output, or chromatogram, depicts the fluorescent peaks corresponding to each nucleotide in the template DNA and is then compared to the reference genome to deduct the variant of interest.

After purification, the PCR products were stored at -20℃ until they were packaged and shipped to LGC Genomics (Berlin, Germany) for Sanger sequencing. The reverse primer for every variant tested was also sent with the samples to be used during the technique. The reverse primer was chosen rather than the forward primer as the variants investigated reside closer to the 5' end of the PCR product and thus might be missed if the forward primer was selected. The obtained chromatograms (refer to section 3.3) were analysed using Codon Code Aligner software (CodonCode Corporation 2019, version 9.0) to ensure that the results obtained by Sanger sequencing confirm those obtained by HTS.

Chapter 3: Results

3.1 Recruitment of research participants

3.1.1 Clinical examination

Six individuals, five females and one male (age range from 21-57 years) from the same ethnically Maltese family were recruited for the study. Apart from IV.9, all exhibited clinical features of hEDS. III.2 describes her late maternal grandfather (I.1) and mother (II.2), as having been hypermobile. III.4, IV.5 and IV.7 are said to be hypermobile but were not recruited for the study as they reside abroad. The proband (IV.2) was initially investigated for Hereditary Haemorrhagic Telangiectasia (HHT) due to a history of recurrent epistaxes, multiple telangiectasias and an arteriovenous (AV) malformation. A potentially pathogenic variant was identified in the Endoglin gene (*ENG*: c.1280T>G, p.Val427Gly) and a diagnosis of HHT was thus confirmed. Other family members at risk were screened for the *ENG* variant by targeted genetic analysis. The variant was subsequently identified in her mother (III.2), maternal aunt (III.5) and maternal cousin (IV.8). A physical examination carried out during her first visit to the genetics clinic, the proband was noted to be hypermobile and also had joint instability and gait disturbances. These features were later found to be presenting other family members indicating a likely hereditary underlying genetic aetiology.

The probands' relatives (refer to pedigree) and an affected unrelated female were subsequently invited to participate in this study. The International EDS Society Diagnostic Criteria checklist (Appendix A) was used to determine if the recruited participants fulfilled the clinical criteria for hEDS (**Table 3.1**).

Table 3.1: Diagnostic outcomes of recruited individuals using the 2017 Diagnostic Criteria for hEDS Diagnosis is based on the presence of two mandatory criteria: presence of hypermobility (criterion 1) and exclusion of alternative diagnoses (criterion 3) in conjunction to having a two-thirds majority of other symptoms (criterion 2). Five out of seven recruited individuals were thus diagnosed with hEDS.

Diagnostic Criteria for Hypermobile Ehlers-Danlos Syndrome (hEDS)	III.2	IV.1	IV.2	III.5	IV.8	IV.9	SDT-005
Criterion 1- Generalised Joint Hypermobility (Mandatory)							
-Positive Beighton Score (≥5/9 up to 50 years of age, ≥4/9 if over 50)	✓	✓	✓	✓	✓	×	✓
Criterion 2- Symptomatic presentation (2/3 features must be present)							
-Feature A: Systemic complications (5/12 features must be present)	× (3/12)	× (4/12)	× (3/12)	× (4/12)	✓ (6/12)	× (3/12)	✓ (6/12)
-Feature B: Positive family history	✓	✓	✓	✓	✓	✓	×
-Feature C: Pain/ joint dislocations	×	✓	✓	✓	×	×	✓
Criterion 3- Exclusion of alternative diagnoses (Mandatory)							
-Absence of unusual skin fragility	✓	✓	✓	✓	✓	✓	✓
-Exclusion of other heritable/ acquired connective tissue diseases	✓	✓	✓	✓	✓	✓	✓
-Exclusion of neuromuscular disorders and skeletal dysplasias	✓	✓	✓	✓	✓	✓	✓
Fulfils diagnostic criteria for hEDS?	**No**	**Yes**	**Yes**	**Yes**	**Yes**	**No**	**Yes**

Four out of six family members (III.5,IV.1,IV.2,IV.8), as well as SDT-005, fulfilled the diagnostic criteria for hEDS. It was noted that one of the family members who did not fulfil the criteria (III.2) failed to be classified as having hEDS only due to the absence of chronic pain or joint dislocations. Nevertheless, she demonstrated joint laxity even after surpassing 50 years of age and has two symptomatic, affected offspring (IV.1 and IV.2). Her husband (III.1) refused participation in this study but he is not known to be hypermobile or to exhibit any of the associated clinical features of hEDS. Therefore, III.5 can be classified as having Generalised Hypermobility Spectrum Disorder (G-HSD, refer to Figure 1.4). SDT-005 gives no family history of HCTDs but fulfils the criteria for hEDS based on her clinical phenotype and medical history.

3.1.2 Medical History

A thorough medical history covering the full extent of hEDS symptomatology as well as lifestyle factors, nutrition and prescription use was obtained for each research participant by the use of a medical questionnaire. Tables showing participants' answers to the questionnaire can be found in Appendix J .

Aside from lifestyle factors, the medical questionnaire was also utilised to get a thorough understanding of the symptoms and functional complaints of the recruited participants, augmenting the symptomatology previously obtained during the diagnostic assessment. There seems to be no difference between affected and non-affected individuals when assessing psychological issues and allergies. All participants report a history of asthma with the exception of IV.1 and IV.9. This same trend was also replicated when enquiring about orthostatic (postural) hypotension. All female participants exhibited most of the associated cutaneous features of hEDS whereas the male participant (IV.1) had no such features. The most common gastro-intestinal complaints were abdominal discomfort and nausea, both of which were absent in IV.1 and IV.9. This pattern was replicated again upon enquiring about fatigue and 'brain fog'- all respondents reported chronic fatigue apart from these two participants in addition to IV.8. All those who responded 'yes' noted that this fatigue starts early in the morning. The majority of affected participants (III.2,III.5,IV.8 and SDT-005) have daily chronic pain. When asked about the location of this pain (which is widespread in some individuals), various areas were mentioned, however the most severe areas were the knees, the shoulders and the neck. One of most characteristic features of hEDS- history of joint dislocations - was only present in 57% of recruited participants, which also included IV.9. In turn, III.2,III.5 and IV.8 had no history of joint dislocations despite having a high pain index. There were no reported functional limitations in IV.9, while SDT-005 responded 'yes' to every criterion. The most common

limitations in hEDS family members were difficulties in balance and coordination, whilst the least common were climbing stairs and lifting/carrying of objects.

3.1.3 Biochemical measurements

Testing of various analytes in peripheral blood were performed at the Pathology department of Mater Dei Hospital; the parathyroid hormone (PTH) levels were performed at an external private laboratory. Phosphate, Magnesium, Albumin, ALP and PTH levels were within the normal ranges issued by the laboratories performing these tests (**Table 3.2**). A minute elevation of ionised calcium was noted in three participants (III.2, IV.2 and III.5) whilst a significant elevation was detected in SDT-005. The majority (%) of research participants had low 25-OH Vitamin D levels. Therefore, no observable trend between hEDS status and these biochemical measurements was observed.

Table 3.2: Results of analyte testing from peripheral blood
A table showing each individual's test results for analytes associated with bone/musculoskeletal metabolism. Values shown in red indicate a deficiency or a value below reference range, while those shown in light blue exceed the normal range.

Analyte tested (normal range)	**III.2**	**IV.1**	**IV.2**	**III.5**	**IV.8**	**IV.9**	**SDT-005**
25-OH Vitamin D (30-100ng/ml)	34	<13	15	17	15	14	31
Ionised calcium (1.12-1.32 mml/l)	1.36	1.37	1.32	1.40	1.32	1.29	2.32
Phosphate (0.87-1.45mmol/l)	1.07	1.15	1.38	1.05	1.38	1.06	1.30
Magnesium (0.65-1.05 mmol/l)	0.90	0.95	0.89	0.83	0.89	0.81	1.02
Albumin (32-52 g/l)	48	52	49	49	51	47	50
Alkaline Phosphatase (40-104 U/l)	95	72	72	58	68	77	74
Parathyroid hormone (15-65 pg/ml)	56	63	25	46	40	51	60

3.1.4 Bone Mineral Density Assessment

BMD data was obtained for all recruited participants except for IV.9, who refused the test. III.2 had a low T-score value at the LS (-1.6), FN (-1.9) and TH (-1.0) while SDT-005 also had low T-score values at these sites (-2.0, -1.8 and -1.9 respectively. These participants, who are both post-menopausal, were classified as osteopenic according to the guidelines issued by WHO (Kanis, Burlet, et al. 2008). The remainder of tested participants have a normal BMD value at both measured sites (**Table 3.3**).

Table 3.3: Results of Bone Mineral Density measurements

A table outlying the demographic characteristics of the recruited participants, their Z- and T-scores for lumbar spine, femoral neck and total hip, the calculated BMD at these sites in grams per centimetre squared (g/cm^2) as well as risk factors affecting bone mass. LS= Lumbar spine, TH= total hip, BMI= Body Mass Index, N/A= Not available. Values outlined in red, belonging to III.2 and SDT-005, indicate low bone mass.

Participant	Gender	Age	LS T-score	LS Z-score	Total LS BMD (g/cm2)	FN T-score	FN Z-score	FN BMD (g/cm2)	TH T-score	TH Z-score	TH BMD (g/cm2)	Risk factors
III.2	F	57	-1.6	-0.5	0.873	-1.9	-0.8	0.635	-1.0	-0.3	0.815	Smoker
IV.1	M	36	0.8	0.8	1.203	0.9	1.2	1.051	1.6	1.8	1.279	Vitamin D deficiency, High BMI, Smoker, Previous fracture
IV.2	F	30	2.6	2.6	1.362	1.2	1.3	0.981	0.9	0.9	1.048	Vitamin D deficiency, High BMI
III.5	F	44	1.7	2.0	1.261	1.4	1.7	0.999	0.9	1.2	1.056	Vitamin D deficiency, Smoker
IV.8	F	21	N/A	N/A	1.237	N/A	1.6	1.032	N/A	0.7	1.031	Vitamin D deficiency, Smoker
SDT-005	F	48	-2.0	-1.4	0.855	-1.8	-1.2	0.652	-1.9	-1.5	0.710	Previous fracture, glucocorticoids

3.2 High-Throughput Sequencing

3.2.1 Quality control of HTS data

In order to avoid the use of low-quality data in the bioinformatic pipeline, several quality control measures were performed after sequencing (Refer to Appendix I), initialised by the trimming of adaptors and removal of short and low-quality sequences. Approximately 25,000 megabytes (MB) of data were trimmed from each sample sequenced by Theragen Bio, resulting in approximately 61,000 to 71,000MB of clean reads per sample in a FASTQ file format. These samples had high Phred scores across all bases in the sequence (**Figure 3.1**) and Q30 scores (**Table 3.4**), defined as the probability of one incorrect base call in a thousand calls. This indicates that 90-91.5% of base calls in these three samples were called at 99.9% accuracy, thus ensuring a low rate of false-positive variant detection resulting from incorrect base calling (Illumina 2011).

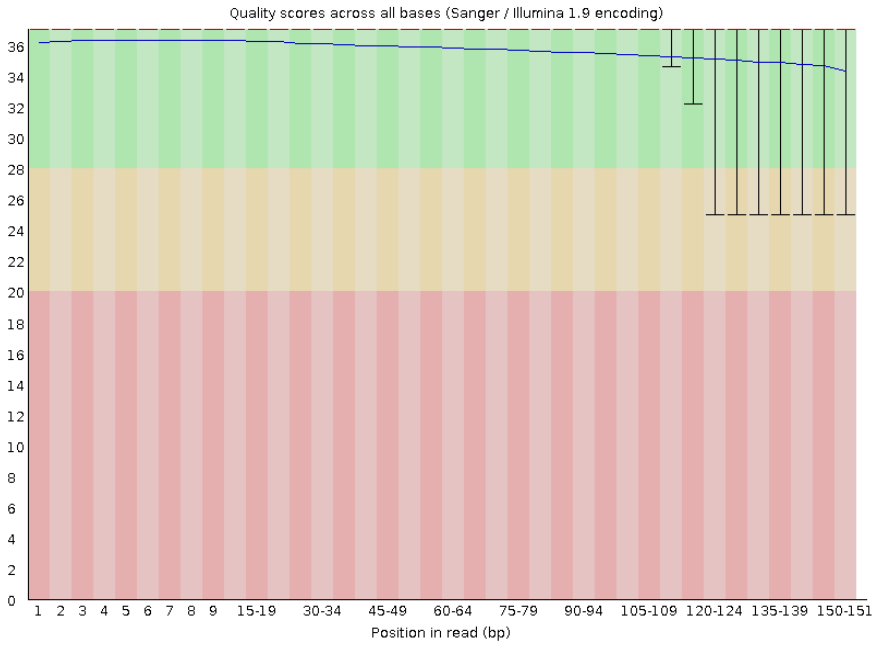


Figure 3.1: Range of quality values across bases
A box plot displaying the Phred score of all sequenced bases in IV.1 relative to the position of the base in the read sequence. The 25-75% interquartile ranges (located above plotted region) of bases in all positions had Phred scores of more than 36, shown residing in the upper green range. This indicates high quality calls throughout the read sequence. Lowest Phred scores were observed at the ends of sequences(right side), lowest being a score of 25, denoting 99.68% accuracy of base calls. HTS data of other sequenced participants had similar quality parameters. Box plot obtained by FastQC software (Andrews, 2010).

All samples had a high rate of alignment to the reference genome (94.94 - 99.86%) and good mean coverage of 20 reads per base for 90.45 - 99.15% of reads . An overview of these parameters can be found in **Table 3.4**. The rate for 'N' calls, base calls which could not be discerned by the sequencing software as A, T, G or C, was 0.0%.

Table 3.4: HTS analysis statistics
A table showing sequencing statistics for samples for which HTS data was obtained. Boxes shaded in dark grey indicate values which were not provided by the sequencing company.

Patient sample	Number of Raw reads	Q30 rate (%)	Alignment mapping rate	Mean sequencing depth	GC content (%)	Coverage 20x (%)
III.5	813951233	95.34	94.94	36.696	41.00	98.25
IV.1	663530846	91.00	97.16	28.121	40.74	90.45
IV.8	65604259	96.27	99.86	27.560	50.39	99.15
IV.9	741040738	90.94	97.63	31.413	41.07	96.73
SDT-005	640250862	90.05	97.52	27.081	40.84	92.37

Percentage of GC-content ranged from 40.74 - 50.39%, with the majority of mapped reads consisting of 30 - 45% GC areas (**Figure 3.2**). Depth of coverage was observed to be satisfactory compared to the reference genome, indicating that there was no large deletions or under/over-representation of certain loci in the acquisition and alignment of data (**Figure 3.3**).

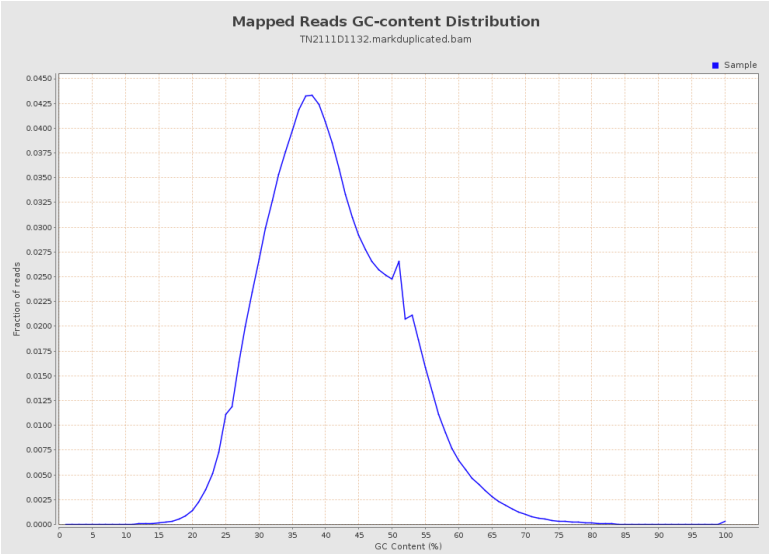


Figure 3.2: Distribution of GC content
A graph showing the distribution of GC content percentage per mapped read. The highest fraction of reads consisted of 30-45% GC bases, denoting normal distribution. Graph obtained using Qualimap 2.2.1 (Okonechnikov et al., 2016).

Figure 3.3: Coverage across reference genome and reference GC content
A graph showing the depth of coverage in a sample (upper graph, red line) across different positions in the genome, superimposed on the depth of coverage of the reference genome (upper graph, light blue areas). Sample coverage is shown to largely mirror reference coverage. Lower graph shows GC content across reference (black line) together with its average value (red dotted line). Figures obtained from Qualimap version 2.2.1 (Okonechnikov et al., 2016).

3.2.2 Processing of data and filtering of variants

Filtering of variants was carried out by merging the .vcf files of sequenced individuals and applying a number of sequential filtering step as discussed in section 2.3.6. Initially, this process was applied by limiting the filtering strategy to variants in three gene panels: variants associated with EDS/hEDS, skin disorders (Centoskin) and skeletal dysplasias only (Refer to Appendix G). The Centoskin and skeletal dysplasias gene panels have been previously compiled by Centogene (Rostock, Germany) to review sequencing data from hypermobile individuals (Prof. I. Borg, pers. comm. 16th March 2022). The EDS/hEDS gene panel was compiled for the purpose of this study by including genes associated with all EDS subtypes (Malfait, Francomano, et al. 2017) as well as *LZTS1* which has also been linked to hEDS (Syx, Symoens, et al. 2015). The filtering process was then expanded by including variants found in all exonic regions and 5'/3' UTRs. The devised filtering pipelines retained only heterozygous

and homozygous variants in III.5, IV.1 and IV.8 which are also absent in IV.9. Additionally, since the primary aim of this study was to explore inherited variations causing the hEDS phenotype in the recruited family, sequencing data of SDT-005 was not included in the filtering pipeline. Sequencing data of this individual was retained in the eventuality that the filtering pipeline yielded a large number of variants from the family, which would require further exclusion. The last filtering step, in which variants were filtered according to the mode of inheritance, was carried out for autosomal dominant (AD) variants only as the family pedigree shows a clear AD mode of inheritance. Steps and counts are depicted in **Table 3.5**.

Table 3.5: Number of variants remaining after each filtering step for every gene panel/ all coding regions

Filtering step	**No. of variants EDS panel**	**No. of variants Centoskin panel**	**No. of variants in Skeletal Dysplasias panel**	**No. of variants in all coding regions**
Total no. of variants	23,853,489	23,853,489	23,853,489	23,853,489
Number remaining after removal of variants with coverage <20	21,371,520	21,371,520	21,371,520	21,371,520
Retained variants in gene panel only	117,83	41,645	196,162	Not applied
Variants in extended coding regions only	491	1,684	4,421	466,113
Removal of synonymous variants	401	1,152	3,590	409,192
Removal of variants with AAF of >0.05%	50	141	485	65,193
Filtered AD variants and according to pedigree	6	4	2	369
Remaining variants	**6**	**4**	**2**	**369**

For the last filtering step, only AD variants which are heterozygous alternative in III.5, IV.1 IV.8 and homozygous reference in IV.9 were retained. As the number of resulting variants from the whole genome analysis was substantial (369, Appendix K), further annotation of variants was required in order to assess the relevance of the affected genes and their associated

disorders in relation to the pathophysiology of hEDS (annotated list of genes in which these variants were found can be found in Appendix L. From this list of variants, only the variants which were also detected using the gene panel approach were deemed to be relevant to the condition, although a small number of variants were either not annotated in databases, were not associated with any condition or the information available is too minimal to be able to assess their effect. Nevertheless, the fact that the variants detected using the gene panel approach were also discovered using the whole genome approach proves in part the viability of the filtering pipeline.

It was noted that after filtering the data for variants in genes related to EDS/hEDS, the variants retained, all found in the *TNXB* gene, affected several *TNXB* transcripts. Only two from these six variants are located at different loci and consequently had unique identification labels (*TNXB* rs140304758 and *TNXB* rs61746206). Similarly, upon analysing the filtered results of the Centoskin gene panel, a variant which affects four transcripts of *PNPLA1* gene was shortlisted. The respective unique rs number was subsequently used to refer to these variants when browsing genomic databases and assessing pathogenicity. Ultimately, five SNVs remained after all filtering steps were applied, of which 3 are missense variants, whilst the remaining are 5' and 3' UTR variants respectively. These shortlisted variants, including their genomic position, alleles with alternative allele frequency and description are described in **Table 3.6**.

Table 3.6: List of shortlisted variants from filtering pipelines

A table showing shortlisted variants and relevant parameters. Allele frequencies obtained from gnomAD, gene description derived from www.genecards.org. AAF= Alternative allele frequency, Chr = chromosome, ENF=European non-Finnish, FA= Fanconi anaemia, Pos= Position, SNV= single nucleotide variant.

Gene	Location (Chr:Pos)	Variant (rs number)	Reference allele	Alternative allele	AAF (gnomAD ENF, %)	Variant type	Gene description
Centoskin gene panel							
Patatin-like phospholipase domain-containing 1 (*PNA1)*	6:36263171	c.745G>A, p.Glu249Lys, (rs45524833)	G	A	0.0030	SNV, Missense	Gene encodes Omega-hydroxyceramide transacylase, a catalytic enzyme for producing lipids in the epidermis and for keratinocyte differentiation
EDS gene panel							
Tenascin XB *(TNXB)*	6:32030129	c.6973C>T, p.Val2325Ile (rs140304758)	C	T	0.0010	SNV, Missense	Produces Tenascin-X, a protein which mediates ECM interactions especially in wound healing, expedites collagen formation and impedes cell migration.
	6:32065113	c.517C>T, p.Ala173Thr (rs61746206)	C	T	0.0194	SNV, Missense	
Skeletal Dysplasias							
SMAD family member 3 (*SMAD3)*	15:67390985	c.206+32287G>A (rs189286879)	G	A	0.0082	SNV, 5′ UTR	Encodes the SMAD3 protein, which is involved in signal transduction and in TGFβ signalling pathway to activate transcription.
Fanconi anaemia complementation group A (*FANCA)*	16:89804166	c.*843C>T (rs201316239)	G	A	0.0012	SNV, 3′UTR	Produces protein subunit in FA pathway associated with DNA repair and stability.
Zinc finger protein 276 (*ZNF276*)							Enables DNA binding activity and is involved in transcriptional regulation

3.3 Confirmatory testing by Sanger sequencing

Sanger sequencing was carried out at LGC Genomics (Germany) whilst the resulting chromatograms were viewed using Codon Code Aligner software (CodonCode Corporation 2019, version 9.0). Examples of homozygous reference and heterozygous genotypes for each tested variant are demonstrated in **Figure 3.4** below:

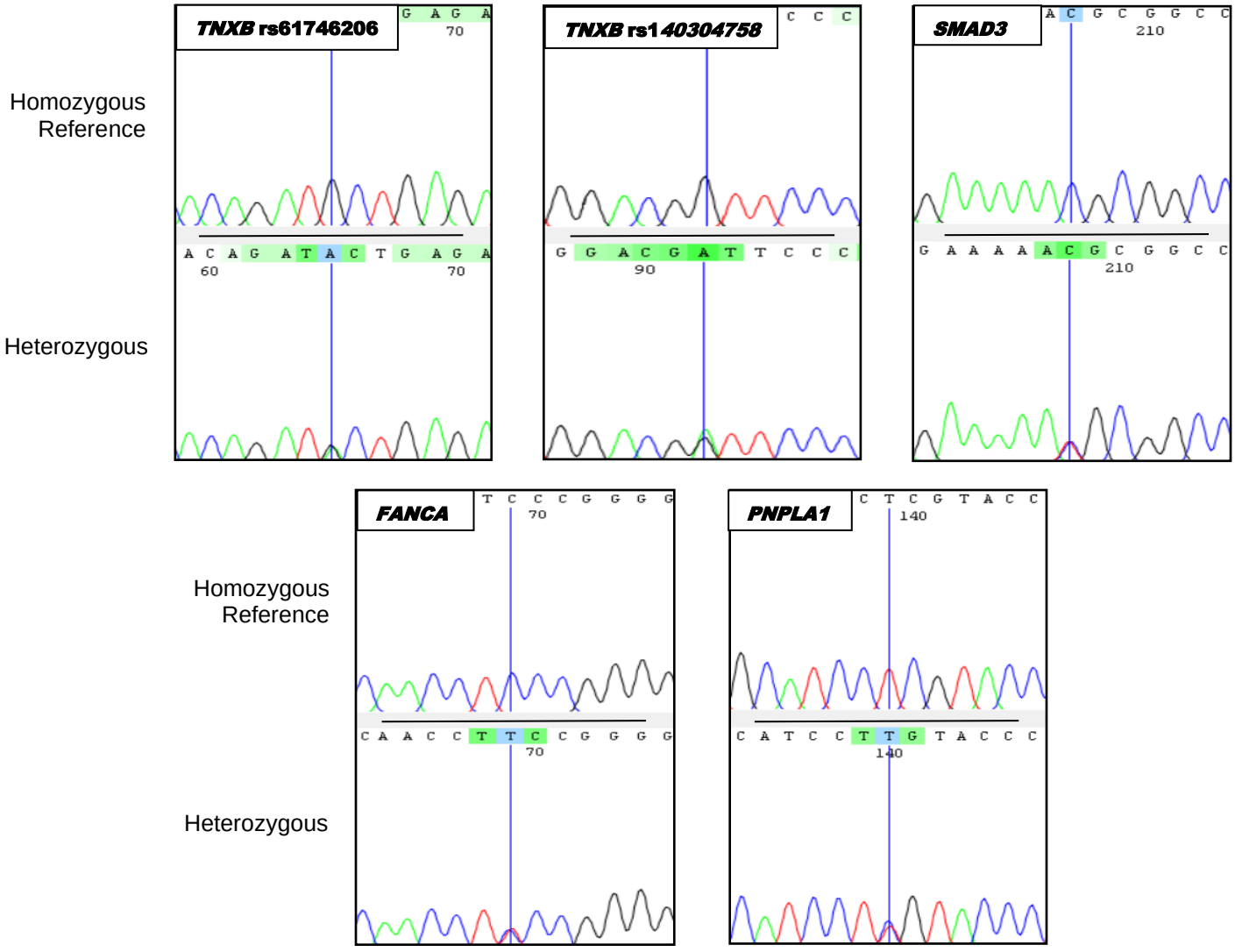


Figure 3.4: Detected genotypes by Sanger sequencing
Examples of chromatograms showing homozygous reference and heterozygous genotypes for each variant; allele position is indicated by a blue line. Two superimposing peaks corresponding to the wild type and alternative alleles are visible in the heterozygous row. Examples obtained from Sanger results of participants IV.9 (unaffected, wild type alleles) and IV.8 (affected, heterozygous alleles). Chromatograms viewed using Codon Code Aligner version 9.0.

The resulting chromatograms confirmed the results previously obtained by HTS. None of the filtered variants were detected in IV.9 and SDT-005. The affected family members who were included in the filtering pipeline (IV.1, III.5 and IV.8) were all heterozygous for the alternative allele, thus confirming the presence of all five variants in these participants. Two affected family members (III.2 and IV.2) were not sequenced by HTS but were tested by Sanger sequencing nevertheless to determine their genotype for these variants. In III.2, the alternative allele was detected for all tested variants, while only the variants in *SMAD3* and *FANCA* were detected in IV.2. Accordingly, the reference allele was detected for *TNXB* rs61746206 (c.517G>A), *TNXB* rs140304758 (c.6973G>A) and *PNPLA1* rs45524833 (c.745G>A) in this research participant (**Table 3.7**). A pedigree showing combined clinical and genetic findings can be found in **Figure 3.5**.

Table 3.7: Variant genotypes of recruited participants
A table showing the genotypes of all individuals for the filtered variants, as verified by Sanger sequencing. These variants were not detected in the non-affected participant (IV.9) and the un-related individual (SDT-005), while most of the affected family members (III.2, IV.1, III.5 and IV.8) were heterozygous for the alternative allele of each variant. Only *SMAD3* rs189286879 and *FANCA* rs201316239 were detected in IV.2.

Research Participant	**Variant & Genotype**				
	TNXB **rs61746206 (c.517G>A)**	***TNXB*** **rs140304758 (c.6973G>A)**	***SMAD 3*** **rs189286879 (c.G206+32287A)**	***FANCA*** **rs201316239 (c.*843C>T)**	***PNPLA1*** **rs45524833 (c.745G>A)**
III.2	CT	CT	GA	GA	GA
IV.1	CT	CT	GA	GA	GA
IV.2	CC	CC	GA	GA	GG
III.5	CT	CT	GA	GA	GA
IV.8	CT	CT	GA	GA	GA
IV.9	CC	CC	GG	GG	GG
SDT-005	CC	CC	GG	GG	GG

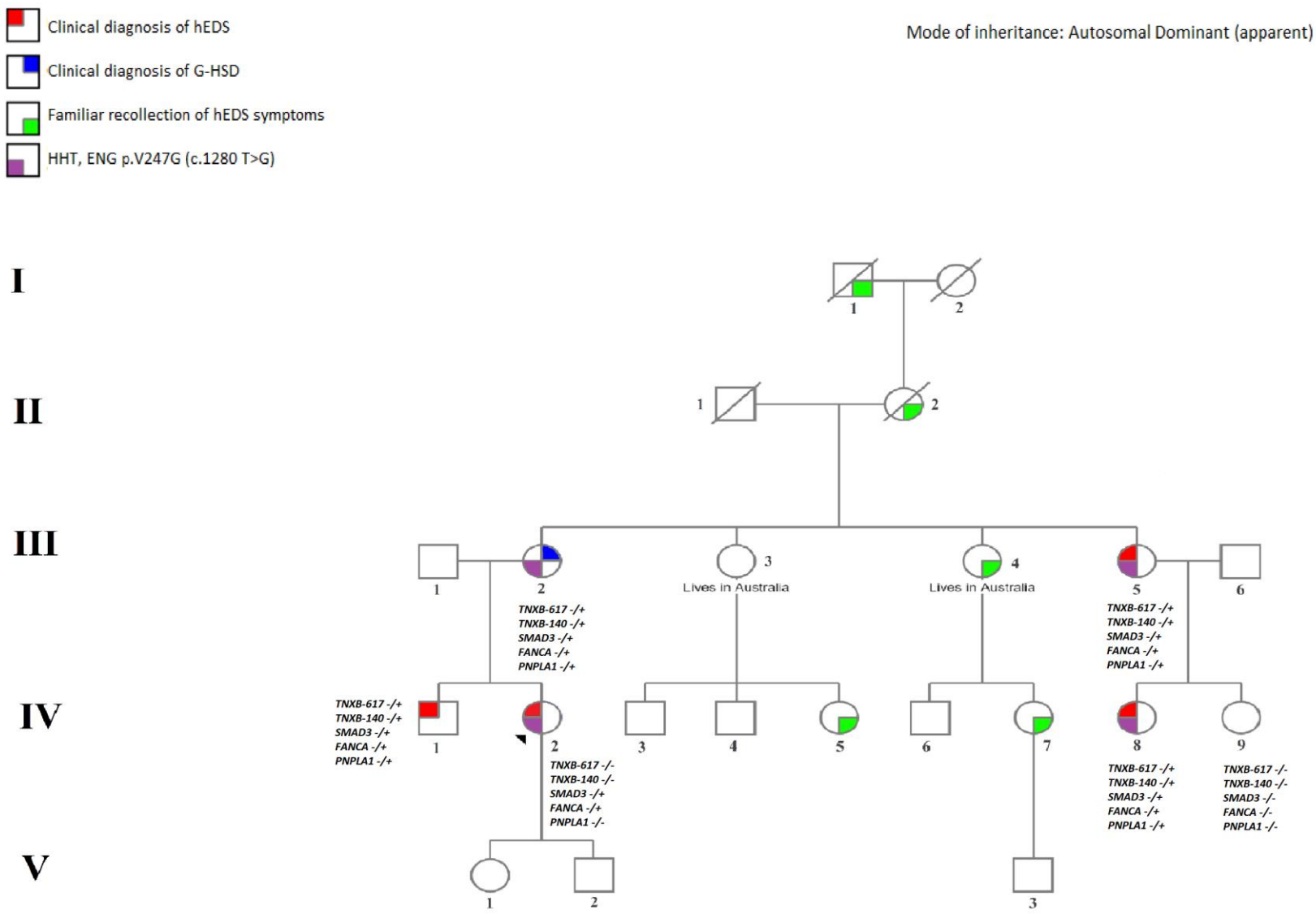


Figure 3.5: Genotyping results of recruited family members

A pedigree with legend of the recruited family, updated with the genotyping results obtained from Sanger sequencing. Heterozygosity is denoted by **-/+** while homozygosity of the reference allele is denoted by **-/-**. Genotyping results of SDT-005 are not shown in this figure, yet mirror those of IV.9. hEDS= Hypermobile Ehlers-Danlos Syndrome, HHT= Hereditary Haemorrhagic Telangiectasia , G-HSD= Generalised Hypermobility Spectrum Disorder, *TNXB-140=TNXB* rs140304758 ,*TNXB-617=TNXB* rs61746206

3.4 Variant analysis

The five shortlisted variants which were confirmed by Sanger Sequencing were analysed in further detail using tools described in Section 2.5 to determine their predicted effect, if any, on the resultant protein and their relevance to the pathophysiology of the disorder .

3.4.1 *TNXB* rs61746206 (c.517G>A)

This missense variant is located in the *TNXB* gene (chromosome 6), at position 32065113 on the forward strand of exon 3. It causes a cytosine to thymine base change at position 517 (c.517G>A), which in turn causes an Alanine to Threonine change at amino acid position 173 (p.Ala173Thr), as shown in **Figure 3.6**. This variant affects four transcripts of the same gene, of which three are protein coding (ENST00000375244.3, ENST00000375247.2, ENST00000479795.1) and one is a non-coding transcript (ENST00000486148.1) that has a retained intron. The transcript with the longest CDS which is usually the most expressed, also known as the canonical transcript, is transcript ENST00000375247.2 as identified by the Ensembl and Varsome databases.

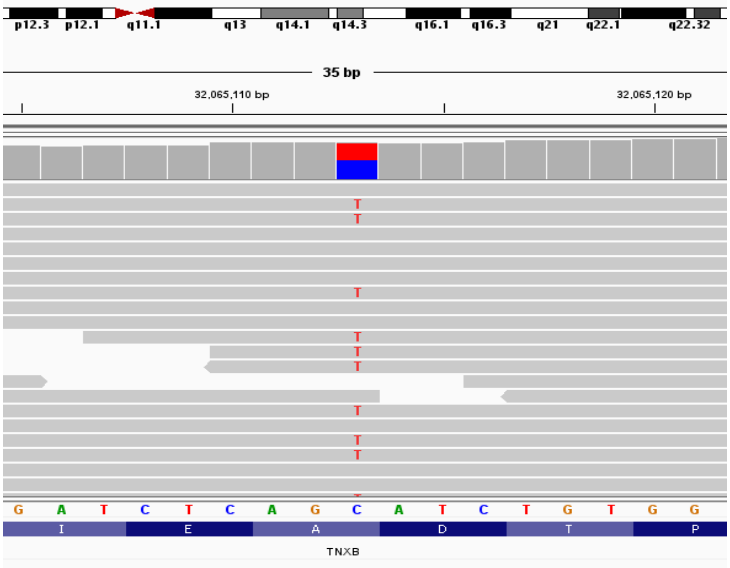


Figure 3.6: Variant *TNXB* rs61746206 in IV.8, as visualised using Integrative Genomics Viewer (IGV)
Variant causes a cytosine to thymine change at position 32065113 of chromosome 6, resulting in an amino acid change from Alanine to Threonine. Sufficient coverage of the area is observed, with 53% of reads detecting the alternative allele in this participant. No other variants were detected in the proximal area. Reference sequence and reference amino acid chain are shown at bottom of image.

It has a genomic AAF of 0.0194% in European non-Finnish populations and 0.0199% across all populations. The variant has four submissions on ClinVar, (last submission on 29th November 2022), one of which was linked to EDS. These submissions describe the variant as having a benign effect by three different classification schemes, including the 2015 ACMG/AMP guidelines.

Using the UCSC Genome Browser, the cytosine reference nucleotide at the variant locus is not conserved in various orthologues and does not appear to constitute an amino acid in chickens, frogs (*X.tropicalis*) and zebrafish, probably due to the significant evolutionary distance between humans and non-mammalian species. The reference amino acid at this position of the TNXB gene, Alanine, appears to be conserved in primates and other mammals, yet is not conserved in murine and canine orthologues (Figure 12). Both GERP++ (-8.83) and PhyloP (-1.17) scores indicate that the reference nucleotide is not conserved across species (**Figure 3.7**).

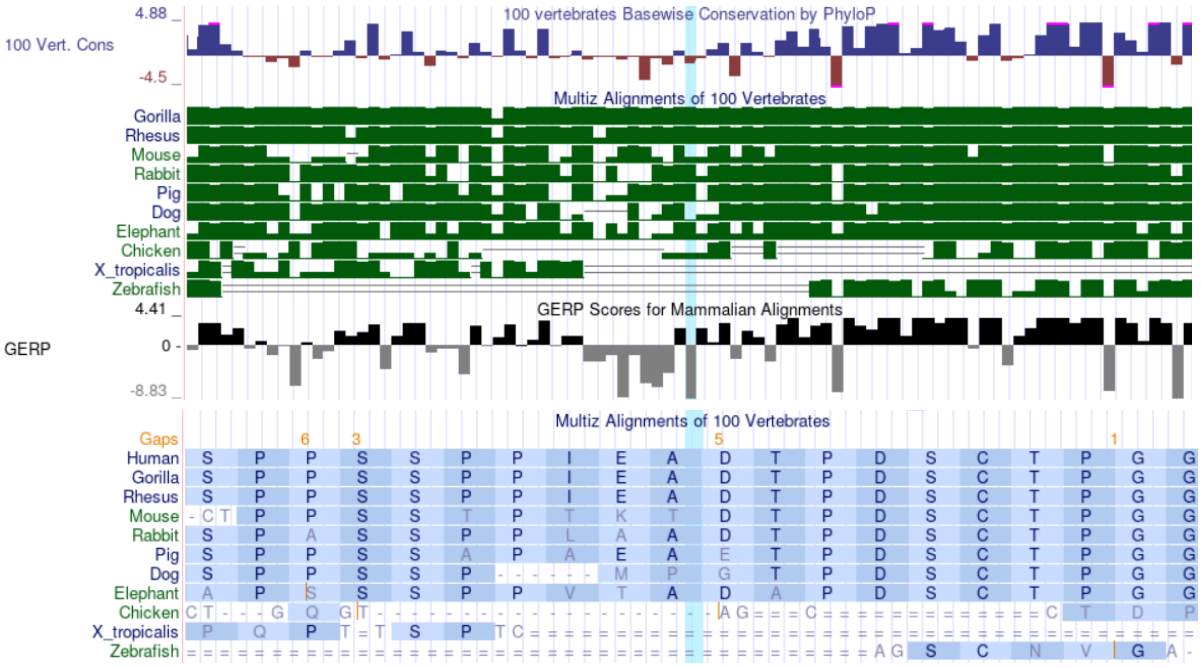


Figure 3.7: Conservation of the cytosine reference allele and coded amino acid across species
Output of conservation assessment tracks on UCSC genome browser for *TNXB* rs61746206 c.517G>A, showing (from top to bottom): a visual representation of base conservation score obtained by PhyloP, alignment of base position with other species, GERP base conservation score and conservation of the affected amino acid between species.

The variant was annotated by several prediction tools as having a benign effect on the transcript and eventual protein product (**Table 3.8**). This variant does not appear to affect protein function according to the Leider Open Variation database (LOVD, v.3.0 build 29, Fokkema, Taschner, et al. 2011).

Table 3.8: Predicted effect of *TNXB* rs61746206 on *TNXB* transcripts by various *in-silico* prediction tools, including pathogenicity scores

Green colouring denotes a benign/tolerated effect, orange a moderate effect, while red implies a deleterious/ probably damaging effect. Abbreviations: Low c= Low confidence, N/A= not available, T=tolerated, VEP= Variant effect predictor. ACMG evidence codes: 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, PP1= Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease, BS1= Allele frequency is greater than expected for disorder, BP4= Multiple lines of computational evidence suggest no impact on gene or gene product , BP6= Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation.

Transcript	ACMG/ AMP	SIFT	Polyphen	CADD	REVEL	MetaLR	Mutation taster	VEP	SNPeff	Mutpred2	FATHMM -XF
ENST00000375247.2 (Canonical)	Likely benign (PM1,PP1 ,BS1,BP4, BP6)	T	Benign	Benign	Benign	Benign	Benign	Moderate	Moderate	Benign	Benign
ENST00000375244.3	N/A	T	Benign	Benign	Benign	Benign	Benign	Moderate	Moderate	N/A	N/A
ENST00000479795.1	N/A	T, Low c (0.27)	Benign	Benign	Benign	Benign	Benign	Moderate	Moderate	N/A	N/A

A 3D protein model for the resultant Tenascin X protein is available on the Alphafold protein database (Jumper et al., 2021; version 2022-11-01,) with accession number C9J7W4. The variant is located at the 173rd residue of the protein out of a total of 4242 amino acids, however, the model is based on amino acids 1-842 (**Figure 3.8**). The reference amino acid at this position (Ala173) has a very low confidence score (pLDDT) of 39.25, which indicated low accuracy in the structure of the modelled region due to its location in an unstructured domain.

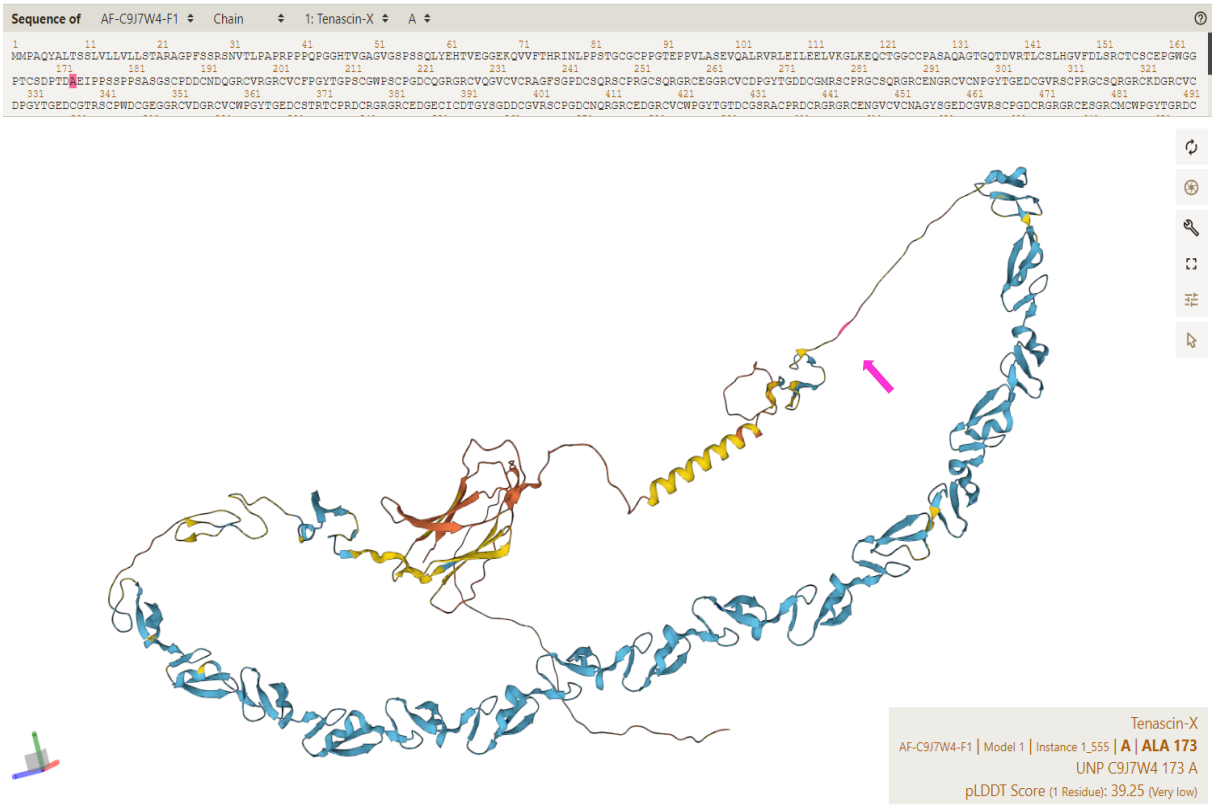


Figure 3.8: A 3D model of the human Tenascin X protein spanning amino acids 1-842
Figure shows the structural model of the first 842 residues of Tenascin X as displayed on Alphafold. Reference amino acid Alanine at residue 173 is highlighted with a pink arrow. Residue confidence score (pLDDT, bottom right of the image) at this position is 39.25, implying low confidence in the accuracy of the predicted structure since the amino acid resides in an unstructured region, as is also evident in the model.

The effect of variant *TNXB* rs61746206 (c.517G>A) on the protein's stability was assessed using CUPSAT (Parthiban, Gromiha, et al. 2006). The substitution of Alanine to Threonine at residue 173 appears to have a destabilising effect on the thermal denaturation of the protein, having a difference in free unfolding energy of -1.8 kcal/mol, while a stabilising effect (+1.91 kcal/mol) is predicted on chemical denaturation. In turn, the predicted structural and

biochemical effect of the Ala173Thr substitution on the protein was determined using Missense 3D (Ittisoponpisan, Islam, et al. 2019), by which no structural damage was identified. The amino acid substitution does not appear to increase cavity size, change the exposure of charged molecules or disrupt covalent bonding. Since the reference amino acid is Alanine, other properties pertaining to specific amino acids were not applicable. This amino acid substitution introduces two additional hydrogen bonds between the side and main chains of the amino acid (**Figure 3.9**).

Figure 3.9: Predicted impact of *TNXB* rs61746206 on the protein structure of TN-X
A 3D modelled structure of TN-X as shown on Missense 3D using model no.C9J7W4, showing the reference amino acid Alanine (top left, in blue) and substituted amino acid Threonine (top right, in red) at the 173rd residue. This alteration introduces two hydrogen bonds between the main chain and side chains of the substituted amino acid.

3.4.2 *TNXB* rs140304758 (c.6973G>A)

This missense variant is located on chromosome 6 at position 32030129 on the forward strand of exon 20, causing a cytosine to thymine nucleotide change at position 6973 (c.6973G>A). This single base substitution results in a Valine to Isoleucine amino acid change at the 2325th codon of the protein (p.Val2325Ile). *TNXB* rs140304758 (c.6973G>A) is multiallelic and thus has two alternative alleles- either adenine or thymine, the latter of which was detected in participants IV.1,III.5 and IV.8, as also shown in **Figure 3.10**.

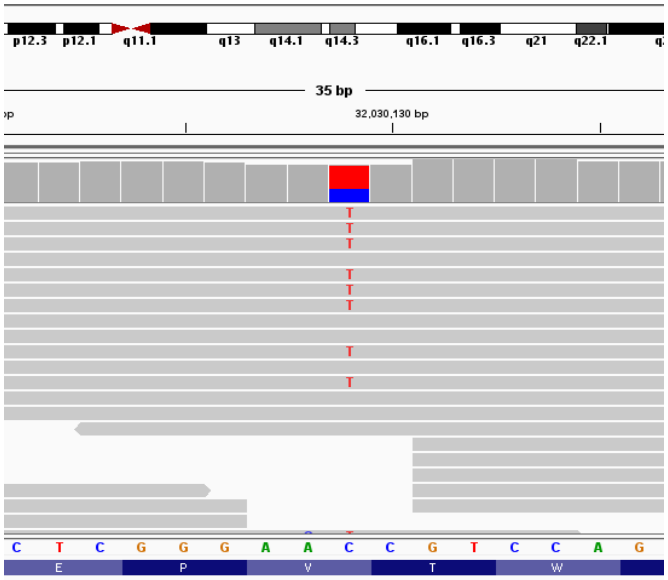


Figure 3.10: Output from IGV, showing variant *TNXB* rs140304758 in IV.8
Variant is visualised by a thymine (T, in red) substitution instead of the cytosine reference allele (show in blue, bottom of figure). The variant was detected in 65% of reads in this participant, while other family members had a similar allele ratio. No other variants were detected in the shown region. SNV causes an amino acid change from Valine to Isoleucine at codon 2325 of the TNXB protein (bottom).

Similar to *TNXB* rs61746206 (c.517G>A), the variant affects two coding transcripts of the *TNXB* gene: ENST00000375244.3 and ENST00000375247.2; however, it does not affect ENST00000479795.1. The substitution does not affect any splice-site regions and the length of the resulting protein is unaffected. This alteration has seven submissions on ClinVar (last submission on April 23rd, 2023), in which it is described as a variant of unknown significance (VUS) in five submissions, three of which associated the variant with classical-like EDS, a cardiovascular phenotype and vesicoureteral reflux. Variant is described as likely benign in the remaining two. Conservation of the base was calculated using PhyloP (0.036) and GERP++

(0.545) algorithms (**Figure 3.11**). The reference amino acid at the variant locus, Valine, is conserved in primates and other mammalian species, however, it appears to be converted to Isoleucine in canine orthologues. This is further demonstrated by PhyloP (0.036) and GERP++ (0.545) scores, both of which imply that the reference nucleotide is generally moderately conserved in the phylogenetic tree. The region does not appear to constitute similar, alignable coding sequences in evolutionarily distant species such as chickens, frogs and zebrafish.

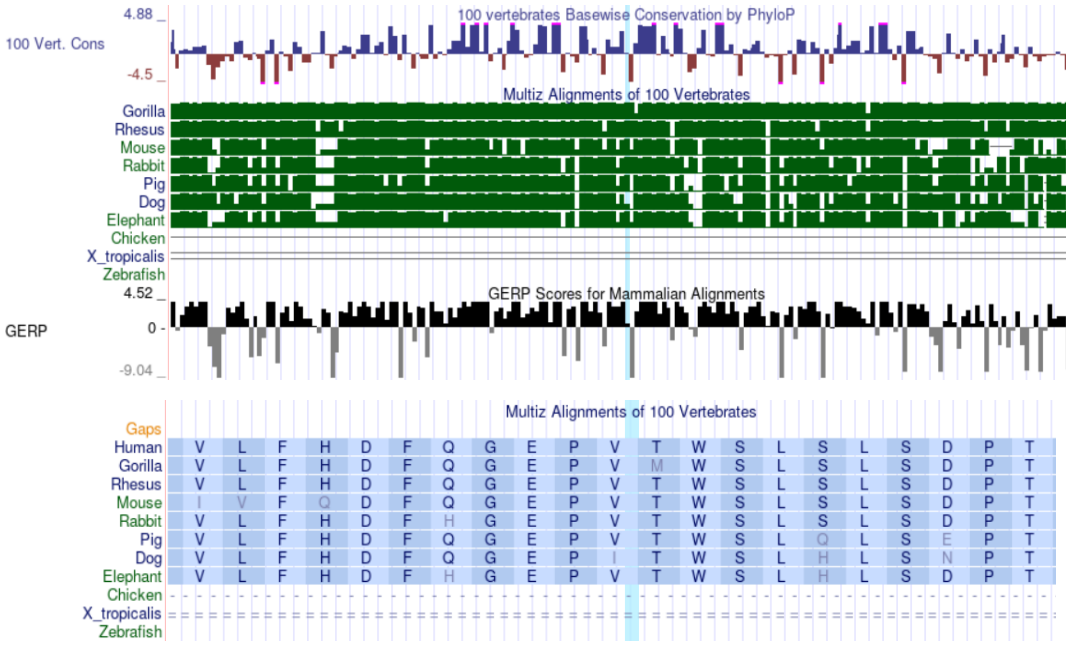


Figure 3.11: Conservation scores of the *TNXB* rs140304758 locus across several species alignments
A figure showing the output obtained by UCSC genome browser, indicating the conservation status of the reference amino acid Valine (denoted by a blue line) in other organisms in addition to base-wise PhyloP (top of the image) and GERP++ (middle) estimates.

Variant is found at a frequency of 0.0010% in European non-Finnish populations and had a global AAF of 0.0006%. The adenine alternative allele at this locus (c.6973G>T) has a less common frequency of 0.000% in European non-Finnish populations and is considered as deleterious by SIFT and Polyphen. In contrast, various variant prediction tools classified the thymine nucleotide change as benign or having a moderate effect (**Table 3.9**). Additionally, as with *TNXB* rs61746206 (c.517G>A), this variant does not appear to affect protein function according to the LOVD database (Fokkema, Taschner, et al. 2011 ; v.3.0 build 29).

Table 3.9: A table showing the *in-silico* putative effect of *TNXB* r140304758 on gene transcripts, using variant effect prediction tools and deleteriousness scores
Green colouring denotes a benign/tolerated effect, orange a moderate effect, while grey implies a variant of uncertain significance (VUS). Abbreviations: N/A= not available, T= Tolerated. ACMG evidence codes: PP1: Co-segregation with disease in multiple affected family members in a gene definitively known to cause the disease, PM2: Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium, BP4: Multiple lines of computational evidence suggest no impact on gene or gene product (conservation, evolutionary, splicing impact, etc.)

Transcript	ACMG	SIFT	Polyphen	CADD	REVEL	MetaLR	Mutation taster	VEP	SNPeff	Mutpred2	FATHMM -XF
ENST00000375247.2 (Canonical)	Uncertain significance (PM2,PP1, BP4)	T	Benign	Benign	Benign	Benign	Benign	Moderate	Moderate	Benign	Benign
ENST00000375244.3	N/A	T	Benign	Benign	Benign	Benign	Benign	Moderate	Moderate	N/A	N/A

3.4.3 *SMAD3* rs189286879 (c.206+32287G>A)

The SNV rs189286879 (c.206+32287G>A), located in the SMAD family member 3 gene (*SMAD3*), was also identified in the family which appears to segregate with the disease phenotype. It is located on the forward strand of chromosome 15 at position 67390985, affecting non-canonical coding transcript ENST00000559092.1, in which it is the 66th residue out of a total of 584 nucleotides. In ENST00000559092.1, the variant is located in a 5' untranslated region, 227 nucleotides prior to the start of the coding sequence of the first exon (c.-227G>A) and 378 bases upstream of the nearest splice-site. In the canonical transcript of *SMAD3 (*ENST00000327367.4), the variant resides in the first intron (c.206+32287G>A). The variant is multiallelic, causing either a guanine to adenine or a guanine to cytosine base substitution. However, in this family, the adenine alternative allele was identified (**Figure 3.12**). Consequently, as the variant is located in an untranslated region, the amino acid sequence of the translated protein is not affected.

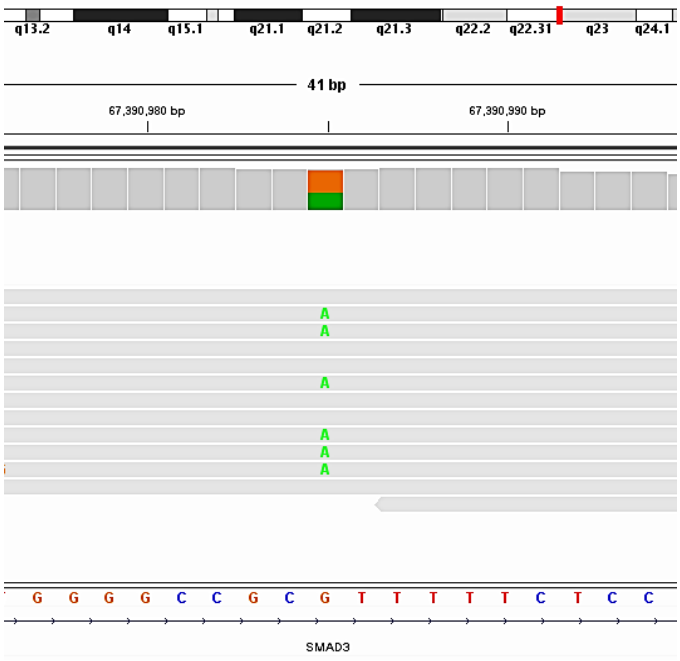


Figure 3.12: Variant *SMAD3* rs189286879) in IV.8
Alternative allele adenine (shown in green) is substituted instead of the guanine reference allele (bottom of image, coloured in orange), which was detected in 45% of reads. Other members of the family have a similar allele ratio. No other variants were detected in the area proximal to the variant locus. Figure obtained from IGV interface.

SMAD3 rs189286879 (c.206+32287G>A) is found at a frequency of 0.0062% in global genomes and had a marginally higher AAF of 0.0082% in European non-Finnish populations. The highest allele frequency is found in Finnish genomes, at 0.0198%. In contrast, the cytosine alternative allele is found at a rarer frequency of 0.00002%. The variant detected in this family (c.-227G>A) is not yet reported on Clinvar nor the LOVD database and thus its pathogenicity in relation to clinical phenotypes cannot be assessed. Similarly, as the variant is upstream of the exon, is not an insertion/deletion or a copy number variation and does not affect the amino acid sequence, its benign or deleterious effect could not be calculated using SIFT, PolyPhen, Intervar or Mutpred2 prediction algorithms. Pathogenicity predictions by Varsome and Fathmm-XF are only available for the canonical transcript only and are thus also not applicable in this case. Nevertheless, MutationTaster classifies this variant as benign, with no perturbations of potential splice-sites while SnpEff denotes a 'modifier' impact. Similarly, the CADD score for this SNV is calculated as 12, which denotes a likely benign effect.

The reference allele appears to be conserved in primates and other mammalian species except in murine orthologues, in which the reference allele is adenine (**Figure 3.13**). However, as observed in the other variants featured in this study, in evolutionarily distant species such as chickens, frogs and zebrafish, the same locus constitutes unalignable bases and therefore conservation scores cannot be estimated. Phylo P score (0.176) implies slight conservation whilst the GERP++ score (-0.756) denotes non-conservation of the reference nucleotide.

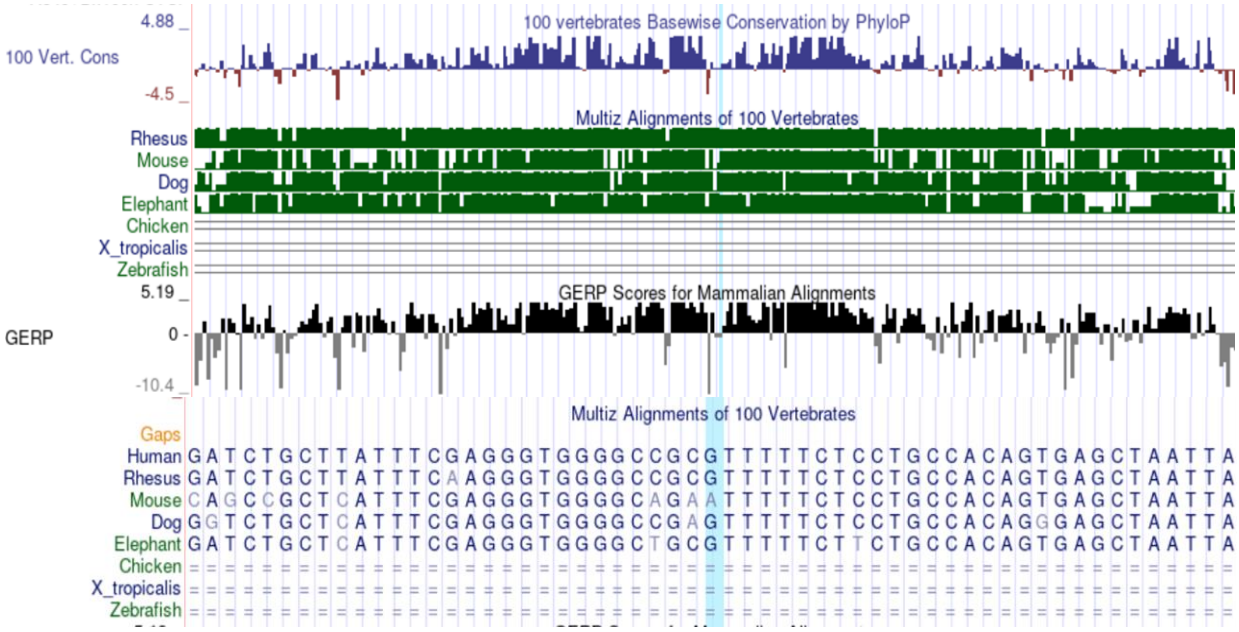


Figure 3.13: Conservation tracks for *SMAD* rs189286879 (highlighted with a blue line) on UCSC genome browser

Figure visualises conservation scores by PhyloP (top) and GERP (middle) as well as comparisons of the conservation scores and the nucleotide sequence across several species (Multiz alignments, bottom).

3.4.4 *FANCA*/*ZNF276* rs201316239 (c.*843C>T)

Another variant detected in the family is *FANCA*/*ZNF276* rs201316239 (c.*843C>T), located at position 89804166 on the forward strand of chromosome 16. This area corresponds to two overlapping genes: FA Complementation Group A (*FANCA)* and Zinc finger protein 276 (*ZNF276*). In *FANCA*, the variant is located in the 3' untranslated region downstream of exon 43 in the canonical transcript ENST00000389301.3, 951 bases downstream of the nearest splice-site. In *ZNF276*, the variant locus resides in the 9th intron of non-canonical transcript ENST00000569901.1, which is a non-coding exon transcript. Similar to other variants in this study, *FANCA/ZNF276* rs201316239 (c.*843C>T) is multiallelic, causing a guanine to adenine (c.*843C>T) or guanine to cytosine (c.*843C>G) base substitution. HTS data indicates that the adenine alternative allele is present in the heterozygous state in hEDS members of the family (**Figure 3.14**). Since the regions in which the variant is located consist of non-translated

sequences, the amino acid sequences of both the FANCA and the ZNF276 proteins remain unaffected.

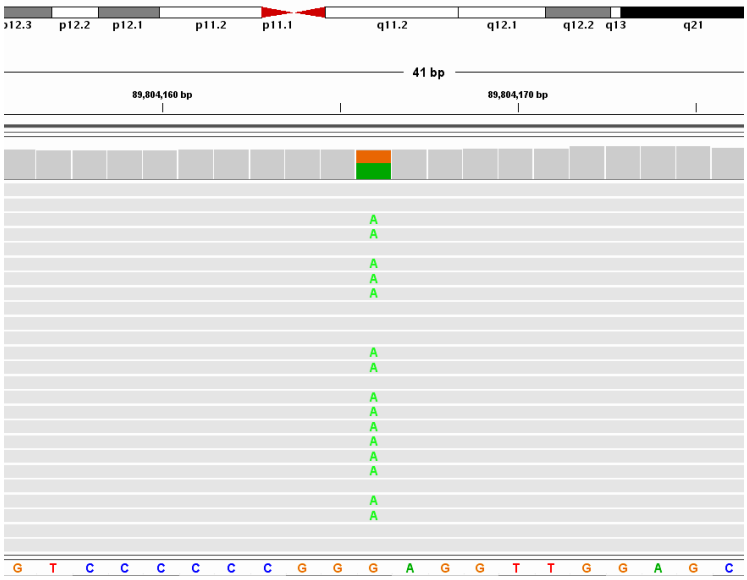


Figure 3.14: Variant c.*843C>T in IV.8
The guanine reference nucleotide (bottom of image, in orange) is substituted to adenine (shown in green).Alternative allele is in a heterozygous state and was detected in 57% of reads. This genomic region is observed by IGV to be adequately covered, with no other detected variants around the variant locus. Output shown above is also representative of data from other members of the family. Figure was obtained from the IGV interface .

On gnomAD, *FANCA/ZNF276* rs201316239 (c.*843C>T) has a rare global AAF of 0.0008% and a frequency of 0.0012% in European non-Finnish genomes, the highest frequency across all populations. The other variant allele at this locus, cytosine (c.*843C>G), is found at an extremely rare frequency of 0.00000882% in European non-Finnish populations and 0.00000398% in global populations. *FANCA/ZNF276* rs201316239 (c.*843C>T) has one submission on Clinvar, submitted on May 31st, 2020, in which it is described as a VUS. The LOVD database also classifies this variant as having an unknown effect. Similar to the *SMAD3* variant, the *in-silico* predicted effect of the variant could not be elucidated by multiple variant effect prediction tools as the variant either does not alter the amino acid sequence or is not the canonical transcript in the case of ENST00000569901.1. However, the Varsome database and Fathmm-XF prediction tools classify this variant in transcript ENST00000389301.3 as likely benign and benign respectively. In addition, MutationTaster also classifies the variant as

benign, with no abrogation of potential splice-sites. Using the UCSC genome browser, the reference allele has a PhyloP score of 1.14 and a GERP++ score of 1.78 (**Figure 3.15**). These scores indicate that the reference allele at the variant locus appears to be moderately conserved across mammalian species. The base is not conserved in murine orthologues, in which the reference allele is adenine rather than guanine. As observed in other variants in this study, the conservation status of the nucleotide at this position cannot be compared to in the case of amphibians as well as in avian and piscine species as these orthologues consist of un-alignable bases or species-specific insertions/deletions at this position.

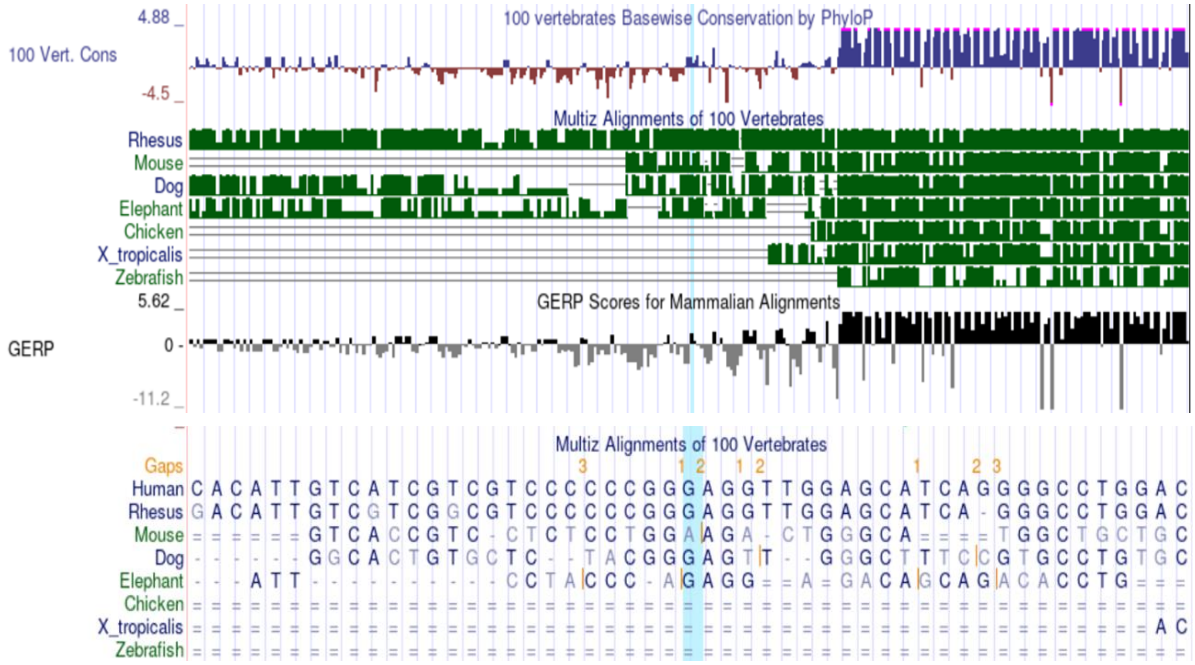


Figure 3.15: Conservation tracks for the region surrounding *FANCA/ZNF276* rs201316239 (locus highlighted with a blue line) as viewed on the UCSC genome browser
Figure shows PhyloP (top) and GERP++ (middle of image) base-wise conservation charts, as well as the nucleotide sequence across several species alignments (bottom). In the latter, the longest nucleotide gap as compared to the human genome is denoted the Gaps row, coloured in orange.

3.4.5 *PNPLA1* rs45524833 (c.745G>A)

This missense variant is located in *PNPLA1* on chromosome 6, at position 36263171 on the forward strand of exon 5, where a guanine nucleotide is substituted with an adenine base. This results in a Glutamic acid to Lysine amino acid change (**Figure 3.16**) in the Omega-hydroxyceramide transacylase protein. However, since this single base alteration affects four

PNPLA1 transcripts at different positions, the position of the alteration differs between these transcripts: in ENST00000312917.5, *PNPLA1* rs45524833 (c.745G>A) is located at the 487th nucleotide (c.487G>A, p.Glu163Lys), in ENST00000394571.2 at the 745th (c.745G>A , p.Glu249Lys) , in transcript ENST00000457797.1 at base 748 (c.748G>A, p.Glu250Lys), whilst in ENST00000388715.3, variant is located at the 460th nucleotide (c.460G>A, p.Glu154Lys). The canonical transcript for *PNPLA1* is ENST00000394571.2, constituting of 532 amino acids, as identified using Ensembl database and as annotated by SNPeff. As *PNPLA1* rs45524833 (c.745G>A) is a missense SNV, the length of the resultant protein is not predicted to be affected.

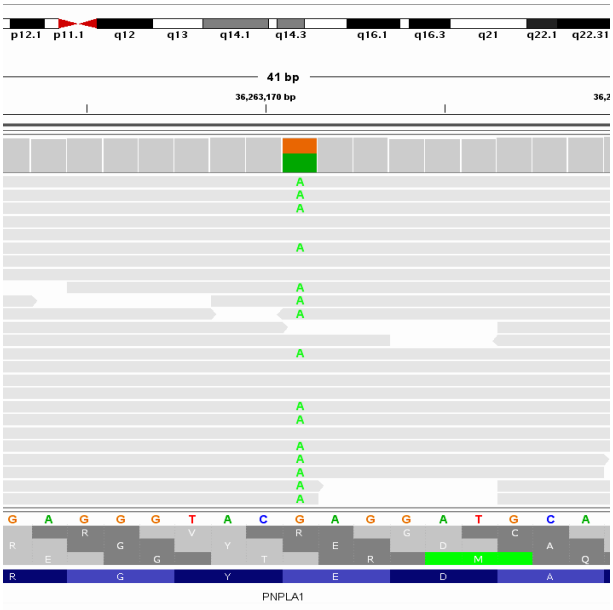


Figure 3.16: Variant *PNPLA1* rs45524833 in IV.8, as visualised by IGV
SNV causes a guanine to adenine (A, in green) nucleotide substitution in 56% of reads, resulting in an amino acid change from Glutamic acid to Lysine (abbreviated as E, in bottom of figure). Sequencing data of other family members show a similar allele ratio. No variants were detected around the variant locus, whilst the region shows an adequate level of coverage.

Using gnomAD data , this *PNPLA1* variant is found at a low frequency of 0.0030% in European non-Finnish population whilst having the same AAF in global genomes. It is described on ClinVar as a likely benign variant using the 2015 ACMG/AMP guidelines. This interpretation is based on five submissions, with the most recent submitted on the 7th February 2023. Of these five entries, two were from patients with AR congenital ichthyosis while the phenotype of the

other three submissions was not specified. The reference allele guanine has a PhyloP score of 2.5077 and GERP++ score of 3.97 (**Figure 3.17**). The reference amino acid as well as the reference allele at position 36263171 appears to be well-conserved in other primates and other mammals yet is not conserved in murine orthologues. The region does not seem to constitute coding elements in evolutionarily distant species such as chickens, frogs and zebrafish.

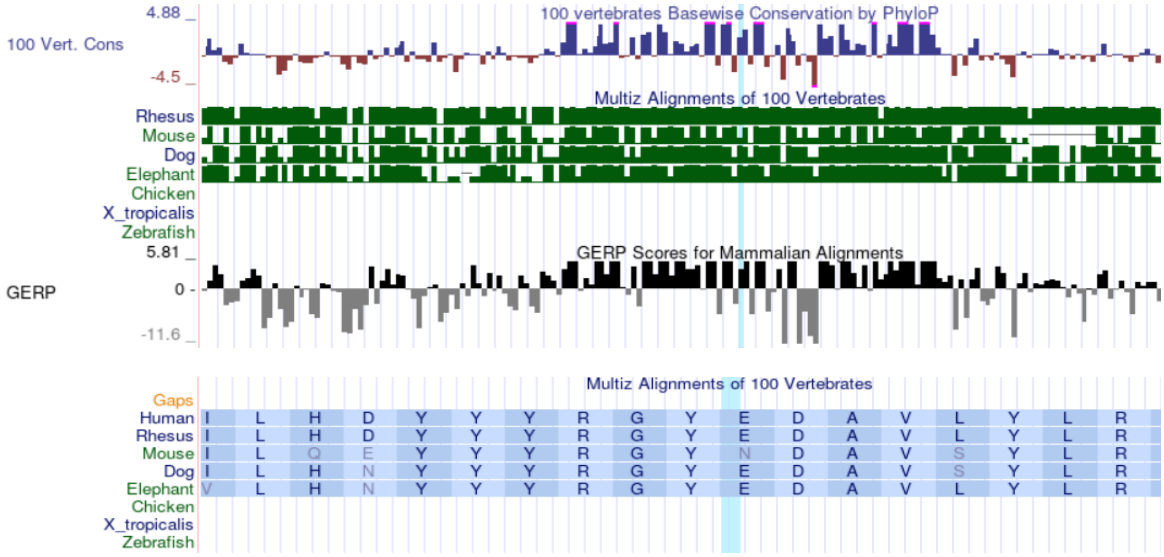


Figure 3.17: Conservation estimates of the reference allele (top of figure, position denoted by a blue line) and reference amino acid (bottom) for the variant locus of *PNPLA1* rs45524833, as displayed on the UCSC genome browser

The MutPred2 prediction tool deems this variant as pathogenic (0.548) due to a pathogenicity cut-off score of 0.5 (**Figure 3.18**). In contrast, multiple *in-silico* variant prediction tools classify this variant as not deleterious (**Table 3.10**). This classification is further supported by the LOVD database, which also classifies the variant as likely benign with no change in protein function.

ID	Substitution	MutPred2 score	Remarks	Affected PROSITE and ELM Motifs
PNPLA1	E249K	0.548	Predicted conservation scores	ELME000081, ELME000233, PS00007
Molecular mechanisms with P-values <= 0.05			Probability	P-value
Altered Metal binding			0.54	3.2e-04
Loss of Sulfation at Y245			0.04	0.01

Figure 3.18: Results of the MutPred2 web server for Glu249Lys in *PNPLA1*, showing a pathogenicity score of 0.548 , putative altered motifs and perturbed molecular mechanisms

Table 3.10: Putative effect and deleteriousness scores of *PNPLA1* rs45524833 on *PNPLA1* transcripts, using several *in-silico* prediction tools

Green colouring denotes a benign/tolerated effect, orange a moderate effect, red colouring signifies a pathogenic effect while grey implies a variant of uncertain significance (VUS). Abbreviations: N/A= not available, T= tolerated. ACMG evidence codes: 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, BP6: Reputable source recently reports variant as benign, but the evidence is not available to the laboratory to perform an independent evaluation.

Transcript	ACMG	SIFT	Polyphen	CADD	REVEL	MetaLR	Mutation taster	VEP	SNPeff	Mutpred2	FATHMM-XF
ENST00000394571.2 (Canonical)	Uncertain significance (PM1, BP6)	T	Benign	Benign	Benign	Benign	Likely Benign	Moderate	Moderate	Pathogenic	Benign
ENST00000457797.1	N/A	T	Moderate	Benign	Benign	Benign	Likely Benign	Moderate	Moderate	N/A	N/A
ENST00000312917.5	N/A	T	Benign	Benign	Benign	Benign	Likely Benign	Moderate	Moderate	N/A	N/A
ENST00000388715.3	N/A	T	Benign	Benign	Benign	Benign	Likely Benign	Moderate	Moderate	N/A	N/A

A 3D protein model for the Omega-hydroxyceramide transacylase protein product of *PNPLA1* is available on the Alphafold protein database (version 2022-11-01, Jumper et al., 2021) with accession number AF-Q8N8W4-F1. The variant locus on the canonical transcript is located at the 249th residue of the protein out of a total of 532 amino acids and resides in an alpha helix domain (**Figure 3.19**). The reference amino acid at this position (Glu249) has a high residue confidence score (pLDDT) of 96.37 while the average confidence across the entire model is calculated as 65.44.

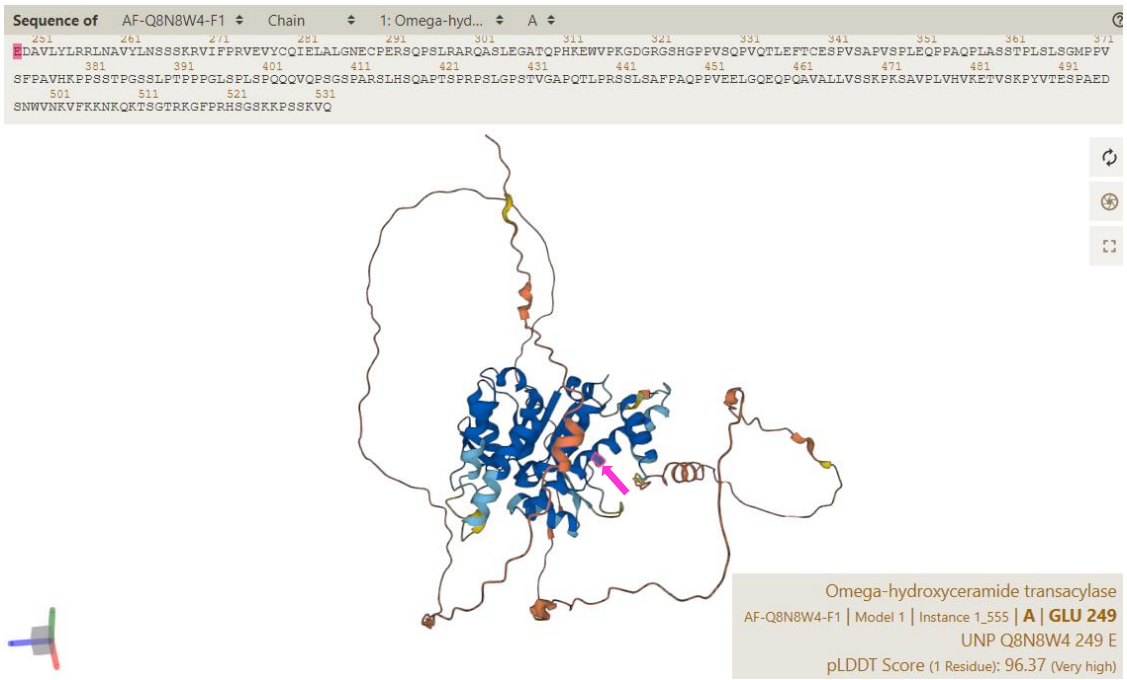


Figure 3.19: A structural model of Omega-hydroxyceramide transacylase, translated from the *PNPLA1* transcript ENST00000394571.2
Figure shows the three-dimensional model structure of the protein as displayed on Alphafold, with the reference amino acid of the 249th residue highlighted by the website's interface and manually by a pink arrow. A residue confidence score (pLDDT, bottom right of the image) of 96.37 implies high confidence in the accuracy of the residue in the 3D model.

The effect of the variant on the protein's stability was assessed using CUPSAT (Parthiban, Gromiha, et al. 2006). The substitution of Glutamic acid for Lysine appears to have a stabilising effect on the thermal denaturation of the protein, having a difference in free unfolding energy of +0.45 kcal/mol, whilst a destabilising effect (-0.13 kcal/mol) is predicted on chemical denaturation.

The putative effect of the variant on this protein model was also assessed using Missense3D (Ittisoponpisan, Islam, et al. 2019). The amino acid substitution is not predicted to alter the secondary structure, does not change the exposure of charged residues nor does it perturb biochemical properties such as hydrophobic/hydrophilic interactions, disulphide bridges or salt bridges. Since the reference amino acid is Glutamic acid, other amino acid-specific changes are also unaffected. Nevertheless, this substitution appears to expand the cavity volume in the region by 171.504 ångströms (Å^3) and disrupt one side chain hydrogen bond (**Figure 3.20**).

Figure 3.20: A 3D model of the Omega-hydroxyceramide transacylase protein
Figure shows the Glutamic Acid reference residue highlighted in blue (top left) and the mutant amino acid Lysine highlighted in red (top right). This substitution results in cavity volume expansion of 171.504 ångströms (bottom left) while disrupting side chain hydrogen bonding between the affected residue and the previous amino acid (bottom right). Images obtained from Missense 3D protein alteration prediction tool.

Chapter 4: Discussion

The advent of HTS and novel molecular techniques have highlighted valid candidate genes and molecular pathways associated with hEDS. However, the number of research studies on the genetics of hEDS using the updated 2017 diagnostic criteria is still limited. The underlying aetiology is still poorly defined, being further confounded by a high degree of phenotypic and genetic heterogeneity.

As the disorder has never been researched in Malta, there is no published clinical and genetic data on hEDS patients in Malta. In consideration of this lacuna in local scientific research, this study sought to identify causative pathogenic gene variants as well as to test the effect of hEDS on bone integrity in local participants. It is also the first study to date to report the presence of hEDS in conjunction with HHT, another rare hereditary disorder (Shovlin, Buscarini, et al. 2022).

A multigenerational Maltese family and an unrelated individual (SDT-005) with an hEDS phenotype were recruited for the study. Their medical histories were obtained using a purposely designed questionnaire and were later investigated by WES (IV.8) and WGS (III.2, IV.1, III.5, and IV.9) analyses. Using an in-house filtering pipeline, five variants were shortlisted, and subsequently analysed using genetic databases and *in-silico* prediction tools to determine their relevance to hEDS. This analysis was also supplemented by bone-related parameters, namely BMD measurements and blood biochemical studies.

The recruited family was selected based on predetermined inclusion and exclusion criteria. A characteristic hEDS phenotype was observed in more than three individuals across two generations, similar to recruited participants in Castori, Dordoni, et al. (2014). The participation of healthy participants from this pedigree was an important asset when filtering identified

genetic variants in affected relatives. At least one clearly affected male (IV.1) could be recruited in order to compare the difference in the severity of symptoms between genders, as hEDS males are reported to have a milder phenotype (Copetti, Morlino, et al. 2019). AD inheritance was clearly demonstrated , consistent with that reported in the literature (Tinkle, Castori, et al. 2017). As hypermobility is prevalent in children under 7 years of age (Beighton, De Paepe, et al. 1998), V.1 and VI.2 were not recruited so as to avoid inaccurate diagnoses due to age, an exclusion criterion previously implemented by Ritter, Atzinger, et al. (2017). All recruited participants were clinically examined by a consultant medical geneticist with substantial experience in HCTDs, who had also followed the clinical progression of these participants for several years.

The medical history of III.2 revealed a positive Beighton score, chronic pain, cutaneous features. and systemic complications associated with hEDS (Refer to Appendix J). However, due to an absence of joint dislocations and other features described in Criterion 2 of the diagnostic checklist, she could not be diagnosed as having hEDS. According to the contemporary nomenclature, this participant was instead classified as having G-HSD (Castori, Tinkle, et al. 2017). While the co-existence of various forms of hypermobility disorders within the same family is documented in the literature, it is possible that this participant would be diagnosed with hEDS if re-assessed using updated diagnostic criteria.

The collection and analysis of clinical data from participants in this study supported the findings of most published studies describing clinical manifestations associated with hEDS. The sexual dimorphism described by Castori and Camerota (2010) and Copetti, Morlino, et al. (2019) was evident when comparing the severity and extent of the participants' symptoms. Additionally, despite the fact that G-HSD is considered less severe than hEDS according to the contemporary 2017 criteria (refer to section 1.4.3), the phenotype of III.2 closely resembled that of her female hEDS relatives. This also supports the conclusions of Copetti, Morlino, et

al. (2019) regarding the lack of distinction between hEDS and HSD in disease severity using the current terminology.

Cutaneous manifestations were present in all female participants including IV.9, but were absent in IV.1. A possible explanation for these two observations relates to the multifactorial threshold model proposed by Castori et al (2011), which describes the emergence of the hEDS phenotype upon the combination of several genetic and environmental factors (Castori, Sperduti, et al. 2011). An underlying variant or a perturbed signalling mechanism affecting the dermal tissue, which has a higher penetrance in females, may be present and further exacerbated by other pleiotropic variants/environmental factors in some individuals to result in a hEDS phenotype. This model might explain why cutaneous features were shared between the female participants of this study regardless of their hEDS/G-HSD status, but not with IV.1.

Despite the fact that III.2, IV.8 and IV.9 as well as SDT-005 reported the use of daily Vitamin D supplements, III.2 and SDT-005 were the only participants with 25-OH Vitamin D levels within the normal range (30-100ng/ml). The discrepancy between the reported supplement intake and 25-OH Vitamin D levels in IV.8 and IV.9 can be attributed to either undisclosed non-compliance with a daily regimen or recent initiation of supplement use, as serum levels of 25-OH Vitamin D take approximately six months to reach adequate levels (Zhao, Gardner, et al. 2015). As Vitamin D deficiency is prevalent in the local population and in other European Countries (Cashman, Dowling, et al. 2016, Times of Malta 2018), Vitamin D deficiency in these cases might not be directly associated with hEDS, as reported by Mazziotti, Dordoni, et al. (2016), but can be attributed to various other genetic and environmental factors in Europeans. The lack of correlation between hEDS and Vitamin D levels would support the results of a study by Eller-Vainicher, Bassotti, et al. (2016), in which there was no difference in serum levels of 25-OH Vitamin D, total Calcium and Phosphorus between hEDS participants and controls.

Plasma PTH and serum levels of magnesium, albumin, phosphate and total ALP in family members, were within the normal ranges, indicating maintained PTH regulation and normal phosphorus haemostasis respectively. A minor elevation of ionised calcium was noted in III.2, IV.2 and III.5 (refer to Table 3.2). Significant elevations in ionised calcium levels suppress PTH secretion in a negative feedback loop (Khan, Jose, et al. 2023). However, this was not observed in PTH measurements in this study as ionised calcium elevations in the family were minute and unlikely to be of any physiological significance. Of particular note is participant III.2, whose serum PTH levels were within the normal range with a slight increase (+0.04mmol/L) in ionised calcium despite having low BMD scores at the LS,FN and TH (see below). Ionised calcium levels in SDT-005 were markedly elevated (2.32mmol/l) yet PTH levels were within the normal range. Furthermore, a decrease in phosphorus levels was also present in SDT-005, despite normal levels of total ALP.

Previous studies which compared BMD measurements between the two Hologic densitometers used in this study found no significant differences in BMD values at multiple anatomical sites (Sasivimol, Amnuaywattakorn, et al. 2022) with comparable parameters of precision (Whittaker, McNamara, et al. 2018, Sasivimol, Amnuaywattakorn, et al. 2022). Thus, whilst the elimination of measurement and instrument bias is ideal and constitutes good research practice (Šimundić 2013), any differences in the methods of BMD measurement by the Horizon and Discovery DXA densitometers in this study are deemed to be minimal and insignificant.

Participants IV.1, IV.2 and III.5 all had Z-scores greater than -1.0, indicating normal bone turnover despite their Vitamin D deficiency (**Table 3.2**) and diagnosis of hEDS. Contrarily, two participants (III.2 and SDT-005) were classified as osteopenic using the WHO classification (Kanis, Burlet, et al. 2008). The low T-scores at the LS, FN and TH in III.2 (-1.6,-1.9, -1.0 respectively) and SDT-005 (-2.0,-1.8,-1.9 ; refer to Table 3.3) support the findings of previous studies (Annapureddy, Block, et al. 2014, Dolan, Arden, et al. 1998, Eller-

Vainicher, Bassotti, et al. 2016, Theodorou, Theodorou, et al. 2012, Gulbahar, Sahin, et al. 2006). This is the first study to test this association using contemporary diagnostic criteria (Malfait, Francomano, et al. 2017).

SDT-005 and III.2 (aged 48 and 57 respectively), who were discovered to be osteopenic, are the eldest participants recruited in this study. They are followed by III.5 (age 44), who had normal BMD values in LS,FN and TH. The natural degeneration of bone mass during and after the fourth decade of life is well documented in the literature describing osteoporosis and bone physiology (Demontiero, Vidal, et al. 2012). Abrupt decline of oestradiol at menopause results in the disruption of the balance between bone resorption and bone formation in the bone remodelling cycle, tilting the balance in favour of bone resorption (Cauley 2015). Enhanced bone resorption results in deterioration of the micro-architecture of bone tissue, leading to osteopenia and osteoporosis in the absence of treatment, such as antiresorptive agent therapy (Chen and Sambrook 2011, Berger, Langsetmo, et al. 2008). As no direct evident association between hEDS status and bone degeneration was observed in this study, age and menopausal status remain a more probable causative factor of low bone mass in these two participants. However, whether the presence of hEDS further accelerates bone degeneration in post-menopausal females is yet to be determined.

Analysis of HTS data filtered using the devised filtering pipeline did not uncover any variants in *COL1A1, COL1A2, COL3A1* ,*COL5A2* and *COL6A3*, confirming the lack of previously associated variants in the collagen-encoding genes (Weerakkody, Vandrovcova, et al. 2016, Henney, Brotherton, et al. 1992, Narcisi, J.Richards, et al. 1994). Furthermore, no variants in *LZTS1* (Syx, Symoens, et al. 2015) were shortlisted, thus excluding the causative role of this gene in the local recruited participants. However, two heterozygous SNVs in *TNXB* were shortlisted in sequenced hEDS participants. *TNXB* variants in hEDS participants have also been previously identified in other studies (Zweers, Bristow, et al. 2003, Zweers, Kucharekova, et

al. 2005, Yamada, Watanabe, et al. 2019). Additionally, three other variants in genes which have never been associated with hEDS (*SMAD3, FANCA* and *PNPLA1*) were also found to segregate with the disease phenotype. Consequently, the putative impact of all five short-listed variants was individually assessed using gene databases, *in-silico* prediction tools and protein models (where available), as discussed in further detail below. The impact of these variants was assessed in all affected gene transcripts where available, as overlooked variants in non-canonical transcripts have been associated with missed diagnoses (Schoch, Tan, et al. 2020).

A missense variant was found in the recruited family which causes an Alanine to Threonine amino acid substitution (p.Ala173Thr) and is located in the 3rd exon of *TNXB*, on chromosome 6. *TNXB* encodes the ECM protein TN-X (OMIM: 600985), which, as annotated by the Genotype Ontology Resource (Ashburner, Ball, et al. 2000), is associated with regulation of cell adhesion (GO:0030155), cell migration (GO:0030334) and regulation of collagen organisation (GO:0030199). *Tnxb* knockout murine models show mild skin hyperextensibility, muscle weakness, reduced collagen fibrils and higher ECM turnover (Mao, Taylor, et al. 2002, Voermans, Verrijp, et al. 2011), a strikingly similar phenotype to that observed in hEDS affected individuals (Rombaut, Malfait, De Wandele, Mahieu, et al. 2012, Hermanns-Lê, Reginster, et al. 2012, Castori, Camerota, et al. 2010). *TNXB* rs61746206 (c.517G>A) was classified as having a benign effect by multiple *in-silico* prediction tools (refer to Table 3.8) and a moderate deleterious effect by VEP and SNPeff.

Lack of conservation was evidenced by low GERP++ (-8.83) and Phylo P (-1.17) scores, as well as the lack of nucleotide conservation in various orthologues. However, lack of sequence conservation is not mutually inclusive with lack of functionality, which is better denoted by selective constraints (Ponting 2017). The level of genetic constraint for the 1kb area around the variant locus (chr 6: 32097000-32098000 in build GHCh38) was calculated to have a missense Z-score of 5.86, while the observed/expected variation ratio is 0.59. This would imply that the

region in which the *TNXB* rs61746206 (c.517G>A) variant is found is more intolerant to variation than other areas (Chen, Francioli, et al. 2022).

The region surrounding the variant (chr6: 32065109-32065121) harbours a transcription factor binding site for Pancreas transcription factor 1 subunit alpha (Ptf1a), as annotated by the JASPAR transcription factor profile database (Castro-Mondragon, Riudavets-Puig, et al. 2022). This transcription factor is predicted to play a crucial role in the differentiation of pancreatic cells as well being involved in cerebellar development (Jin and Xiang 2019).The effect, if any, of this substitution on putative transcription factor binding motifs including that of Ptf1a was then analysed using HaploReg (version 4.2, Ward & Kellis, 2012), which also provides data on histone binding of the region in several cell lines. This variant does not appear to alter Ptf1a binding, yet it appears to alter the binding motifs of transcriptional repressor 58 (RP58) and TAL BHLH Transcription Factor 1, Erythroid Differentiation Factor (TAL1), as evidenced by the significant difference in the logarithm of the odds (LOD) scores between the reference and alternate alleles (**Figure 4.1**).

Regulatory motifs altered

Position Weight Matrix ID (Library from Kheradpour and Kellis, 2013)	Strand	Ref	Alt	Match on: Ref: TGAGGGTGGGGAAGAGGGAGGGATCTCAGCATCTGTGGGGTCTGAGCAGGTGGGCCCAC Alt: TGAGGGTGGGGAAGAGGGAGGGATCTCAGTATCTGTGGGGTCTGAGCAGGTGGGCCCAC
RP58	+	9.6	-2.4	NVMACATCTGKM
TAL1_known2	-	10.7	-1.2	NNNRMCATCTGKTVVB

Figure 4.1: HaploReg output for regulatory motif perturbations by *TNXB* rs61746206
The variant alters binding of transcription factors RP58 (forward strand) and TAL1 (reverse strand), having LOD discrepancies between the reference and alternate alleles of 12 and 11.9 respectively.

Transcription factor RP58 (also known as ZNF238) is associated with neuronal differentiation, migration and neuronal positioning (Heng, Qu, et al. 2015). In turn, TAL1 is essential for the differentiation of murine hematopoietic progenitors and cytoskeleton arrangement in erythroblasts (Porcher, Chagraoui, et al. 2017), with *Tal1*$^{-/-}$ mice models also showing disrupted vascular remodelling (Visvader, Fujiwara, et al. 1998). Haploreg also indicates that *TNXB*

rs61746206 (c.517G>A) appears to overlap histone enhancer marks in 7 tissue types, including foreskin fibroblasts, foetal abdominal muscle, foetal leg muscle, foetal stomach and adult skeletal muscle. It also overlaps a promoter histone region in human embryonic stem cell 1-derived mesenchymal stem cells, which are involved in differentiation of a multitude of cell lineages including osteogenic cells (Ding, Shyu, et al. 2011).

Using 3D *in-silico* model predictions, this variant does not appear to disrupt the protein structure and amino acid interactions in TN-X. The TN-X protein contains multiple epidermal growth factor (EGF)-like domains and Fibronectin Type III repeats interspaced by unstructured regions , as well as a Fibrinogen C-terminal (Zahn-Zabal, Michel, et al. 2020; neXtProt protein database version v2.52.0). The Alanine to Threonine amino acid substitution caused by *TNXB* rs61746206 (c.517G>A) is located in an unstructured region (amino acids 169-189), which is flanked on either side by EGF-like domains (**Figure 4.2**).

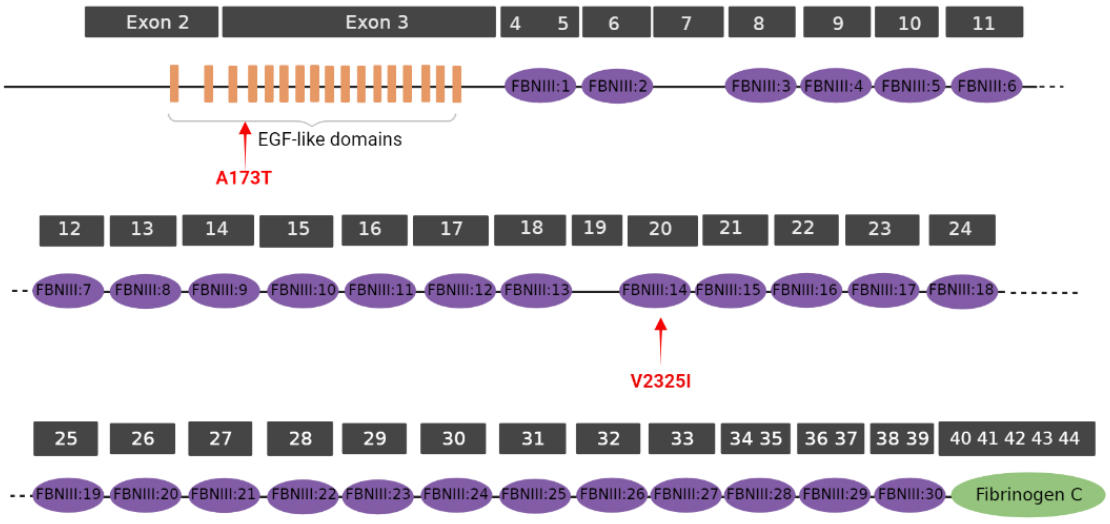


Figure 4.2: Position of filtered variants in the linear structure of Tenascin X (TN-X)
TN-X is composed of multiple EGF-like domains spanning exons 2-3, several Fibronectin type III repeats and a terminal Fibrinogen-C domain. Exon 1 is spliced from the final gene product and is thus not shown. Genomic positions of *TNXB* variants p.A173T (*TNXB* rs61746206 c.517G>A) and p.V2325I (*TNXB* rs140304758 c.6973G>A) are denoted by a red arrow . Gene sequence information based on canonical *TNXB* transcript ENST00000375247.2 which was obtained from Ensembl. Figure created using BioRender.com, with permission to use under BioRender.com licensing and usage rights.

In contrast to the Alanine, Threonine has side chains containing hydroxyl groups, by which it forms new hydrogen bonds (Scheiner, Kar, et al. 2002; Figure 3.9). These may contribute to increased stability of the altered protein upon chemical degradation, as calculated by CUPSAT.

A second shortlisted *TNXB* missense variant, *TNXB* rs140304758 (c.6973G>A), causes a Valine to Isoleucine substitution (p.Val2325Ile) in exon 20 of *TNXB*. This variant was described as benign by several variant prediction tools (Table 3.9), yet was classified as a VUS by ACMG/AMP guidelines. The reference amino acid Valine was shown to be conserved in primates, whilst in mammalian and non-mammalian species the reference nucleotide shows a slight measure of conservation as calculated by PhyloP (0.036) and GERP++(0.545). Accordingly, genomic constraint for the 1kb area surrounding *TNXB* rs140304758 (c.6973G>A), spanning the positions chr6: 32062000-32063000 using GHCh38, is calculated to have a missense Z-score of 4.26, while the observed/expected variation ratio is 0.73. In consideration of these scores, similar to the other *TNXB* variant detected in research participants, this SNP resides in an area that is relatively more intolerant to variation.

According to the JASPAR database (Castro-Mondragon, Riudavets-Puig, et al. 2022), the region surrounding the variant (chr6: 32030115-320130) is predicted to harbour a transcription factor binding site for Early B-cell Factors 2 (EBF2) and 3 (EBF3). EBF2 and EBF3 play a crucial role in neuron myelination (Moruzzo, Nobbio, et al. 2017) and neuronal cell differentiation (Green and Vetter 2011) while EBF2 is also associated with initiation of adipogenesis (Rajakumari, Wu, et al. 2013) and osteoclast differentiation (Kieslinger, Folberth, et al. 2005). However, according to annotation by HaploReg, no transcription factor binding motifs are perturbed by this SNV. Nevertheless, enhancer histone marks H3K4me1 and H3K27ac are described to contribute to the chromatin state at the variant locus in right ventricular tissue.

According to the neXtProt database (Zahn-Zabal, Michel, et al. 2020), the variant's position is located in the 14th fibronectin type III domain (**Figure 4.2**), which spans from amino acids 2305 to 2398. As no complete predicted model of the human Tenascin X protein is available on protein databases, damage to secondary structures, covalent bonding and charge difference could not be elucidated for *TNXB* rs140304758 (c.6973G>A). Available model structures of human Tenascin X only cover domains containing amino acid residues either upstream (Uniprot model accession numbers C9J7W4 and O95680) or downstream (Uniprot model accession number P22105) of the substituted amino acid. Thus, until a complete structural model is made available for this region, the effect of this variant on protein structure cannot be determined.

Both of these *TNXB* variants were classified as benign by most *in-silico* prediction tools and were also absent in one family member with hEDS (IV.2). Furthermore, *TNXB* rs61746206 (c.517G>A) does not appear to perturb structural and biochemical properties in the resultant protein. Consequently, the individual impact of *TNXB* rs140304758 (c.6973G>A) and *TNXB* rs61746206 (c.517G>A) in the molecular aetiology of hEDS is unlikely. According to the inheritance pattern inferred from the family pedigree (**Figure 3.5**), both SNVs are being maternally inherited on the same allele (in *cis*-), which is also concordant with the observed AD inheritance pattern. It may thus be possible that the cumulative effect of these variants may be sufficient to ablate the function of this allele. This hypothesis may be investigated further by the recruitment of hEDS participants with both SNVs and non-hEDS participants harbouring either one of these variants.

Variant *SMAD3* rs189286879 (c.206+32287G>A) was also shortlisted in the recruited family, causing a guanine to adenine nucleotide substitution in *SMAD3* (c.206+32287G>A) on chromosome 15. *SMAD3* (OMIM: 603109) encodes the SMAD family member 3 (SMAD3) protein, which belongs to the ubiquitously-expressed SMAD protein family (Moustakas,

Souchelnytskyi, et al. 2001). This gene is composed of 9 exons, which are translated to form MAD homology 1 (MH1) and MAD homology 2 (MH2) domains separated by a linker region separating the two domains (**Figure 4.3**). *SMAD3* rs189286879 (c.206+32287G>A) is located in intron 1 in the canonical transcript ENST00000327367.4, 32287 bases upstream of the nearest splice site, and in the 5' UTR of the short transcript ENST00000559092.1.

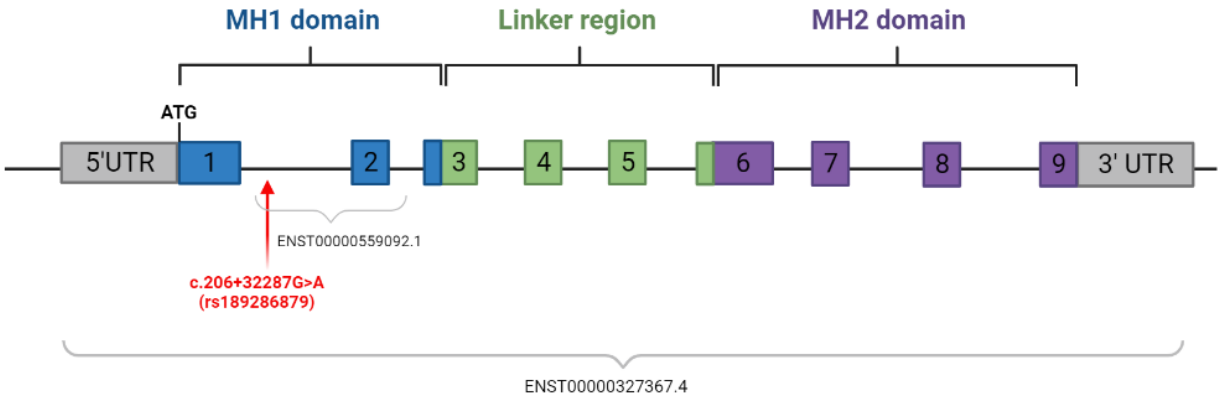


Figure 4.3 Linear structure of *SMAD3* and affected transcripts
A schematic representation of *SMAD3*, showing exons 1-9 which are organised in MH1 (blue boxes), linker region (green) and MH2 (purple) domains, flanked by UTR regions. This structure is based on canonical transcript ENST00000327367.4 which spans the entire coding sequence. The genomic position of *SMAD3* rs189286879 c.206+32287G>A (indicated with a red arrow) resides in intron 1 of the canonical transcript and in the 5' UTR region (not indicated) of shorter transcript ENST00000559092.1. Figure based on gene sequence annotation available on Ensembl and protein domain information on neXtProt (Entro no: NX_P84022). Assembled using BioRender.com, with permission to use under BioRender.com licensing and usage rights.

SMAD3 is associated with a multitude of protein interactions and biological processes, yet is most notably associated with the regulation (GO:0017015) and downstream mediation (GO:0007179) of the TGFβ receptor signalling cascade. It also enables sequence-specific DNA binding (GO:0043565) as well as binding with common mediator-SMAD proteins (GO:0070410) and collagen (GO:0005518). *Smad3*$^{-/-}$ murine models show impaired healing of the intestinal mucosa (Owen, Yuan, et al. 2008), kyphosis, osteoarthritis, aortic aneurysms, and premature death (van der Pluijm, van Vliet, et al. 2016).

The deleterious impact of SMAD3 deficiency as evidenced in mouse knockout models is attributed to the essential role of this protein in the TGFβ signalling pathway. The TGFβ

pathway is a signalling cascade (**Figure 4.4**) which is essential for cell proliferation, lineage differentiation and apoptosis (Liu 2010). Upon binding of a transforming growth factor (TGF) beta ligand to TGF Type I and Type II cell membrane receptors, TGF type II phosphorylates and activates TGF type I (Macias, Martin-Malpartida, et al. 2015). In turn, this activated heterotetrametric complex recruits intracellular SMAD2 and SMAD3 proteins by the mediation of Smad anchor for receptor activation (SARA) proteins (Liu 2010).

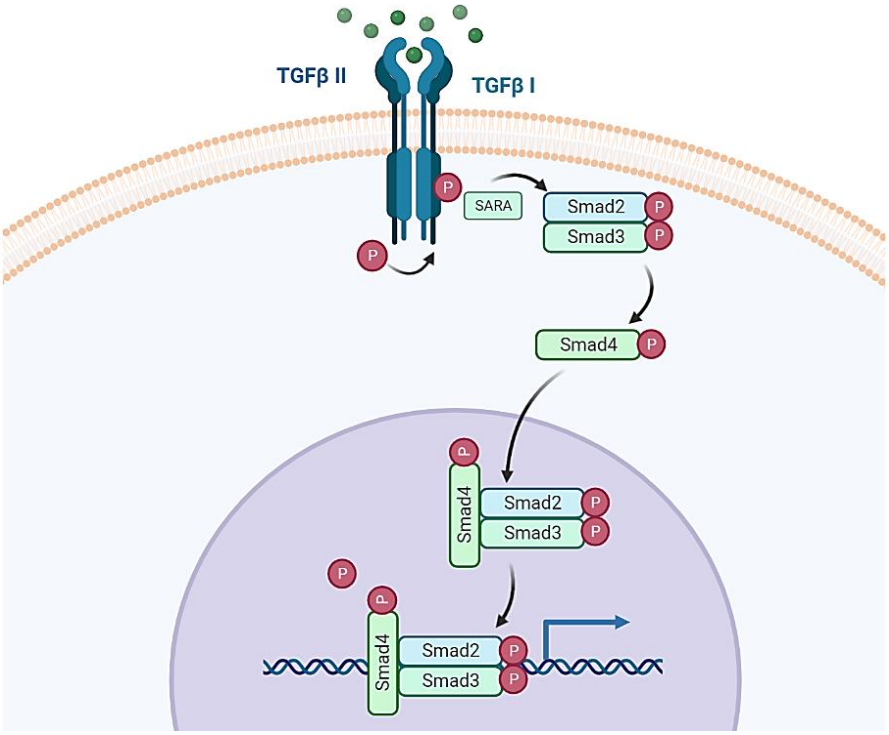


Figure 4.4: Canonical TGFβ signalling
A figure showing the recruitment and phosphorylation of *SMAD2* and *SMAD3* upon TGFβ receptor activation, leading to complex formation with *SMAD4*. This complex translocates to the nucleus, where it binds to DNA to activate transcriptional processes. Figure modified from pre-existing template using BioRender.com, with permission to use under BioRender.com licensing and usage rights.

SMAD2 and SMAD3 are then phosphorylated via their MH2 domain (Macias, Martin-Malpartida, et al. 2015). Once phosphorylated, SMAD2/3 form a complex with co-SMAD4, which translocates to the nucleus to regulate transcriptional activation of target genes (Owen, Yuan, et al. 2008) via their MH1 transcription factor- and DNA-binding domain (Martin-Malpartida, Batet, et al. 2017). TN-X is also hypothesised to be involved in the initiation and regulation of this pathway (Valcourt, Alcaraz, et al. 2015). During tissue ECM remodelling,

proteases cleave the Fibrinogen C terminal of the TN-X protein (**Figure 4.2**) which is then fused with a latent TGFβ complex. This enables the binding of this complex with transmembrane α11-β1 integrin receptors, inducing a conformational change in the TGFβ complex. The now-available TGFβ ligand binds to TGF Type I and Type II, thus initiating the SMAD3 cascade (Valcourt, Alcaraz, et al. 2015).

The observed discrepancy in conservation estimates for *SMAD3* rs189286879 (c.206+32287G>A) (**Figure 3.13**) is marginal and can be attributed to differences in the algorithms by which these scores are calculated. The level of genomic constraint in the 1kb region surrounding *SMAD3* rs189286879 (c.206+32287G>A), spanning positions 15:67098000-67099000 in build GHCh38, has an observed/expected ratio of 0.91 with a missense Z-score of 1.37. Thus, since the Z-score is below the threshold of 2.18, the region in which the variant resides is not constrained and is thus more tolerant to variation than other highly constrained areas of the genome.

The variant locus (Chr15: 67,390,400-67,391,801) overlaps regulatory feature ENSR00001872503. This promoter region does not appear to be active in any cell line, yet can be epigenetically activated in brain tissue (RoadmapEpigenomics dataset; Satterlee, Chadwick, et al. 2019). According to the JASPAR database, the 12 base-pair region upstream of the variant and the variant locus itself (chr 15: 67390973-67390985) is a binding site for the Plagl1 transcription protein. Plagl1 regulates gene expression by binding to proximal promoter regions of several genes involved in cell signalling (GO:0007267), cell adhesion (GO:0007155) and ECM organisation (GO:0030198) (Varrault, Dantec, et al. 2017). Putative alteration of transcription factor binding motifs including that of Plagl1 by *SMAD3* rs189286879 (c.206+32287G>A) was subsequently checked using HaploReg, which indicated that the SNV does not impede Plagl1 binding yet disrupts the binding motif of E2 Transcription Factor (E2F), thus decreasing binding activity. This transcription factor has been documented to regulate

gene expression in islet cells (Varshney, Scott, et al. 2017), which is not biologically significant to the pathophysiology of hEDS. Nevertheless, the variant locus is observed to reside in an area which overlaps multiple enhancer and promoter histone marks (H3K4me1, H3K4me3,H3K27ac and H3K9ac) in almost all cell lines featured on HaploReg. Additionally, this variant overlaps DNase hypersensitive sites in mesenchymal cells, epithelial and muscle tissue, enabling chromatin cleavage by DNase 1 thus increasing the accessibility of DNA to transcription machinery in these cell lines. This ChIP-Seq data implies that *SMAD3* rs189286879 (c.206+32287G>A) resides in an area which is essential for gene regulation and expression.

Another variant segregating with the disease phenotype in the recruited family is *FANCA*/*ZNF276* rs201316239 (c.*843C>T), a guanine to adenine substitution in the 3' UTR region of the ENST00000389301.3 canonical transcript of *FANCA* on chromosome 16. In this gene, the variant locus is a 3'UTR, which also overlaps an intronic region in non-canonical transcript ENST00000569901.1 of *ZNF276 (***Figure 4.5***)*.

FANCA (OMIM: 607139) encodes the Fanconi anaemia complementation group A protein (FANCA), which forms a protein complex with other complementation group members (Benitez, Liu, et al. 2018). Variants in *FANCA* are the most common cause of Fanconi anaemia (FA, OMIM: 227650), an AR disorder causing chromosomal abnormalities, premature ageing and bone marrow failure (Deakyne and Mazin 2011). *Fanca*$^{-/-}$ murine models displayed significant growth arrest (Vanuytsel, Cai, et al. 2014), which is attributed to the essential role of FANCA in DNA repair pathways (Deakyne and Mazin 2011). In turn, *ZNF276* (OMIM: 608460) codes for Zinc finger protein 276 (ZNF276, also known as ZFP276), a transcription regulator which regulates oligodendrocyte myelination (Aberle, Piefke, et al. 2022).

Accordingly, *ZNF276* knockout mice show delayed neuronal differentiation and myelination (Aberle, Piefke, et al. 2022).

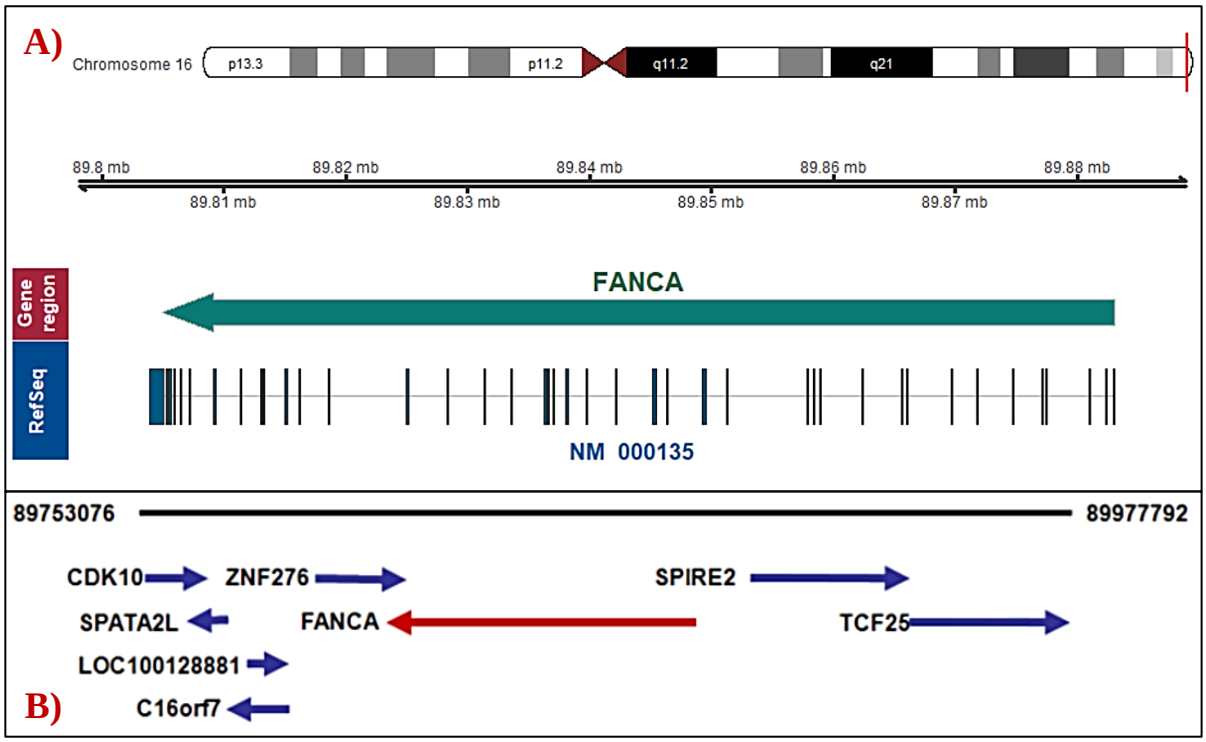


Figure 4.5: Genomic locations of *FANCA* and *ZNF276* within chromosome 16
Schematic representations of A) the position of *FANCA* in 16q24.3, direction of transcription (indicated by green arrow) as well as its gene structure, which is composed of 44 exons and B) a diagram showing the presence of several genes in the proximal area of *FANCA*, which overlaps with *ZNF276* at its 3' end. Image A obtained from the Japanese version of the Cancer Genome Atlas (Serizawa, Mizuguchi, et al. 2021), while image B was obtained from the Atlas of Genetics and Cytogenetics in Oncology and Haematology (Casado and Bueren 2012)

As *FANCA/ZNF276* rs201316239 (c.*843C>T) is located outside of the coding sequence in both *FANCA* and *ZNF276*, assessment of its deleteriousness using *in-silico* prediction tools was limited. However, submissions on ClinVar and the LOVD database both classify this SNV as a VUS. Functional conservation of the reference nucleotide was thus assessed using PhyloP and GERP++ , which indicated that the guanine reference nucleotide is moderately conserved and is thus under selective pressure. Nevertheless, the genomic constraint in the 1kb area surrounding the variant (chr16: 89737000-89738000 in build GHCh38) has a missense Z-score of -2.24, with an observed/expected ratio of 0.16, implying that the region is not under strong selection compared to other areas of the genome.

The variant is located in a region which overlaps the regulatory feature ENSR00001887189 (chr16: 89802002-89804599), a promoter flanking region. ENSR00001887189 is active in several cell lines related to immune function, colon carcinoma cell lines as well as brain tissue (RoadmapEpigenomics dataset; Satterlee, Chadwick, et al. 2019). Furthermore, the variant locus at chr16:89804166 forms part of transcription factor binding areas for transcription factors EBF3 (as *TNXB* rs140304758, discussed above), Zinc finger protein 460 (ZNF460) and Zinc finger protein (ZNF135) according to the JASPAR database (Castro-Mondragon, Riudavets-Puig, et al. 2022). ZNF460 and ZNF135 are transcriptional regulators belonging to the Zinc finger protein family, with the latter also being associated with regulation of transcription, cell morphology and organisation of the cytoskeleton (neXtProt entry NX_P52742; Zahn-Zabal, Michel, et al. 2020). However, upon enquiry using HaploReg, the SNV does not appear to alter the binding of these transcription factors. Instead, the variant alters the binding motifs of CTCF_disc7, Zic_4 and TR4_disc1 (**Figure 4.6**).

Transcription factor binding motifs altered by rs201316239

Alteration of transcription factor (TF) binding may explain the regulatory effect of a variant.

Position Weight Matrix	Delta	Position	Strand	Reference Score	Alternate Score
CTCF_disc7	-4.37	89,804,166	-	11.12	6.76
TR4_disc1	11.97	89,804,166	+	0.34	12.30
Zic_4	-4.50	89,804,166	+	14.74	10.24

Figure 4.6: List of transcription factor binding motifs altered by the FANCA/ZNF276 variant
Figure showing the results obtained by Haploreg indicating that the variant alters the binding sites of transcription factors CTCF_disc7 (reverse strand) as well as TR4_disc1 and Zic_4 (forward strand), having LOD discrepancies of -4.37, +11.97 and -4.50 respectively.

CTCF_disc7 appears to be involved in the regulation of gene expression in islet cells (Varshney, Scott, et al. 2017), whilst Zic_4 is an inhibitor of tumour proliferation (Chen, Tang, et al. 2020). Unfortunately, the function of TR4_disc1, the transcription factor with the highest LOD discrepancy between the reference and alternate alleles, is not annotated yet. The area in which *FANCA*/*ZNF276* rs201316239 (c.*843C>T) resides is annotated by Haploreg to overlap enhancer histone marks in multiple cell lines related to haematological and immune function.

In consideration of these regulatory effects and lack of correlation with connective tissue metabolism, this variant is unlikely to be associated with genetic mechanisms causing hEDS.

Lastly, another shortlisted missense variant results in a Glutamic acid to Lysine substitution, *PNPLA1* rs45524833 (c.745G>A), in various transcripts of *PNPLA1* (OMIM: 612121) in chromosome 6. In turn, *PNPLA1* encodes the Omega-hydroxyceramide transacylase protein, a catalytic enzyme involved in lipid haemostasis (GO:0055088), triglyceride lipase activity (GO:0004806), acyltransferase activity (GO:0016747) and regulation of keratinocyte differentiation (GO:0045616) (Ashburner, Ball, et al. 2000). The crucial role of Omega-hydroxyceramide transacylase in cutaneous integrity is further evidenced in *Pnpla1*$^{-/-}$ models, whereby these knockout mice displayed severe epidermal permeability which lead to premature death (Miyamoto, Itoh, et al. 2020, Pichery, Huchenq, et al. 2017). Knockout of *Pnpla1* is also documented to be lethal in data obtained by the International Mouse Phenotyping Consortium (IMPC; Groza, Gomez, et al. 2023).

The Omega-hydroxyceramide transacylase protein is composed of a patatin-like phospholipase domain spanning residues 16-185 (**Figure 4.7**), with occasional short-spanning sequence motifs, which are interspaced by large areas of disordered regions and a hydrophobic region (neXtProt v2.52.0; Zahn-Zabal, Michel, et al. 2020). The patatin-like phospholipase domain constitutes the active site of Omega-hydroxyceramide transacylase, by which it exerts its acyl hydrolytic activity on lipids (Rydel, Williams, et al. 2003, Kienesberger, Oberer, et al. 2009).

Both Polyphen and GERP++ conservation estimates (2.51 and 3.97 respectively) indicate that the reference nucleotide is well-conserved across species. Additionally, the reference amino acid is also observed to be under positive selective pressure as it is well conserved in mammalian species (**Figure 3.17**). However, in the 1kb region surrounding the variant (chr6: 36295000-36296000), the missense Z-score is calculated to be 0.62, with an observed/expected

ratio of 0.95. As the observed/expected ratio is close to 1 and the calculated Z-score marginally exceeds 0, the level of genetic constraint for this region in *PNPLA1* is low and therefore, the region is prone to variation.

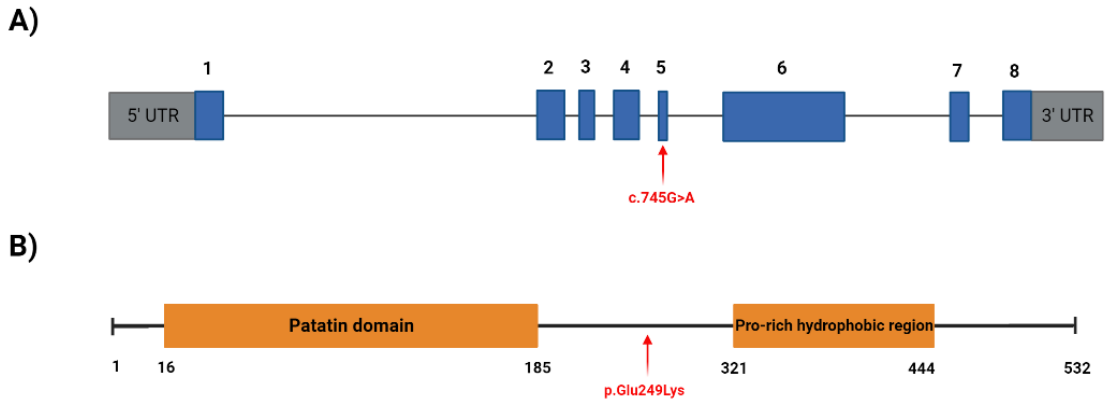


Figure 4.7: Position of *PNPLA1* rs45524833 in *PNPLA1* and Omega-hydroxyceramide transacylase
A schematic diagram of (A) the PNPLA1 gene, showing the variant locus of *PNPLA1* rs45524833 (c.745G>A) in exon 5 and (B) position of the same variant within Omega-hydroxyceramide transacylase. The protein is composed of a Patatin-like domain and a proline-rich hydrophobic region, interspaced by disordered residues. Figure based on gene sequence information of transcript ENST00000394571.2 available on Ensembl and Omega-hydroxyceramide transacylase on neXtProt (entry: NX_Q8N8W4). Assembled using BioRender.com, with permission to use under BioRender.com licensing and usage rights.

This variant was classified as a VUS by the ACMG/AMP guidelines based on the functionally critical location of the variant, which is contrasted with the benign prediction from other sources. Indeed, *PNPLA1* rs45524833 (c.745G>A) was predicted to have a likely benign effect by multiple *in-silico* prediction tools, with the exception of Polyphen-2, which denoted a moderate effect in the non-canonical transcript ENST00000457797.1 and Mutpred2 (**Figure 3.18**). Mutpred2 classified this variant as pathogenic due to a significant probability that *PNPLA1* rs45524833 (c.745G>A) alters metal ion chelation and post-translational sulfation of Tyrosine at residue 245. However, it is worth noting that according to the developers of this prediction model, a probability score of 0.68 carries a 10% false positive rate, while that of 0.8 yields a false positive rate of 5% (MutPred2 n.d.). This would imply that the MutPred2 prediction for this variant carries a false positive rate of >10%, which must be taken into consideration. This SNV is also predicted to affect binding sites associated with Tyrosine

kinase phosphorylation signalling (motifs PS00007 and ELME000081) as well as a binding site associated with mitogen-activated protein kinase (MAPK) signalling (ELME000233;Kumar, Michael, et al. 2022, Paysan-Lafosse, Blum, et al. 2023).

According to the JASPAR database, no transcription factor binds to the variant locus (chr6: 36263171). However, in Haploreg, *PNPLA1* rs45524833 (c.745G>A) appears to alter the binding motifs of transcription factors Ets_known9, Nr2f2 and Pou1f1_2, having LOD score differences between the reference and alternate alleles of 9.85,11.97 and 1.87 respectively (**Figure 4.8**).

Transcription factor binding motifs altered by rs45524833

Alteration of transcription factor (TF) binding may explain the regulatory effect of a variant.

Position Weight Matrix	Delta	Position	Strand	Reference Score	Alternate Score
Ets_known9	-9.85	36,263,171	+	9.13	-0.72
Nr2f2	11.97	36,263,171	+	-4.69	7.28
Pou1f1_2	1.87	36,263,171	+	10.73	12.60

Figure 4.8: ***PNPLA1* rs45524833 alters transcription factor binding motifs of Ets_known9, Nr2f2 and Pou1f1_2 on the forward strand**
Output obtained from Haploreg showing LOD scores for these transcription factors of the reference and alternate alleles.

The transcription factor Ets_known9, also known as GA-binding protein (GABP) is an important transcriptional regulator of several genes associated with oxidative phosphorylation, control of the cell cycle and cell metabolism (Rosmarin, Resendes, et al. 2004, Grubert, Judith, et al. 2015). In turn, Nuclear receptor subfamily 2, group F, member 2 (Nr2f2) is essential for the structural development of cardiac and valvular tissue (Wang, Abhinav, et al. 2019) while Pou1f1 regulates pituitary differentiation and transcription (Brue and Camper 2021). The variant locus is also annotated by Haploreg to reside within an H3K4me1 enhancer peak in peripheral blood neutrophils and monocytic cells. Nevertheless, these cell types as well as the annotated function of Ets_known9, Nr2f2 and Pou1f1_2 are not considered relevant to the pathophysiology of hEDS.

The amino acid substitution in the canonical transcript, p.Glu249Lys, does not appear to reside within a specified protein domain. Yet, it is shown to form part of an α-helix motif on the 3D model of the protein (**Figure 3.20**). This SNV was predicted to disrupt a hydrogen bond between the reference amino acid and a Tyrosine residue in a proximal α-helix due to the lack of a hydrogen bond acceptor in the substituted amino acid. Furthermore, the difference in structure of glutamic acid and lysine leads to expansion of the cavity volume in the region, which has been shown to decrease protein stability (Jenkins, Fossat, et al. 2018, Xue, Wakamoto, et al. 2019). This cavity expansion might explain the destabilising effect of the variant on chemical denaturation calculated by CUPSAT (refer to section 3.4.5).

Individually, each of the shortlisted variants analysed in this dissertation were predicted to have a benign or non-deleterious effect, and are thus unlikely to be the causative factor driving the emergence of the hEDS phenotype. The reported genetic heterogeneity in hEDS individuals (Levy 2018, Tinkle, Castori, et al. 2017) complicates the elucidation of genetic variants associated with the disorder. Individuals within the same family might have different gene-environmental interactions, impeding the efficacy of filtering pipelines by the possible exclusion of variants which might be contributing to the hEDS phenotype in one individual, but are absent in another (Tinkle, Castori, et al. 2017). In turn, genetic factors might also be further influenced by the effect of hormones and gender on pain thresholds, occupational strain/trauma and other lifestyle factors (Tinkle, Castori, et al. 2017, Martin 2019). These intrinsic characteristics of the disorder might indeed explain the significant differences in hEDS phenotypes as well as the persistent lack of correlation between gene variants and disease aetiology.

In the recruited family, 4 family members (III.3, III.5, IV.2 and IV.8) also harbour the missense variant c.1280T>G (p.Val427Gly) in the endoglin (*ENG*) gene, which is associated with Hereditary Haemorrhagic Telangiectasia (HHT). HHT is an AD condition characterised

by spontaneous and recurrent epistaxes, large visceral arteriovenous malformations (AVMs) and telangiectasias (McDonald and Stevenson 2021). HHT is reportedly caused by variants in *ENG*, activin receptor-like kinase 1 (*ALK1*) and *SMAD4* (McDonald and Stevenson 2021, Berg, Gallione, et al. 1997). *ENG* variants cause haploinsufficiency of the endoglin transmembrane protein, which mediates non-canonical TGFβ signal transduction (Rossi, Lopez-Novoa, et al. 2015, Wheeler, Ikonomidis, et al. 2021). This reduction of transmembrane endoglin receptors impedes the infiltration of leucocytes upon inflammatory stimuli, resulting in abnormal vascular repair and reduction of the capillary network (Rossi, Lopez-Novoa, et al. 2015).

In addition to endoglin, non-canonical TGFβ signal transduction can also be initiated by activin receptor-like kinase 1 (ALK-1*;* Bertolino, Deckers, et al. 2005), which enhances matrix metalloproteinase 13 levels (MMP13; Wheeler, Ikonomidis, et al. 2021). MMP13 is a major cleaving agent of collagen type II, IV and IV as well as bone ECM proteoglycans (Wang, Sampson, et al. 2013), resulting in the altered collagen network and ECM degradation (Wheeler, Ikonomidis, et al. 2021). It might thus be possible that, in the presence of endoglin haploinsufficiency caused by this variant in III.3, III.5, IV.2 and IV.8, non-canonical TGFβ signalling is compensated by ALK-1, resulting in collagen cleavage and ECM degradation. This proposed mechanism does not explain fully the emergence of the hEDS phenotype in these participants, as hEDS participant IV.1 does not have the *ENG* variant. Furthermore, this alternative mechanism is unlikely to cause the cutaneous features observed in female research participants (Refer to Appendix J), as IV.9 also doesn't have the *ENG* familial variant. However, it might be possible that HHT may contribute to or aggravate the clinical manifestations of hEDS in this family. The comparatively diminished severity of the hEDS phenotype in IV.1 might not be entirely due to gender-dependant differences but also due to higher maintenance of ECM integrity compared to his familial counterparts.

Limitations of the study

There are several factors worth mentioning which limited the extent, interpretation and applicability of the results obtained. Notably, this study recruited only one Maltese family and a separate non-related affected individual, which limited the assessment of genetic heterogeneity between Maltese families and between individuals with hEDS. The lack of other males in the recruited group of participants limited, in part, the conclusions pertaining to gender-dependent differences in hEDS phenotypes, whilst the recruitment of additional non-affected individuals would have been beneficial to filter further non-causative variants. The number of family studies investigating the genetics of hEDS is scant, which limited the comparability of the findings obtained in this study. Furthermore, due to the lack of previous research on the disorder in Malta, the results obtained also could not be compared with other local studies.

As a significant phenotypic overlap between hEDS and other EDS subtypes is observed, the specificity of the newly adopted diagnostic criteria used to diagnose participants (International Consortium on EDS & Related Disorders 2017) is also a potential limitation of this study. Significant diagnostic weight on clinical characteristics has been proposed to result in subjective interpretations (Martin 2019). The authors themselves have acknowledged shortcomings in the structure of the diagnostic checklist and identified the need to base future updates of the criteria on evidence-based practice, which is currently in progress (The International Consortium on EDS & Related Disorders 2023).

The lack of serum TN-X measurement in recruited participants with variants in *TNXB* also hindered the feasibility of correlating these variants with protein levels. In the literature, serum TN-X was measured by LC/MS/MS (Yamada, Watanabe, et al. 2019), Western blot analysis from cultured skin fibroblasts (Merke, Chen, et al. 2013) and by immunohistochemistry using

primary rabbit polyclonal antihuman TN-X antibody on skin biopsy sections (Merke, Chen, et al. 2013). These analyses were not performed due to time constraints, feasibility of obtaining consent for collection of tissue biopsies in addition to logistical difficulties pertaining to the availability of local LC/MS/MS analysers.

The use of the GRCh37 reference genome to align sequencing data instead of the newer GRCh38 is also a limiting factor of this study. As the sequencing data previously analysed by Centogene was aligned to GRCh37, the sequencing data of other participants in this study was also aligned to the same reference genome to avoid any discrepancies stemming from several differences between builds. The GRCh38 build improved the annotation of the exome and enhanced read alignment by arranging misassembled sequences and filling numerous gaps between contiguous sequences found in GRCh37, as well as including centromere regions and adding alternate loci in several genomic regions (Guo, Dai, et al. 2017). Thus, GRCh38 is considered as a more comprehensive representation of the human genome than GRCh37, which improves the accuracy of genomic analysis using this build (Guo, Dai, et al. 2017).

Additionally, while the function of all the genes included in the gene panel approach is well known, five variants (which were filtered by a filtering pipeline covering extended coding regions) are located in three genes (*C5orf58, C6orf229* and *C9orf131*) which are not yet annotated. According to data obtained from the Genotype-Tissue Expression project (GTEx; Carithers, Ardlie, et al. 2015), *C5orf58* and *C6orf229* are predominantly expressed in the testes, spinal cord and lungs, while *C9orf131* are mostly expressed in the testes, retina and cerebellum. No mice or zebrafish knockout models are available for these three genes on IMPC (Groza, Gomez, et al. 2023), Mouse Genome Informatics (MGI;Blake, Baldarelli, et al. 2021) and the Zebrafish Information Network (ZFIN; Bradford, Van Slyke, et al. 2022). Thus, the functional role of these genes and their relevance to hEDS could not be assessed. Additionally, several

variants were found in genes which, although annotated, are not associated with any disorder or phenotype yet (refer to Appendix L).

This lack of correlation between shortlisted variants and hEDS may also be attributed to other variables which require consideration. Most *in-silico* prediction tools ,such as the ones utilised in this research study, assess the impact of a variant by the degree of imparted loss-of-function (LOF). However, gain-of-function (GOF) alterations may also introduce novel functions in protein isoforms (Li, Zhang, et al. 2019). The functional extent of GOF variants is not yet fully understood (Li, Zhang, et al. 2019) as, in contrast to LOF alterations which are usually located in structured domains, GOF variants are found in intrinsically disordered regions (Li and Babu 2018, Vaughan, Singh, et al. 2017, Li, Zhang, et al. 2019). Furthermore, GOF variants are comparatively less-common and their effect is less evident, resulting in the classification of GOF variants as benign by *in-silico* prediction tools due to limited data sets and lower accuracy of prediction (Flanagan, Patch, et al. 2010).

Gene expression varies significantly between cell lines due to alternative mechanisms such as epigenetic regulation (Stephens, Miaskowski, et al. 2013). Transcriptional processes are epigenetically regulated by methylation/demethylation of DNA and histone proteins, histone de/acetylation as well as remodelling of chromatin structure, all of which alter the availability of the DNA strand to transcriptional machinery (Stephens, Miaskowski, et al. 2013, Chen, Li, et al. 2017). In turn, post-transcriptional epigenetic regulation is mediated by the action of non-coding RNA fragments, which mediate mRNA transcript stability (Chen, Li, et al. 2017, Stephens, Miaskowski, et al. 2013). Thus, as this study did not assess the role of epigenetic regulation or possible GOF in short-listed variants, the putative role of these mechanisms in hEDS might have been overlooked.

Furthermore, since deep intronic variants possibly located in upstream promoter sequences and other cis-regulatory elements (CREs) were excluded from bioinformatic analysis, this study might have also missed putative causative variants affecting regulation of gene expression in hEDS. Non-coding regions harbour multiple regulatory elements within topologically associated domains (TADs) which are wide-spanning, high-interaction chromatin regions flanking the coding sequence of the gene as well as its regulatory elements (Krude, Mundlos, et al. 2021). TADs are confined by distinct boundary/insulator sequences on each end which are then gathered during transcription to form a DNA loop, thus restricting the activity and contact of CREs within their target genes (Lupiáñez, Kraft, et al. 2015). Nucleotide alteration in TADs can affect transcription by alteration of promoter/ enhancer sequences which hinder transcription factor binding, 5' and 3' UTR disruptions and gain/loss of splice site regions (Almeida, Tavares, et al. 2017, Krude, Mundlos, et al. 2021). If disrupted, modifications in TAD structure can have crucial ramifications on gene expression and disease (Ellingford, Ahn, et al. 2022, Vaz-Drago, Custódio, et al. 2017). Indeed, variants in deep intronic regions have been associated with rare disorders (Almeida, Tavares, et al. 2017, Vaz-Drago, Custódio, et al. 2017, Bax, Sangermano, et al. 2015) as well as affecting degree of penetrance and phenotypic differences (Perez-Becerril, Evans, et al. 2021), which might explain the gender-dependant and protean clinical manifestations of hEDS. In the absence of a clear correlation between variants in coding sequences and hEDS in this study, readily available WGS data of these research participants can be analysed in future studies to determine the presence, if any, of variants in TADs and their impact on the hEDS phenotype.

Recommendations for future research

Optimally, future molecular research on hEDS should include the recruitment of multiple hEDS families as well as isolated affected and non-affected participants to better assess the inter-individual, intra- and inter-familial variability of discovered variants. WGS data from

participants should be used in multiple filtering pipelines, including 1. a filtering strategy for each family, 2. a combination of multiple families, and 3. a separate filtering pipeline for individual participants and unaffected, unrelated controls. In the absence of associated deleterious variants in the coding exome, short-listed variants in non-coding areas should be thoroughly assessed to determine the impact, or lack thereof, on regulatory elements. The need for further assessment of epigenetic changes affecting gene expression in hEDS has already been suggested (Ritelli, Chiarelli, et al. 2022), which can be further investigated by transcriptomic analysis of skin and muscle tissue in future work.

Molecular studies should also be supplemented with biochemical and radiological analyses, which may aid researchers to determine the effect of variants on clinical phenotypes. Specifically, the determination of serum TN-X would be beneficial to test the impact of *TNXB* variants on protein levels. Studies which include immunofluorescent and transcriptome analysis of cultured skin fibroblasts (Chiarelli, Carini, et al. 2016, Zoppi, Chiarelli, et al. 2018) as well as transmission electron microscopy of skin tissue (Hermanns-Lê, Reginster, et al. 2012) are also ideal to detect perturbed molecular pathways. Of particular interest is the expression and role of TGFβ signalling, *SMAD3* expression and metalloproteinase levels in hEDS, which can be analysed in transcriptome analysis from tissue samples. The functional, physiological effect of short-listed variants in this study as well as any variants identified in future molecular work on the disorder can also be assessed using *in-vivo* techniques such as animal models.

The prevalence, clinical manifestations, and genetic aetiology of hEDS in Maltese individuals require further elucidation. This is the first research project ever to investigate hEDS in Malta, possibly commencing the establishment of a Maltese hEDS cohort to be recruited in future research on the disorder. As previously stated, hEDS patients are often misdiagnosed or dismissed by their clinicians (Castori 2016b) and thus, this study might highlight further the

clinical manifestations of hEDS in local participants, which can then be applied in clinical practice. Future studies on the disorder would also enable the comparison of research findings between cohorts, thus resulting in a more comprehensive assessment of hEDS in the local population. Data from these studies can subsequently be included in larger international consortia such as the Hypermobile Ehlers Danlos Genetic Evaluation (HEDGE) network. The lack of reproducibility between different molecular studies on hEDS can be counteracted by the establishment of an international database (Tinkle, Castori, et al. 2017) containing phenotypic data from participants attending various medical centres, annotated using standardised terminology such as the Human Phenotype Ontology (HPO) system. Identified phenotypic patterns can be then integrated with genomic and transcriptomic data to detect genotype-phenotype patterns using developed statistical algorithms (Scicluna, Formosa, et al. 2022; Appendix M).

Conclusion

Several notable advances have been made in the diagnostic delineation and molecular understanding of hEDS, which have highlighted several valid candidate mechanisms possibly associated with the disorder. Nevertheless, as is evident from this research, uncovering the genetic aetiology of hEDS remains a complex process which is further aggravated by a multitude of clinical phenotypes. This study has described, for the first time, the clinical manifestations of hEDS in local individuals and tested the hypothesised association of this condition with bone-related parameters. Most importantly, this research study also utilised a variety of *in silico* tools to determine the relevance of shortlisted variants in *TNXB,SMAD3,FANCA* and *PNPLA1* identified through HTS. However, the individual impact and relevance of these variants with the molecular mechanism of this disorder is unlikely or still unclear. The utilisation of HTS technology, combined with the newly defined diagnostic criteria and novel genetic techniques, present ample opportunities for researchers to finally identify the genetic aetiology of this condition. Future local studies may thus augment and expand the work conducted in this research study, with the ultimate objective of establishing a potential molecular diagnosis and improving the clinical management of hEDS patients.

References

Aberle,T., Piefke,S., et al. ,2022. Transcription factor Zfp276 drives oligodendroglial differentiation and myelination by switching off the progenitor cell program. *Nucleic Acids Research* 50(4): p.1951–1968.

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

Afrin,L. ,2016. *Never Bet Against Occam: Mast Cell Activation Disease and the Modern Epidemics of Chronic Illness and Medical Complexity*. SistersMedia, LLC.

Alberts,B., Johnson,A., et al. ,2002. Fibroblasts and Their Transformations: The Connective-Tissue Cell Family. *Molecular Biology of the Cell. 4th edition*. Available at: https://www.ncbi.nlm.nih.gov/books/NBK26889/ [Accessed April 27, 2020].

Almeida,R.M. de, Tavares,J., et al. ,2017. Whole gene sequencing identifies deep-intronic variants with potential functional impact in patients with hypertrophic cardiomyopathy. *PLOS ONE* 12(8): p.e0182946.

Ambardar,S., Gupta,R., Trakroo,D., Lal,R., and Vakhlu,J. ,2016. High Throughput Sequencing: An Overview of Sequencing Chemistry. *Indian Journal of Microbiology* 56(4): p.394–404.

Ancillao,A., Galli,M., et al. ,2012. Temporomandibular joint mobility in adult females with Ehlers-Danlos syndrome, hypermobility type (also known as joint hypermobility syndrome). *Journal of Cranio-Maxillary Diseases* 1(2). Available at: https://www.researchgate.net/publication/262403659_Temporomandibular_joint_mobility_in_adult_females_with_Ehlers-Danlos_syndrome_hypermobility_type_also_known_as_joint_hypermobility_syndrome [Accessed May 15, 2020].

Andrews,S. ,2010. FastQC: a quality control tool for high throughput sequence data.

Annapureddy,N., Block,J.A., and Khandelwal,S. ,2014. Decreased Bone Mineral Density in Patients with Ehler-Danlos Syndrome. *Arthritis & Rheumatology* 66(10): p.S961–S962.

Arendt-Nielsen,L., Kaalund,S., Bjerring,P., and Høgsaa,B. ,1990. Insufficient effect of local analgesics in Ehlers Danlos type III patients (connective tissue disorder). *Acta Anaesthesiologica Scandinavica* 34(5): p.358–361.

Arendt-Nielsen,L., Kaalund,S., Høgsaa,B., Bjerring,P., and Grevy,C. ,1991. The response to local anaesthetics (EMLA-cream) as a clinical test to diagnose between hypermobility and Ehlers Danlos type III syndrome. *Scandinavian Journal of Rheumatology* 20(3): p.190–195.

Ashburner,M., Ball,C.A., et al. ,2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 25(1): p.25–29.

Asher,S.B., Chen,R., and Kallish,S. ,2018. Mitral valve prolapse and aortic root dilation in adults with hypermobile Ehlers-Danlos syndrome and related disorders. *American Journal of Medical Genetics. Part A* 176(9): p.1838–1844.

Atzinger,C.L., Meyer,R.A., Khoury,P.R., Gao,Z., and Tinkle,B.T. ,2011. Cross-sectional and longitudinal assessment of aortic root dilation and valvular anomalies in hypermobile and classic Ehlers-Danlos syndrome. *The Journal of Pediatrics* 158(5): p.826-830.e1.

Baeza-Velasco,C., Bourdon,C., et al. ,2018. Low- and high-anxious hypermobile Ehlers-Danlos syndrome patients: comparison of psychosocial and health variables. *Rheumatology International* 38(5): p.871–878.

Bagalad,B.S., Mohan Kumar,K.P., and Puneeth,H.K. ,2017. Myofibroblasts: Master of disguise. *Journal of Oral and Maxillofacial Pathology : JOMFP* 21(3): p.462–463.

Bascom,R., Schubart,J.R., et al. ,2019. Heritable disorders of connective tissue: Description of a data repository and initial cohort characterization. *American Journal of Medical Genetics. Part A* 179(4): p.552–560.

Basel,D., and Steiner,R.D. ,2009. Osteogenesis imperfecta: Recent findings shed new light on this once well-understood condition. *Genetics in Medicine* 11(6): p.375–385.

Bax,N.M., Sangermano,R., et al. ,2015. Heterozygous Deep-Intronic Variants and Deletions in *ABCA4* in Persons with Retinal Dystrophies and One Exonic *ABCA4* Variant. *Human Mutation* 36(1): p.43–47.

Beighton,P., de Paepe,A., et al. ,1988. International Nosology of Heritable Disorders of Connective Tissue, Berlin, 1986. *American Journal of Medical Genetics* 29(3): p.581–594.

Beighton,P., De Paepe,A., Steinmann,B., Tsipouras,P., and Wenstrup,R.J. ,1998. Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos Support Group (UK). *American Journal of Medical Genetics* 77(1): p.31–37.

Beighton,P., Solomon,L., and Soskolne,C.L. ,1973. Articular mobility in an African population. *Annals of the Rheumatic Diseases* 32(5): p.413–418.

Bendik,E.M., Tinkle,B.T., et al. ,2011. Joint hypermobility syndrome: A common clinical disorder associated with migraine in women. *Cephalalgia: An International Journal of Headache* 31(5): p.603–613.

Bénistan,K., and Martinez,V. ,2019. Pain in hypermobile Ehlers-Danlos syndrome: New insights using new criteria. *American Journal of Medical Genetics. Part A* 179(7): p.1226–1234.

Benitez,A., Liu,W., et al. ,2018. FANCA Promotes DNA Double-Strand Break Repair by Catalyzing Single-Strand Annealing and Strand Exchange. *Molecular Cell* 71(4): p.621-628.e4.

Berg,J.N., Gallione,C.J., et al. ,1997. The activin receptor-like kinase 1 gene: genomic structure and mutations in hereditary hemorrhagic telangiectasia type 2. *American Journal of Human Genetics* 61(1): p.60–67.

Berger,C., Langsetmo,L., et al. ,2008. Change in bone mineral density as a function of age in women and men and association with the use of antiresorptive agents. *CMAJ : Canadian Medical Association Journal* 178(13): p.1660–1668.

Berglund,B., Pettersson,C., Pigg,M., and Kristiansson,P. ,2015. Self-reported quality of life, anxiety and depression in individuals with Ehlers-Danlos syndrome (EDS): a questionnaire study. *BMC musculoskeletal disorders* 16: p.89.

Bertolino,P., Deckers,M., Lebrin,F., and ten Dijke,P. ,2005. Transforming growth factor-beta signal transduction in angiogenesis and vascular disorders. *Chest* 128(6 Suppl): p.585S-590S.

Beyens,A., Albuisson,J., et al. ,2018. Arterial tortuosity syndrome: 40 new families and literature review. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 20(10): p.1236–1245.

Bioneer. AccuPrep® PCR/Gel Purification Kit. Available at: https://us.bioneer.com/pagecat1/accuprep/pcr-gel-purification-kit?tab=overview [Accessed January 16, 2023].

Blake,J.A., Baldarelli,R., et al. ,2021. Mouse Genome Database (MGD): Knowledgebase for mouse-human comparative biology. *Nucleic Acids Research* 49(D1): p.D981–D987.

Bolger,A.M., Lohse,M., and Usadel,B. ,2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30(15): p.2114–2120.

Botstein,D., and Risch,N. ,2003. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature Genetics* 33(S3): p.228–237.

Bradford,Y.M., Van Slyke,C.E., et al. ,2022. Zebrafish information network, the knowledgebase for Danio rerio research. *Genetics* 220(4): p.iyac016.

Brockway,L. ,2016. Gastrointestinal problems in hypermobile Ehlers-Danlos syndrome and hypermobility spectrum disorders. Available at: https://www.ehlers-danlos.org/information/gastrointestinal-problems-in-hypermobile-ehlers-danlos-syndrome-and-hypermobility-spectrum-disorders/ [Accessed June 6, 2019].

Brue,T., and Camper,S.A. ,2021. Novel mechanism of pituitary hormone deficiency: genetic variants shift splicing to produce a dominant negative transcription factor isoform. *European Journal of Endocrinology* 185(6): p.C19–C25.

Bulbena,A., Baeza-Velasco,C., et al. ,2017. Psychiatric and psychological aspects in the Ehlers-Danlos syndromes. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175(1): p.237–245.

Burrows,M., and Wheeler,D.J. ,1994. A Block-sorting Lossless Data Compression Algorithm.

Byers,P.H., and Murray,M.L. ,2012. Heritable Collagen Disorders: The Paradigm of the Ehlers—Danlos Syndrome. *Journal of Investigative Dermatology* 132: p.E6–E11.

Byers,P.H., and Murray,M.L. ,2014. Ehlers–Danlos syndrome: A showcase of conditions that lead to understanding matrix biology. *Matrix Biology* 33: p.10–15.

Campbell,I.D., and Humphries,M.J. ,2011. Integrin structure, activation, and interactions. *Cold Spring Harbor Perspectives in Biology* 3(3).

Carbone,L., Tylavsky,F.A., et al. ,2000. Bone Density in Ehlers-Danlos Syndrome. *Osteoporosis International; London* 11(5): p.388–92.

Carithers,L.J., Ardlie,K., et al. ,2015. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking* 13(5): p.311–319.

Casanova,E.L., Sharp,J.L., Edelson,S.M., Kelly,D.P., and Casanova,M.F. ,2018. A Cohort Study Comparing Women with Autism Spectrum Disorder with and without Generalized Joint Hypermobility. *Behavioral Sciences (Basel, Switzerland)* 8(3).

Cashman,K.D., Dowling,K.G., et al. ,2016. Vitamin D deficiency in Europe: pandemic?12. *The American Journal of Clinical Nutrition* 103(4): p.1033–1044.

Castori. ,2016a. Hypermobile Ehlers-Danlos Syndrome and Hypermobility Spectrum Disorders. Available at: https://www.ehlers-danlos.com/pdf/2018-Maastricht/2018-Maastricht-Saturday-Castori-S.pdf.

Castori. ,2016b. Pain in Ehlers-Danlos syndromes: manifestations, therapeutic strategies and future perspectives. *Expert Opinion on Orphan Drugs* 4(11): p.1145–1158.

Castori,M., Camerota,F., et al. ,2010. Natural history and manifestations of the hypermobility type Ehlers–Danlos syndrome: A pilot study on 21 patients. *American Journal of Medical Genetics Part A* 152A(3): p.556–564.

Castori,M. ,2012. Ehlers-Danlos Syndrome, Hypermobility Type: An Underdiagnosed Hereditary Connective Tissue Disorder with Mucocutaneous, Articular, and Systemic Manifestations. *ISRN Dermatology; New York*. Available at: http://search.proquest.com/docview/1282257075/abstract/94CFC3D05EA7430EPQ/1 [Accessed May 26, 2019].

Castori,M., Morlino,S., et al. ,2013. Re-writing the natural history of pain and related symptoms in the joint hypermobility syndrome/Ehlers-Danlos syndrome, hypermobility type. *American Journal of Medical Genetics. Part A* 161A(12): p.2989–3004.

Castori,M., Dordoni,C., et al. ,2014. Nosology and inheritance pattern(s) of joint hypermobility syndrome and Ehlers-Danlos syndrome, hypermobility type: A study of intrafamilial and interfamilial variability in 23 Italian pedigrees. *American Journal of Medical Genetics Part A* 164(12): p.3010–3020.

Castori,M., Dordoni,C., et al. ,2015. Spectrum of mucocutaneous manifestations in 277 patients with joint hypermobility syndrome/Ehlers-Danlos syndrome, hypermobility

type. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 169(1): p.43–53.

Castori,M., Tinkle,B., et al. ,2017. A framework for the classification of joint hypermobility and related conditions. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175(1): p.148–157.

Castori,M., and Camerota,F. ,2010. Ehlers-Danlos syndrome hypermobility type and the excess of affected females: possible mechanisms and perspectives. *American Journal of Medical Genetics. Part A* 152A(9): p.2406–2408.

Castori,M., and Hakim,A. ,2017. Contemporary approach to joint hypermobility and related disorders. *Current Opinion in Pediatrics* 29(6): p.640–649.

Castori,M., Morlino,S., Pascolini,G., Blundo,C., and Grammatico,P. ,2015. Gastrointestinal and nutritional issues in joint hypermobility syndrome/Ehlers-Danlos syndrome, hypermobility type. *American Journal Of Medical Genetics. Part C, Seminars In Medical Genetics* 169C(1): p.54–75.

Castori,M., Sperduti,I., Celletti,C., Camerota,F., and Grammatico,P. ,2011. Symptom and joint mobility progression in the joint hypermobility syndrome (Ehlers-Danlos syndrome, hypermobility type). *Clinical and Experimental Rheumatology* 29: p.9.

Castro-Mondragon,J.A., Riudavets-Puig,R., et al. ,2022. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* 50(D1): p.D165–D173.

Celletti,C., Galli,M., et al. ,2012. Relationship between fatigue and gait abnormality in Joint Hypermobility Syndrome/Ehlers-Danlos Syndrome Hypermobility type. *Research in Developmental Disabilities* 33(6): p.1914–1918.

Celletti,C., Camerota,F., et al. ,2017. Orthostatic Intolerance and Postural Orthostatic Tachycardia Syndrome in Joint Hypermobility Syndrome/Ehlers-Danlos Syndrome, Hypermobility Type: Neurovegetative Dysregulation or Autonomic Failure? *BioMed Research International* 2017. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5329674/ [Accessed July 10, 2019].

Celletti,C., Castori,M., La Torre,G., and Camerota,F. ,2013. Evaluation of kinesiophobia and its correlations with pain and fatigue in joint hypermobility syndrome/Ehlers-Danlos syndrome hypermobility type. *BioMed Research International* 2013: p.580460.

Cesare,A.E., Rafer,L.C., Myler,C.S., and Brennan,K.B. ,2019. Anesthetic Management for Ehlers-Danlos Syndrome, Hypermobility Type Complicated by Local Anesthetic Allergy: A Case Report. *The American Journal of Case Reports* 20: p.39–42.

Chacon-Cortes,D., and Griffiths,L.R. ,2014. Methods for extracting genomic DNA from whole blood samples: current perspectives. *Journal of Biorepository Science for Applied Medicine* 2: p.1–9.

Challal,S., Minichiello,E., Funalot,B., and Boissier,M.-C. ,2015. Ehlers–Danlos syndrome in rheumatology: Diagnostic and therapeutic challenges. *Joint Bone Spine* 82(5): p.305–307.

Chen,J.S., and Sambrook,P.N. ,2011. Antiresorptive therapies for osteoporosis: a clinical overview. *Nature Reviews. Endocrinology* 8(2): p.81–91.

Chen,S., Francioli,L.C., et al. ,2022. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv*: p.2022.03.20.485034.

Chen,W., Tang,Donge, Tang,Dongxin, and Dai,Y. ,2020. Epigenetic silencing of ZIC4 contributes to cancer progression in hepatocellular carcinoma. *Cell Death & Disease* 11(10): p.1–11.

Chen,Z., Li,S., Subramaniam,S., Shyy,J.Y.-J., and Chien,S. ,2017. Epigenetic Regulation: A New Frontier for Biomedical Engineers. *Annual Review of Biomedical Engineering* 19(1): p.195–219.

Cheung,I., and Vadas,P. ,2015. A New Disease Cluster: Mast Cell Activation Syndrome, Postural Orthostatic Tachycardia Syndrome, and Ehlers-Danlos Syndrome. *Journal of Allergy and Clinical Immunology* 135(2): p.AB65.

Chiarelli,N., Carini,G., et al. ,2016. Transcriptome-Wide Expression Profiling in Skin Fibroblasts of Patients with Joint Hypermobility Syndrome/Ehlers-Danlos Syndrome Hypermobility Type. *PLOS ONE* 11(8): p.e0161347.

Chiarelli,N., Carini,G., Zoppi,N., Ritelli,M., and Colombi,M. ,2019. Molecular insights in the pathogenesis of classical Ehlers-Danlos syndrome from transcriptome-wide expression profiling of patients' skin fibroblasts. *PLOS ONE* 14(2): p.e0211647.

Chiarelli,N., Ritelli,M., Zoppi,N., and Colombi,M. ,2019. Cellular and Molecular Mechanisms in the Pathogenesis of Classical, Vascular, and Hypermobile Ehlers–Danlos Syndromes. *Genes* 10(8). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6723307/ [Accessed March 22, 2020].

Chopra,P., Tinkle,B., et al. ,2017. Pain management in the Ehlers-Danlos syndromes. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* 175(1): p.212–219.

Church,D.M., Schneider,V.A., et al. ,2011. Modernizing reference genome assemblies. *PLoS biology* 9(7): p.e1001091.

Cimolin,V., Galli,M., et al. ,2014. Foot type analysis based on electronic pedobarography data in individuals with joint hypermobility syndrome/Ehlers-Danlos syndrome hypermobility type during upright standing. *Journal of the American Podiatric Medical Association* 104(6): p.588–593.

Cingolani,P., Platts,A., et al. ,2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6(2): p.80–92.

CodonCode Corporation. ,2019. Sequence Assembly and Alignment Software - CodonCode. Available at: https://www.codoncode.com/index.htm [Accessed April 18, 2023].

Cohen,S., and Markham,F. ,2017. Ehlers–Danlos hypermobility type in an adult with chronic pain and fatigue: a case study. *Clinical Case Reports* 5(8): p.1248–1251.

Colombi,M., Dordoni,C., Chiarelli,N., and Ritelli,M. ,2015. Differential diagnosis and diagnostic flow chart of joint hypermobility syndrome/ehlers-danlos syndrome hypermobility type compared to other heritable connective tissue disorders. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* 169C(1): p.6–22.

Copetti,M., Morlino,S., et al. ,2019. Severity classes in adults with hypermobile Ehlers-Danlos syndrome/hypermobility spectrum disorders: a pilot study of 105 Italian patients. *Rheumatology (Oxford, England)*.

Cortini,F., Villa,C., et al. ,2019. Understanding the basis of Ehlers-Danlos syndrome in the era of the next-generation sequencing. *Archives of Dermatological Research* 311(4): p.265–275.

Cunningham,F., Allen,J.E., et al. ,2022. Ensembl 2022. *Nucleic Acids Research* 50(D1): p.D988–D995.

Daens,S., Grossin,D., Hermanns-Lê,T., Peeters,D., and Manicourt,D. ,2018. [Severe Mast Cell Activation Syndrome in a 15-year-old patient with an hypermobile Ehlers-Danlos syndrome]. *Revue Medicale De Liege* 73(2): p.61–64.

Danlos,H. ,1908. Un cas de cutis laxa avec tumeurs par contusion chronique des coudes et des genoux. *Bull Soc Fr Dermatol Syphiligr* 19: p.70–72.

Davydov,E.V., Goode,D.L., et al. ,2010. Identifying a High Fraction of the Human Genome to be under Selective Constraint Using GERP++ W. W. Wasserman (ed). *PLoS Computational Biology* 6(12): p.e1001025.

De Cario,R., Kura,A., et al. ,2020. Sanger Validation of High-Throughput Sequencing in Genetic Diagnosis: Still the Best Practice? *Frontiers in Genetics* 11. Available at: https://www.frontiersin.org/articles/10.3389/fgene.2020.592588 [Accessed January 24, 2023].

De Paepe,A., Devereux,R.B., Dietz,H.C., Hennekam,R.C., and Pyeritz,R.E. ,1996. Revised diagnostic criteria for the Marfan syndrome. *American Journal of Medical Genetics* 62(4): p.417–426.

De Paepe,A., and Malfait,F. ,2004. Bleeding and bruising in patients with Ehlers-Danlos syndrome and other collagen vascular disorders. *British Journal of Haematology* 127(5): p.491–500.

De Wandele,I., Rombaut,L., et al. ,2013. Clinical Heterogeneity in Patients with the Hypermobility Type of Ehlers-Danlos Syndrome. *Research in Developmental Disabilities: A Multidisciplinary Journal* 34(3): p.873–881.

De Wandele,I., Rombaut,L., et al. ,2014. Dysautonomia and its underlying mechanisms in the hypermobility type of Ehlers-Danlos syndrome. *Seminars in Arthritis and Rheumatism* 44(1): p.93–100.

Deakyne,J.S., and Mazin,A.V. ,2011. Fanconi anemia: at the Crossroads of DNA repair. *Biochemistry (Moscow)* 76(1): p.36–48.

Demirdas,S., Dulfer,E., et al. ,2017. Recognizing the tenascin-X deficient type of Ehlers-Danlos syndrome: a cross-sectional study in 17 patients. *Clinical Genetics* 91(3): p.411–425.

Demmler,J.C., Atkinson,M.D., et al. ,2019. Diagnosed prevalence of Ehlers-Danlos syndrome and hypermobility spectrum disorder in Wales, UK: a national electronic cohort study and case–control comparison. *BMJ Open* 9(11): p.e031365.

Demontiero,O., Vidal,C., and Duque,G. ,2012. Aging and bone loss: new insights for the clinician. *Therapeutic Advances in Musculoskeletal Disease* 4(2): p.61–76.

Dietz,H.C., Cutting,G.R., et al. ,1991. Marfan syndrome caused by a recurrent de novo missense mutation in the fibrillin gene. *Nature* 352(6333): p.337–339.

Ding,D.-C., Shyu,W.-C., and Lin,S.-Z. ,2011. Mesenchymal Stem Cells. *Cell Transplantation* 20(1): p.5–14.

Dolan,A.L., Arden,N.K., Grahame,R., and Spector,T.D. ,1998. Assessment of bone in Ehlers Danlos syndrome by ultrasound and densitometry. *Annals of the Rheumatic Diseases* 57(10): p.630–633.

Dong,C., Wei,P., et al. ,2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Human Molecular Genetics* 24(8): p.2125–2137.

Ehlers,E. ,1901. Cutis laxa, Neigung zu Haemorrhagien in der Haut, Lockerung mehrerer Artikulationen. *Dermat Zachr* 8: p.173–174.

Eller-Vainicher,C., Bassotti,A., et al. ,2016. Bone involvement in adult patients affected with Ehlers-Danlos syndrome. *Osteoporosis international: a journal established as result of cooperation between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA* 27(8): p.2525–2531.

Ellingford,J.M., Ahn,J.W., et al. ,2022. Recommendations for clinical interpretation of variants found in non-coding regions of the genome. *Genome Medicine* 14(1): p.73.

El-Maouche,D., Arlt,W., and Merke,D.P. ,2017. Congenital adrenal hyperplasia. *The Lancet* 390(10108): p.2194–2210.

EMQN and Certus Technology. SNPCheck V3. *SNPCheck*. Available at: https://genetools.org/SNPCheck/snpcheck.htm [Accessed November 8, 2022].

Farmer,A.D., Fikree,A., and Aziz,Q. ,2014. Addressing the confounding role of joint hypermobility syndrome and gastrointestinal involvement in postural orthostatic tachycardia syndrome. *Clinical Autonomic Research* 24(3): p.157–158.

Fikree,A., Grahame,R., et al. ,2014. A prospective evaluation of undiagnosed joint hypermobility syndrome in patients with gastrointestinal symptoms. *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association* 12(10): p.1680-1687.e2.

Fikree,A., Aktar,R., et al. ,2015. Functional gastrointestinal disorders are associated with the joint hypermobility syndrome in secondary care: a case-control study. *Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society* 27(4): p.569–579.

Fikree,A., Aktar,R., et al. ,2017. The association between Ehlers-Danlos syndrome—hypermobility type and gastrointestinal symptoms in university students: a cross-sectional study. *Neurogastroenterology & Motility* 29(3): p.e12942.

Flanagan,S.E., Patch,A.-M., and Ellard,S. ,2010. Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations. *Genetic Testing and Molecular Biomarkers* 14(4): p.533–537.

Fokkema,I.F.A.C., Taschner,P.E.M., et al. ,2011. LOVD v.2.0: the next generation in gene variant databases. *Human Mutation* 32(5): p.557–563.

Forghani,I. ,2019. Updates in Clinical and Genetics Aspects of Hypermobile Ehlers Danlos Syndrome. *Balkan Medical Journal* 36(1): p.12–16.

Francomano,C.A., Liberfarb,R.M., et al. ,1987. The Stickler syndrome: evidence for close linkage to the structural gene for type II collagen. *Genomics* 1(4): p.293–296.

Fraser,R.D., MacRae,T.P., and Suzuki,E. ,1979. Chain conformation in the collagen molecule. *Journal of Molecular Biology* 129(3): p.463–481.

Frevert,C.W., Felgenhauer,J., Wygrecka,M., Nastase,M.V., and Schaefer,L. ,2018. Danger-Associated Molecular Patterns Derived From the Extracellular Matrix Provide Temporal Control of Innate Immunity: *Journal of Histochemistry & Cytochemistry*. Available at: https://journals.sagepub.com/doi/10.1369/0022155417740880 [Accessed April 28, 2020].

Gazit,Y., Jacob,G., and Grahame,R. ,2016. Ehlers–Danlos Syndrome—Hypermobility Type: A Much Neglected Multisystemic Disorder. *Rambam Maimonides Medical Journal* 7(4). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5101008/ [Accessed July 8, 2019].

Gazit,Y., Nahir,A.M., Grahame,R., and Jacob,G. ,2003. Dysautonomia in the joint hypermobility syndrome. *The American Journal of Medicine* 115(1): p.33–40.

Gelse,K. ,2003. Collagens—structure, function, and biosynthesis. *Advanced Drug Delivery Reviews* 55(12): p.1531–1546.

Gibbons,C.H. ,2019. Basics of autonomic nervous system function. In *Handbook of Clinical Neurology*, 407–418. Elsevier Available at: https://linkinghub.elsevier.com/retrieve/pii/B9780444640321000278 [Accessed May 11, 2020].

Gjaltema,R.A.F., and Bank,R.A. ,2017. Molecular insights into prolyl and lysyl hydroxylation of fibrillar collagens in health and disease. *Critical Reviews in Biochemistry and Molecular Biology* 52(1): p.74–95.

Grahame,R. ,2000. Heritable disorders of connective tissue. *Best Practice & Research Clinical Rheumatology* 14(2): p.345–361.

Green,Y.S., and Vetter,M.L. ,2011. EBF factors drive expression of multiple classes of target genes governing neuronal development. *Neural Development* 6(1): p.19.

Groza,T., Gomez,F.L., et al. ,2023. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Research* 51(D1): p.D1038–D1045.

Grubert,F., Judith, et al. ,2015. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell* 162(5): p.1051–1065.

Guarnieri,V., and Castori,M. ,2018. Clinical Relevance of Joint Hypermobility and Its Impact on Musculoskeletal Pain and Bone Mass. *Current Osteoporosis Reports* 16(4): p.333–343.

Gulbahar,S., Sahin,E., et al. ,2006. Hypermobility syndrome increases the risk for low bone mass. *Clinical Rheumatology* 25(4): p.511–514.

Guo,Y., Dai,Y., et al. ,2017. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics* 109(2): p.83–90.

Hagemann,I.S. ,2015. Overview of Technical Aspects and Chemistries of Next-Generation Sequencing. In *Clinical Genomics*, 3–19. Elsevier Available at: https://linkinghub.elsevier.com/retrieve/pii/B9780124047488000010 [Accessed January 24, 2023].

Hakim,A., O'Callaghan,C., et al. ,2017. Cardiovascular autonomic dysfunction in Ehlers-Danlos syndrome-Hypermobile type. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175(1): p.168–174.

Hakim,A., De Wandele,I., O'Callaghan,C., Pocinki,A., and Rowe,P. ,2017. Chronic fatigue in Ehlers-Danlos syndrome-Hypermobile type. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175(1): p.175–180.

Hakim,A.J., and Grahame,R. ,2004. Non-musculoskeletal symptoms in joint hypermobility syndrome. Indirect evidence for autonomic dysfunction? *Rheumatology (Oxford, England)* 43(9): p.1194–1195.

Hakim, and Sahota. ,2006. Joint hypermobility and skin elasticity: the hereditary disorders of connective tissue. *Clinics in Dermatology* 24(6): p.521–533.

Hay,E.D. ,1981. Extracellular matrix. *The Journal of Cell Biology* 91(3 Pt 2): p.205s–223s.

Heng,J.I.-T., Qu,Z., et al. ,2015. The zinc finger transcription factor RP58 negatively regulates Rnd2 for the control of neuronal migration during cerebral cortical development. *Cerebral Cortex (New York, N.Y.: 1991)* 25(3): p.806–816.

Henney,A.M., Brotherton,D.H., Child,A.H., Humphries,S.E., and Grahame,R. ,1992. SEGREGATION ANALYSIS OF COLLAGEN GENES IN TWO FAMILIES WITH JOINT HYPERMOBILITY SYNDROME. *Rheumatology* 31(3): p.169–174.

Hermanns-Lê,T., Reginster,M.-A., et al. ,2012. Dermal ultrastructure in low Beighton score members of 17 families with hypermobile-type Ehlers-Danlos syndrome. *Journal of Biomedicine & Biotechnology* 2012: p.878107.

Hermanns-Lê,T., and Piérard,G.E. ,2007. Ultrastructural alterations of elastic fibers and other dermal components in ehlers-danlos syndrome of the hypermobile type. *The American Journal of Dermatopathology* 29(4): p.370–373.

Hermanns-Lê,T., Piérard,G.E., and Angenot,P. ,2013. [Fibromyalgia: an unrecognized Ehlers-Danlos syndrome hypermobile type?]. *Revue Medicale De Liege* 68(1): p.22–24.

Hermanns-Lê,T., Piérard,G.E., and Piérard-Franchimont,C. ,2015. [EHLERS-DANLOS SYNDROME OF THE HYPERMOBILE TYPE: A MULTISYSTEMIC DISORDER. CONTRIBUTION OF SKIN ULTRASTRUCTURE TO INDIVIDUAL MANAGEMENT]. *Revue Medicale De Liege* 70(5–6): p.325–330.

Hernández,J.L., Olmos,J.M., et al. ,2013. Influence of Vitamin D Status on Vertebral Fractures, Bone Mineral Density, and Bone Turnover Markers in Normocalcemic Postmenopausal Women With High Parathyroid Hormone Levels. *The Journal of Clinical Endocrinology & Metabolism* 98(4): p.1711–1717.

Hoffman,G.G., Dodson,G.E., Cole,W.G., and Greenspan,D.S. ,2008. Absence of apparent disease causing mutations in COL5A3 in 13 patients with hypermobility Ehlers–Danlos syndrome. *American Journal of Medical Genetics Part A* 146A(24): p.3240–3241.

Hugon-Rodin,J., Lebègue,G., Becourt,S., Hamonet,C., and Gompel,A. ,2016. Gynecologic symptoms and the influence on reproductive life in 386 women with hypermobility type ehlers-danlos syndrome: a cohort study. *Orphanet Journal of Rare Diseases* 11(1): p.124.

Hynes,R.O. ,2009. Extracellular matrix: not just pretty fibrils. *Science (New York, N.Y.)* 326(5957): p.1216–1219.

Illumina. ,2011. Quality Scores for Next-Generation Sequencing.

Illumina. ,2017. TruSeq DNA PCR-Free catalogue. : p.4.

International Consortium on EDS & Related Disorders. ,2017. hEDS Diagnostic Criteria Checklist. Available at: https://ehlers-danlos.com/wp-content/uploads/hEDS-Dx-Criteria-checklist-1.pdf [Accessed July 12, 2019].

Ioannidis,N.M., Rothstein,J.H., et al. ,2016. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *The American Journal of Human Genetics* 99(4): p.877–885.

Ittisoponpisan,S., Islam,S.A., et al. ,2019. Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? *Journal of Molecular Biology* 431(11): p.2197–2212.

Jacobs,J.W.G., Cornelissens,L.J.M., and Veenhuizen,M.C. ,2018. *Ehlers-Danlos Syndrome: A Multidisciplinary Approach*. IOS Press.

Jenkins,K.A., Fossat,M.J., et al. ,2018. The consequences of cavity creation on the folding landscape of a repeat protein depend upon context. *Proceedings of the National Academy of Sciences* 115(35): p.E8153–E8161.

Jesudas,R., Chaudhury,A., and Laukaitis,C.M. ,2019. An update on the new classification of Ehlers-Danlos syndrome and review of the causes of bleeding in this population. *Haemophilia* 25(4): p.558–566.

Jin,K., and Xiang,M. ,2019. Transcription factor Ptf1a in development, diseases and reprogramming. *Cellular and Molecular Life Sciences* 76.

Jumper,J., Evans,R., et al. ,2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596(7873): p.583–589.

Kanis,J.A., Burlet,N., et al. ,2008. European guidance for the diagnosis and management of osteoporosis in postmenopausal women. *Osteoporosis International* 19(4): p.399–428.

Kent,W.J., Sugnet,C.W., et al. ,2002. The Human Genome Browser at UCSC. *Genome Research* 12(6): p.996–1006.

Khan,M., Jose,A., and Sharma,S. ,2023. Physiology, Parathyroid Hormone. In *StatPearls*, Treasure Island (FL): StatPearls Publishing Available at: http://www.ncbi.nlm.nih.gov/books/NBK499940/ [Accessed June 19, 2023].

Kibbe,W.A. ,2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* 35(Web Server issue): p.W43-46.

Kienesberger,P.C., Oberer,M., Lass,A., and Zechner,R. ,2009. Mammalian patatin domain containing proteins: a family with diverse lipolytic activities involved in multiple biological functions. *Journal of Lipid Research* 50(Suppl): p.S63–S68.

Kieslinger,M., Folberth,S., et al. ,2005. EBF2 regulates osteoblast-dependent differentiation of osteoclasts. *Developmental Cell* 9(6): p.757–767.

Köhler,S., Gargano,M., et al. ,2021. The Human Phenotype Ontology in 2021. *Nucleic Acids Research* 49(D1): p.D1207–D1217.

Kohn,A., and Chang,C. ,2019. The Relationship Between Hypermobile Ehlers-Danlos Syndrome (hEDS), Postural Orthostatic Tachycardia Syndrome (POTS), and Mast Cell Activation Syndrome (MCAS). *Clinical Reviews in Allergy & Immunology*.

Kopanos,C., Tsiolkas,V., et al. ,2019. VarSome: the human genomic variant search engine. *Bioinformatics* 35(11): p.1978–1980.

Krahe,A.M., Adams,R.D., and Nicholson,L.L. ,2018. Features that exacerbate fatigue severity in joint hypermobility syndrome/Ehlers-Danlos syndrome - hypermobility type. *Disability and Rehabilitation* 40(17): p.1989–1996.

Krude,H., Mundlos,S., Øien,N.C., Opitz,R., and Schuelke,M. ,2021. What can go wrong in the non-coding genome and how to interpret whole genome sequencing data. *Medizinische Genetik* 33(2): p.121–131.

Kuivaniemi,H., Tromp,G., and Prockop,D.J. ,1997. Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type IX), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels. *Human Mutation* 9(4): p.300–315.

Kumar,B., and Lenert,P. ,2017. Joint Hypermobility Syndrome: Recognizing a Commonly Overlooked Cause of Chronic Pain. *The American Journal of Medicine* 130(6): p.640–647.

Kumar,M., Michael,S., et al. ,2022. The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Research* 50(D1): p.D497–D508.

Lamandé,S.R., and Bateman,J.F. ,2019. Genetic Disorders of the Extracellular Matrix. *The Anatomical Record*. Available at: https://anatomypubs.onlinelibrary.wiley.com/doi/abs/10.1002/ar.24086 [Accessed April 28, 2020].

Landrum,M.J., Lee,J.M., et al. ,2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Research* 46(D1): p.D1062–D1067.

Layer,R.M., Chiang,C., Quinlan,A.R., and Hall,I.M. ,2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biology* 15(6): p.R84.

Leblanc,D.R., Schneider,M., Angele,P., Vollmer,G., and Docheva,D. ,2017. The effect of estrogen on tendon and ligament metabolism and function. *The Journal of Steroid Biochemistry and Molecular Biology* 172: p.106–116.

Lee,D., and Mueller,E. ,2017. Mast Cell Activation Features in Ehlers-Danlos/Joint Hypermobility Patients: A Retrospective Analysis in Light of an Emerging Disease Cluster. Available at: https://acrabstracts.org/abstract/mast-cell-activation-features-in-ehlers-danlosjoint-hypermobility-patients-a-retrospective-analysis-in-light-of-an-emerging-disease-cluster/ [Accessed March 15, 2020].

Levy,H.P. ,2018. Hypermobile Ehlers-Danlos Syndrome. In M. P. Adam et al. (eds) *GeneReviews*®, Seattle (WA): University of Washington, Seattle Available at: http://www.ncbi.nlm.nih.gov/books/NBK1279/ [Accessed June 12, 2019].

Li,H., Handsaker,B., et al. ,2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16): p.2078–2079.

Li,H., and Durbin,R. ,2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14): p.1754–1760.

Li,Q., and Wang,K. ,2017. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *The American Journal of Human Genetics* 100(2): p.267–280.

Li,X.-H., and Babu,M.M. ,2018. Human Diseases from Gain-of-Function Mutations in Disordered Protein Regions. *Cell* 175(1): p.40–42.

Li,Y., Zhang,Y., Li,X., Yi,S., and Xu,J. ,2019. Gain-of-Function Mutations: An Emerging Advantage for Cancer Biology. *Trends in biochemical sciences* 44(8): p.659–674.

Liu,F.-Y. ,2010. Smad anchor for receptor activation SARA in TGF-beta signaling. *Frontiers in Bioscience* E2(3): p.857–860.

Liu,X., Jian,X., and Boerwinkle,E. ,2011. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Human Mutation* 32(8): p.894–899.

Lupiáñez,D.G., Kraft,K., et al. ,2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161(5): p.1012–1025.

Lyons,J.J., Yu,X., et al. ,2016. Elevated basal serum tryptase identifies a multisystem disorder associated with increased *TPSAB1* copy number. *Nature Genetics* 48(12): p.1564–1569.

Macias,M.J., Martin-Malpartida,P., and Massagué,J. ,2015. Structural determinants of SMAD function in TGF-β signaling. *Trends in biochemical sciences* 40(6): p.296–308.

Majewski,J., Schwartzentruber,J., Lalonde,E., Montpetit,A., and Jabado,N. ,2011. What can exome sequencing do for you? *Journal of Medical Genetics* 48(9): p.580–589.

Malfait,F., Francomano,C., et al. ,2017. The 2017 international classification of the Ehlers-Danlos syndromes. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175(1): p.8–26.

Malfait,F., Hakim,A.J., De Paepe,A., and Grahame,R. ,2006. The genetic basis of the joint hypermobility syndromes. *Rheumatology* 45(5): p.502–507.

Mao,J.R., Taylor,G., et al. ,2002. Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. *Nature Genetics* 30(4): p.421–425.

Martin,A. ,2019. An acquired or heritable connective tissue disorder? A review of hypermobile Ehlers Danlos Syndrome. *European Journal of Medical Genetics*: p.103672.

Martin-Malpartida,P., Batet,M., et al. ,2017. Structural basis for genome wide recognition of 5-bp GC motifs by SMAD transcription factors. *Nature Communications* 8(1): p.2070.

Masago,K., Fujita,S., et al. ,2021. Comparison between Fluorimetry (Qubit) and Spectrophotometry (NanoDrop) in the Quantification of DNA and RNA Extracted from Frozen and FFPE Tissues from Lung Cancer Patients: A Real-World Use of Genomic Tests. *Medicina (Kaunas).* 57(12): p.1375.

Mast,K.J., Nunes,M.E., Ruymann,F.B., and Kerlin,B.A. ,2009. Desmopressin responsiveness in children with Ehlers-Danlos syndrome associated bleeding symptoms. *British Journal of Haematology* 144(2): p.230–233.

Matlock,B. ,2015. Assessment of Nucleic Acid Purity. *Technical Note 52646*: p.3.

Mazziotti,G., Dordoni,C., et al. ,2016. High prevalence of radiological vertebral fractures in adult patients with Ehlers–Danlos syndrome. *Bone* 84: p.88–92.

McDonald,J., and Stevenson,D.A. ,2021. Hereditary Hemorrhagic Telangiectasia. In M. P. Adam et al. (eds) *GeneReviews®*, Seattle (WA): University of Washington, Seattle

Available at: http://www.ncbi.nlm.nih.gov/books/NBK1351/ [Accessed June 24, 2023].

McGillis,L., Mittal,N., et al. ,2020. Utilization of the 2017 diagnostic criteria for hEDS by the Toronto GoodHope Ehlers–Danlos syndrome clinic: A retrospective review. *American Journal of Medical Genetics Part A* 182(3): p.484–492.

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 Reports* 9(1): p.1–15.

McKenna,A., Hanna,M., et al. ,2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20(9): p.1297–1303.

McKusick-Nathans Institute of Genetic Medicine. Online Mendelian Inheritance in Man, OMIM®. *OMIM*®. Available at: World Wide Web URL: https://omim.org/ [Accessed January 19, 2021].

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. ,2023. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. *Nucleic Acids Research*. Available at: https://omim.org/ [Accessed June 6, 2023].

McLaren,W., Gil,L., et al. ,2016. The Ensembl Variant Effect Predictor. *Genome Biology* 17(1): p.122.

Menys,A., Keszthelyi,D., et al. ,2017. A magnetic resonance imaging study of gastric motor function in patients with dyspepsia associated with Ehlers-Danlos Syndrome-Hypermobility Type: A feasibility study. *Neurogastroenterology and Motility: The Official Journal of the European Gastrointestinal Motility Society* 29(9).

Merke,D.P., Chen,W., et al. ,2013. Tenascin-X haploinsufficiency associated with Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. *The Journal of Clinical Endocrinology and Metabolism* 98(2): p.E379-387.

Miller,A.J., Stiles,L.E., et al. ,2020. Prevalence of hypermobile Ehlers-Danlos syndrome in postural orthostatic tachycardia syndrome. *Autonomic Neuroscience: Basic & Clinical* 224: p.102637.

Miller,S.A., Dykes,D.D., and Polesky,H.F. ,1988. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 16(3): p.1215.

Miyamoto,M., Itoh,N., Sawai,M., Sassa,T., and Kihara,A. ,2020. Severe Skin Permeability Barrier Dysfunction in Knockout Mice Deficient in a Fatty Acid ω-Hydroxylase Crucial to Acylceramide Production. *Journal of Investigative Dermatology* 140(2): p.319-326.e4.

Mizuno,K., Boudko,S., Engel,J., and Bächinger,H.P. ,2013. Vascular Ehlers-Danlos Syndrome Mutations in Type III Collagen Differently Stall the Triple Helical Folding. *The Journal of Biological Chemistry* 288(26): p.19166.

Moruzzo,D., Nobbio,L., et al. ,2017. The Transcription Factors EBF1 and EBF2 Are Positive Regulators of Myelination in Schwann Cells. *Molecular Neurobiology* 54(10): p.8117–8127.

Moustakas,A., Souchelnytskyi,S., and Heldin,C.-H. ,2001. Smad regulation in TGF-β signal transduction. *Journal of Cell Science* 114(24): p.4359–4369.

Mullis,K., Faloona,F., et al. ,1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51 Pt 1: p.263–273.

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): p.344–354.

Murray,B., Yashar,B.M., Uhlmann,W.R., Clauw,D.J., and Petty,E.M. ,2013. Ehlers-Danlos syndrome, hypermobility type: A characterization of the patients' lived experience. *American Journal of Medical Genetics. Part A* 161A(12): p.2981–2988.

MutPred2. MutPred2: a predictor of impactful missense variants- Output Help. Available at: http://mutpred2.mutdb.org/help.html#interpret [Accessed May 24, 2023].

Myllyharju,J., and Kivirikko,K.I. ,2001. Collagens and collagen-related diseases. *Annals of Medicine* 33(1): p.7–21.

Narcisi,P., J.Richards,A., Ferguson,S.D., and Pope,F.M. ,1994. A family with Ehlers — Danlos syndrome type III/articular hypermobility syndrome has a glycine 637 to serine substitution in type III collagen. *Human Molecular Genetics* 3(9): p.1617–1620.

Ofluoglu,D., Gunduz,O.H., Kul-Panza,E., and Guven,Z. ,2006. Hypermobility in women with fibromyalgia syndrome. *Clinical Rheumatology* 25(3): p.291–293.

Ogawa,M., LaRue,A.C., Watson,P.M., and Watson,D.K. ,2010. Hematopoietic stem cell origin of connective tissues. *Experimental Hematology* 38(7): p.540–547.

Okonechnikov,K., Conesa,A., and García-Alcalde,F. ,2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics (Oxford, England)* 32(2): p.292–294.

Owen,C.R., Yuan,L., and Basson,M.D. ,2008. Smad3 knockout mice exhibit impaired intestinal mucosal healing. *Laboratory Investigation* 88(10): p.1101–1109.

Palmer,S. ,2014. The Impact of Joint Hypermobility Syndrome.

Palmer,S., Manns,S., Cramp,F., Lewis,R., and Clark,E.M. ,2017. Test-retest reliability and smallest detectable change of the Bristol Impact of Hypermobility (BIoH) questionnaire. *Musculoskeletal Science and Practice* 32: p.64–69.

Parthiban,V., Gromiha,M.M., and Schomburg,D. ,2006. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Research* 34(Web Server issue): p.W239-242.

Paysan-Lafosse,T., Blum,M., et al. ,2023. InterPro in 2022. *Nucleic Acids Research* 51(D1): p.D418–D427.

Pejaver,V., Urresti,J., et al. ,2020. Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nature Communications* 11(1): p.5918.

Perez-Becerril,C., Evans,D.G., and Smith,M.J. ,2021. Pathogenic noncoding variants in the neurofibromatosis and schwannomatosis predisposition genes. *Human Mutation* 42(10): p.1187–1207.

Petersen,J.W., and Douglas,J.Y. ,2013. Tenascin-X, collagen, and Ehlers–Danlos syndrome: Tenascin-X gene defects can protect against adverse cardiovascular events. *Medical Hypotheses* 81(3): p.443–447.

Pichery,M., Huchenq,A., et al. ,2017. PNPLA1 defects in patients with autosomal recessive congenital ichthyosis and KO mice sustain PNPLA1 irreplaceable function in epidermal omega-O-acylceramide synthesis and skin permeability barrier. *Human Molecular Genetics* 26(10): p.1787–1800.

van der Pluijm,I., van Vliet,N., et al. ,2016. Defective Connective Tissue Remodeling in Smad3 Mice Leads to Accelerated Aneurysmal Growth Through Disturbed Downstream TGF-β Signaling. *EBioMedicine* 12: p.280–294.

Pollard,K.S., Hubisz,M.J., Rosenbloom,K.R., and Siepel,A. ,2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research* 20(1): p.110–121.

Ponting,C.P. ,2017. Biological function in the twilight zone of sequence conservation. *BMC Biology* 15(1): p.71.

Porcher,C., Chagraoui,H., and Kristiansen,M.S. ,2017. SCL/TAL1: a multifaceted regulator from blood development to disease. *Blood* 129(15): p.2051–2060.

Quail,M.A., Smith,M., et al. ,2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13(1): p.341.

Quatman,C.E., Ford,K.R., Myer,G.D., Paterno,M.V., and Hewett,T.E. ,2008. The effects of gender and pubertal status on generalized joint laxity in young athletes. *Journal of Science and Medicine in Sport* 11(3): p.257–263.

Rajakumari,S., Wu,J., et al. ,2013. EBF2 Determines and Maintains Brown Adipocyte Identity. *Cell Metabolism* 17(4): p.562–574.

Rappaport,N., Twik,M., et al. ,2017. MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Research* 45(D1): p.D877–D887.

Remvig,L., Jensen,D.V., and Ward,R.C. ,2007. Epidemiology of general joint hypermobility and basis for the proposed criteria for benign joint hypermobility syndrome: review of the literature. *The Journal of Rheumatology* 34(4): p.804–809.

Rentzsch,P., Schubach,M., Shendure,J., and Kircher,M. ,2021. CADD-Splice—improving genome-wide variant effect prediction using deep learning-derived splice scores. *Genome Medicine* 13(1): p.31.

Reuter,J.A., Spacek,D., and Snyder,M.P. ,2015. High-Throughput Sequencing Technologies. *Molecular cell* 58(4): p.586–597.

Ritelli,M., Chiarelli,N., et al. ,2022. RNA-Seq of Dermal Fibroblasts from Patients with Hypermobile Ehlers–Danlos Syndrome and Hypermobility Spectrum Disorders Supports Their Categorization as a Single Entity with Involvement of Extracellular Matrix Degrading and Proinflammatory Pathomechanisms. *Cells* 11(24): p.4040.

Ritter,A., Atzinger,C., et al. ,2017. Natural history of aortic root dilation through young adulthood in a hypermobile Ehlers-Danlos syndrome cohort. *American Journal of Medical Genetics. Part A* 173(6): p.1467–1472.

Robinson,J.T., Thorvaldsdóttir,H., et al. ,2011. Integrative genomics viewer. *Nature Biotechnology* 29(1): p.24–26.

Rodgers,K.R., Gui,J., Dinulos,M.B.P., and Chou,R.C. ,2017. Ehlers-Danlos syndrome hypermobility type is associated with rheumatic diseases. *Scientific Reports* 7: p.39636.

Rogers,M.F., Shihab,H.A., et al. ,2018. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. *Bioinformatics (Oxford, England)* 34(3): p.511–513.

Rombaut,L., Malfait,F., et al. ,2011. Balance, gait, falls, and fear of falling in women with the hypermobility type of Ehlers-Danlos syndrome. *Arthritis Care & Research* 63(10): p.1432–1439.

Rombaut,L., Malfait,F., De Wandele,I., Mahieu,N., et al. ,2012. Muscle-tendon tissue properties in the hypermobility type of Ehlers-Danlos syndrome. *Arthritis Care & Research* 64(5): p.766–772.

Rombaut,L., Malfait,F., De Wandele,I., Taes,Y., et al. ,2012. Muscle mass, muscle strength, functional performance, and physical impairment in women with the hypermobility type of Ehlers-Danlos syndrome. *Arthritis Care & Research* 64(10): p.1584–1592.

Rombaut,L., Malfait,F., Cools,A., Paepe,A.D., and Calders,P. ,2010. Musculoskeletal complaints, physical activity and health-related quality of life among patients with the Ehlers–Danlos syndrome hypermobility type. *Disability and Rehabilitation* 32(16): p.1339–1345.

Rosmarin,A.G., Resendes,K.K., Yang,Z., McMillan,J.N., and Fleming,S.L. ,2004. GA-binding protein transcription factor: a review of GABP as an integrator of intracellular signaling and protein-protein interactions. *Blood Cells, Molecules & Diseases* 32(1): p.143–154.

Ross,J., and Grahame,R. ,2011. Joint hypermobility syndrome. *BMJ* 342: p.c7167.

Rossi,E., Lopez-Novoa,J.M., and Bernabeu,C. ,2015. Endoglin involvement in integrin-mediated cell adhesion as a putative pathogenic mechanism in hereditary hemorrhagic telangiectasia type 1 (HHT1). *Frontiers in Genetics* 5. Available at:

https://www.frontiersin.org/articles/10.3389/fgene.2014.00457 [Accessed June 24, 2023].

Rossi,V., Lee,B., and Marom,R. ,2019. Osteogenesis imperfecta: advancements in genetics and treatment. *Current Opinion in Pediatrics* 31(6): p.708–715.

Rozario,T., and DeSimone,D.W. ,2010. The extracellular matrix in development and morphogenesis: A dynamic view. *Developmental Biology* 341(1): p.126–140.

Rydel,T.J., Williams,J.M., et al. ,2003. The Crystal Structure, Mutagenesis, and Activity Studies Reveal that Patatin Is a Lipid Acyl Hydrolase with a Ser-Asp Catalytic Dyad. *Biochemistry* 42(22): p.6696–6708.

Sanger,F., Nicklen,S., and Coulson,A.R. ,1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74(12): p.5463–5467.

Sasivimol, Amnuaywattakorn, et al. ,2022. Cross calibration between two dual energy X-ray absorptiometry systems: Horizon A and Discovery A. *Chulalongkorn Medical Journal* 66: p.145151.

Satterlee,J.S., Chadwick,L.H., et al. ,2019. The NIH Common Fund/Roadmap Epigenomics Program: Successes of a comprehensive consortium. *Science Advances* 5(7): p.eaaw6507.

Sayers,E.W., Beck,J., et al. ,2021. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 49(D1): p.D10–D17.

Schalkwijk,J., Zweers,M.C., et al. ,2001. A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X deficiency. *The New England Journal of Medicine* 345(16): p.1167–1175.

Scheiner,S., Kar,T., and Pattanayak,J. ,2002. Comparison of Various Types of Hydrogen Bonds Involving Aromatic Amino Acids. *Journal of the American Chemical Society* 124(44): p.13257–13264.

Scheper,M., Rombaut,L., et al. ,2017. The association between muscle strength and activity limitations in patients with the hypermobility type of Ehlers-Danlos syndrome: the impact of proprioception. *Disability and Rehabilitation* 39(14): p.1391–1397.

Scheper,M.C., de Vries,J.E., Verbunt,J., and Engelbert,R.H. ,2015. Chronic pain in hypermobility syndrome and Ehlers–Danlos syndrome (hypermobility type): it is a challenge. *Journal of Pain Research* 8: p.591–601.

Schoch,K., Tan,Q.K.-G., et al. ,2020. Alternative transcripts in variant interpretation: the potential for missed diagnoses and misdiagnoses. *Genetics in medicine : official journal of the American College of Medical Genetics* 22(7): p.1269–1275.

Schubart,J.R., Schaefer,E., Hakim,A.J., Francomano,C.A., and Bascom,R. ,2019. Use of Cluster Analysis to Delineate Symptom Profiles in Ehlers-Danlos Syndrome Patient Population. *Journal of Pain and Symptom Management*.

Scicluna,K., Formosa,M.M., Farrugia,R., and Borg,I. ,2022. Hypermobile Ehlers–Danlos syndrome: A review and a critical appraisal of published genetic research to date. *Clinical Genetics* 101(1): p.20–31.

Seneviratne,S.L., Maitland,A., and Afrin,L. ,2017. Mast cell disorders in Ehlers-Danlos syndrome. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* 175(1): p.226–236.

Shovlin,C.L., Buscarini,E., et al. ,2022. The European Rare Disease Network for HHT Frameworks for management of hereditary haemorrhagic telangiectasia in general and speciality care. *European Journal of Medical Genetics* 65(1): p.104370.

Siepel,A., Bejerano,G., et al. ,2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research* 15(8): p.1034–1050.

Šimundić,A.-M. ,2013. Bias in research. *Biochemia Medica* 23(1): p.12–15.

Sinibaldi,L., Ursini,G., and Castori,M. ,2015. Psychopathological manifestations of joint hypermobility and joint hypermobility syndrome/ Ehlers–Danlos syndrome, hypermobility type: The link between connective tissue and psychological distress revised. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 169(1): p.97–106.

Slatko,B.E., Gardner,A.F., and Ausubel,F.M. ,2018. Overview of Next-Generation Sequencing Technologies. *Current Protocols in Molecular Biology* 122(1). Available at: https://onlinelibrary.wiley.com/doi/10.1002/cpmb.59 [Accessed December 1, 2022].

Smith,T.O., Easton,V., et al. ,2014. The relationship between benign joint hypermobility syndrome and psychological distress: a systematic review and meta-analysis. *Rheumatology (Oxford, England)* 53(1): p.114–122.

Steinhaus,R., Proft,S., et al. ,2021. MutationTaster2021. *Nucleic Acids Research* 49(W1): p.W446–W451.

Steinmann,B., Royce,P.M., and Superti-Furga,A. ,2003. The Ehlers-Danlos Syndrome. In *Connective Tissue and Its Heritable Disorders*, 431–523. John Wiley & Sons, Ltd Available at: http://onlinelibrary.wiley.com/doi/abs/10.1002/0471221929.ch9 [Accessed June 20, 2019].

Stelzer,G., Rosen,N., et al. ,2016. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics* 54: p.1.30.1-1.30.33.

Stephens,K.E., Miaskowski,C.A., Levine,J.D., Pullinger,C.R., and Aouizerat,B.E. ,2013. Epigenetic Regulation and Measurement of Epigenetic Changes. *Biological research for nursing* 15(4): p.373–381.

Sulli,A., Talarico,R., et al. ,2018. Ehlers-Danlos syndromes: state of the art on clinical practice guidelines. *RMD Open* 4(Suppl 1): p.e000790.

Syx,D., Symoens,S., et al. ,2015. Ehlers-Danlos Syndrome, Hypermobility Type, Is Linked to Chromosome 8p22-8p21.1 in an Extended Belgian Family. *Disease Markers* 2015: p.828970.

Talevich,E., Shain,A.H., Botton,T., and Bastian,B.C. ,2016. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLOS Computational Biology* 12(4): p.e1004873.

Tamilselvam,T.N., and Malarvizhi. ,2017. Biochemical marker changes benign hypermobility syndrome (BHMS). *International Archives of Integrated Medicine* 4(2). Available at: http://iaimjournal.com/wp-content/uploads/2017/02/iaim_2017_0402_04.pdf [Accessed March 6, 2020].

The International Consortium on EDS & Related Disorders. ,2023. Update on the diagnostic criteria for hEDS, the definition of HSD, and EDS diagnostic pathway work. *The Ehlers Danlos Society*. Available at: https://www.ehlers-danlos.com/criteria-and-diagnostic-pathway-update/ [Accessed June 12, 2023].

The UniProt Consortium. ,2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Research* 51(D1): p.D523–D531.

Theocharis,A.D., Manou,D., and Karamanos,N.K. ,2019. The extracellular matrix as a multitasking player in disease. *The FEBS journal* 286(15): p.2830–2869.

Theodorou,S.J., Theodorou,D.J., Kakitsubata,Y., and Adams,J.E. ,2012. Low Bone Mass in Ehlers-Danlos Syndrome. *Internal Medicine* 51(22): p.3225–3226.

ThermoFisher Scientific. Sanger sequencing history, overview, and workflow. Available at: https://www.thermofisher.com/tr/en/home/life-science/sequencing/sequencing-learning-center/capillary-electrophoresis-information/what-is-sanger-sequencing.html [Accessed January 24, 2023a].

ThermoFisher Scientific. Sequencing Coverage and Throughput. *Sequencing Coverage and Throughput*. Available at: https://www.thermofisher.com/tr/en/home/life-science/sequencing/sequencing-learning-center/next-generation-sequencing-information/ngs-basics/importance-coverage-throughput.html [Accessed February 8, 2023b].

Times of Malta. ,2018. Over 57% of people tested had low Vitamin D levels. *Times of Malta*. Available at: https://timesofmalta.com/articles/view/over-57-of-people-tested-had-low-vitamin-d-levels.679720.

Tinkle, Castori,M., et al. ,2017. Hypermobile Ehlers-Danlos syndrome (a.k.a. Ehlers-Danlos syndrome Type III and Ehlers-Danlos syndrome hypermobility type): Clinical description and natural history. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 175(1): p.48–69.

Tusnády,G.E., Simon,I., Váradi,A., and Arányi,T. ,2005. BiSearch: primer-design and search tool for PCR on bisulfite-treated genomes. *Nucleic Acids Research* 33(1): p.e9.

Unger,S., Ferreira,C.R., et al. ,2023. Nosology of genetic skeletal disorders: 2023 revision. *American Journal of Medical Genetics Part A* 191(5): p.1164–1209.

Untergasser,A., Cutcutache,I., et al. ,2012. Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40(15): p.e115.

Valcourt,U., Alcaraz,L.B., Exposito,J.-Y., Lethias,C., and Bartholin,L. ,2015. Tenascin-X: beyond the architectural function. *Cell Adhesion & Migration* 9(1–2): p.154–165.

Vanuytsel,K., Cai,Q., et al. ,2014. FANCA knockout in human embryonic stem cells causes a severe growth disadvantage. *Stem Cell Research* 13(2): p.240–250.

Varrault,A., Dantec,C., et al. ,2017. Identification of Plagl1/Zac1 binding sites and target genes establishes its role in the regulation of extracellular matrix genes and the imprinted gene network. *Nucleic Acids Research* 45(18): p.10466–10480.

Varshney,A., Scott,L.J., et al. ,2017. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. *Proceedings of the National Academy of Sciences* 114(9): p.2301–2306.

Vaser,R., Adusumalli,S., Leng,S.N., Sikic,M., and Ng,P.C. ,2016. SIFT missense predictions for genomes. *Nature Protocols* 11(1): p.1–9.

Vaughan,C.A., Singh,S., et al. ,2017. Gain-of-function p53 activates multiple signaling pathways to induce oncogenicity in lung cancer cells. *Molecular Oncology* 11(6): p.696–711.

Vaz-Drago,R., Custódio,N., and Carmo-Fonseca,M. ,2017. Deep intronic mutations and human disease. *Human Genetics* 136(9): p.1093–1111.

Visvader,J.E., Fujiwara,Y., and Orkin,S.H. ,1998. Unsuspected role for the T-cell leukemia protein SCL/tal-1 in vascular development. *Genes & Development* 12(4): p.473–479.

Voermans,N.C., Verrijp,K., et al. ,2011. Mild muscular features in tenascin-X knockout mice, a model of Ehlers-danlos syndrome. *Connective Tissue Research* 52(5): p.422–432.

Voermans,N.C., Knoop,H., Bleijenberg,G., and van Engelen,B.G. ,2010. Pain in ehlers-danlos syndrome is common, severe, and associated with functional impairment. *Journal of Pain and Symptom Management* 40(3): p.370–378.

Wallace,R.B., Shaffer,J., et al. ,1979. Hybridization of synthetic oligodeoxyribonucleotides to phi chi 174 DNA: the effect of single base pair mismatch. *Nucleic Acids Research* 6(11): p.3543–3557.

Wang,J., Abhinav,P., et al. ,2019. NR2F2 loss-of-function mutation is responsible for congenital bicuspid aortic valve. *International Journal of Molecular Medicine* 43(4): p.1839–1846.

Wang,M., Sampson,E.R., et al. ,2013. MMP13 is a critical target gene during the progression of osteoarthritis. *Arthritis Research & Therapy* 15(1): p.R5.

Ward,L.D., and Kellis,M. ,2012. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Research* 40(D1): p.D930–D934.

Watanabe,A., Satoh,K., Maniwa,T., and Matsumoto,K.-I. ,2016. Proteomic analysis for the identification of serum diagnostic markers for joint hypermobility syndrome. *International Journal of Molecular Medicine* 37(2): p.461–467.

Waters,D.L.E., and Shapter,F.M. ,2014. The polymerase chain reaction (PCR): general methods. *Methods in Molecular Biology (Clifton, N.J.)* 1099: p.65–75.

Weerakkody,R.A., Vandrovcova,J., et al. ,2016. Targeted next-generation sequencing makes new molecular diagnoses and expands genotype-phenotype relationship in Ehlers-Danlos syndrome. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 18(11): p.1119–1127.

Wehrwein,E.A., Orer,H.S., and Barman,S.M. ,2016. Overview of the Anatomy, Physiology, and Pharmacology of the Autonomic Nervous System. In R. Terjung (ed) *Comprehensive Physiology*, 1239–1278. Hoboken, NJ, USA: John Wiley & Sons, Inc. Available at: http://doi.wiley.com/10.1002/cphy.c150037 [Accessed May 11, 2020].

Wheeler,J.B., Ikonomidis,J.S., and Jones,J.A. ,2021. Connective Tissue Disorders and Cardiovascular Complications: The Indomitable Role of Transforming Growth Factor-Beta Signaling. *Advances in experimental medicine and biology* 1348: p.161–184.

Whittaker,L.G., McNamara,E.A., et al. ,2018. Direct Comparison of the Precision of the New Hologic Horizon Model With the Old Discovery Model. *Journal of Clinical Densitometry* 21(4): p.524–528.

Xue,M., Wakamoto,T., et al. ,2019. How internal cavities destabilize a protein. *Proceedings of the National Academy of Sciences of the United States of America* 116(42): p.21031–21036.

Yamada,K., Watanabe,A., et al. ,2019. Measurement of Serum Tenascin-X in Joint Hypermobility Syndrome Patients. *Biological and Pharmaceutical Bulletin* 42(9): p.1596–1599.

Yen,J.-L., Lin,S.-P., Chen,M.-R., and Niu,D.-M. ,2006. Clinical Features of Ehlers-Danlos Syndrome. *Journal of the Formosan Medical Association* 105(6): p.475–480.

Zahn-Zabal,M., Michel,P.-A., et al. ,2020. The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Research* 48(D1): p.D328–D334.

Zarate,N., Farmer,A.D., et al. ,2010. Unexplained gastrointestinal symptoms and joint hypermobility: is connective tissue the missing link? *Neurogastroenterology & Motility* 22(3): p.252-e78.

Zhang,D., Tai,L.K., et al. ,2005. Proteomic Study Reveals That Proteins Involved in Metabolic and Detoxification Pathways Are Highly Expressed in HER-2/neu-positive Breast Cancer*. *Molecular & Cellular Proteomics* 4(11): p.1686–1696.

Zhao,S., Gardner,K., Taylor,W., Marks,E., and Goodson,N. ,2015. Vitamin D assessment in primary care: changing patterns of testing. *London Journal of Primary Care* 7(2): p.15–22.

Zoppi,N., Chiarelli,N., Binetti,S., Ritelli,M., and Colombi,M. ,2018. Dermal fibroblast-to-myofibroblast transition sustained by αvß3 integrin-ILK-Snail1/Slug signaling is a common feature for hypermobile Ehlers-Danlos syndrome and hypermobility spectrum disorders. *Biochimica Et Biophysica Acta. Molecular Basis of Disease* 1864(4 Pt A): p.1010–1023.

Zoppi,N., Gardella,R., De Paepe,A., Barlati,S., and Colombi,M. ,2004. Human fibroblasts with mutations in COL5A1 and COL3A1 genes do not organize collagens and fibronectin in the extracellular matrix, down-regulate alpha2beta1 integrin, and recruit alphavbeta3 Instead of alpha5beta1 integrin. *The Journal of Biological Chemistry* 279(18): p.18157–18168.

Zweers,M.C., Bristow,J., et al. ,2003. Haploinsufficiency of TNXB Is Associated with Hypermobility Type of Ehlers-Danlos Syndrome. *American Journal of Human Genetics* 73(1): p.214–217.

Zweers,M.C., Dean,W.B., van Kuppevelt,T.H., Bristow,J., and Schalkwijk,J. ,2005. Elastic fiber abnormalities in hypermobility type Ehlers-Danlos syndrome patients with tenascin-X mutations. *Clinical Genetics* 67(4): p.330–334.

Zweers,M.C., Kucharekova,M., and Schalkwijk,J. ,2005. Tenascin-X: a candidate gene for benign joint hypermobility syndrome and hypermobility type Ehlers-Danlos syndrome? *Annals of the Rheumatic Diseases* 64(3): p.504–505.

Appendix

Appendix A- hEDS diagnostic criteria

Appendix B-Medical questionnaire

Appendix C-Protocols and volumes for preparation of buffers

Appendix D- Theragen sample QC report

Appendix E- TruSeq DNA PCR-free Protocol

Appendix F- TruSeq DNA Nano Protocol

Appendix G-List of analysed genes in gene panels

Appendix H-Sample concentration prior to Sanger

Appendix I-Theragen WGS analysis QC report

Appendix J-Medical questionnaire answer list

Appendix K- Filtered variants in extended coding regions

Appendix L- Annotated gene list

Appendix M-Publication by author

Appendix A- hEDS diagnostic criteria

The International Consortium on Ehlers-Danlos Syndromes & Related Disorders
In Association with The Ehlers-Danlos Society

Diagnostic Criteria for Hypermobile Ehlers-Danlos Syndrome (hEDS)

This diagnostic checklist is for doctors across all disciplines to be able to diagnose EDS

Distributed by The Ehlers Danlos Society

Patient name: _____ DOB: _____ DOV: _____ Evaluator: _____

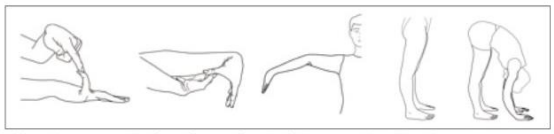
The clinical diagnosis of hypermobile EDS needs the simultaneous presence of all criteria, 1 ***and*** 2 ***and*** 3.

CRITERION 1 – Generalized Joint Hypermobility

One of the following selected:

- ☐ ≥6 pre-pubertal children and adolescents
- ☐ ≥5 pubertal men and woman to age 50
- ☐ ≥4 men and women over the age of 50

Beighton Score: _____/9



If Beighton Score is one point below age- and sex-specific cut off, two or more of the following must also be selected to meet criterion:

- ☐ Can you now (or could you ever) place your hands flat on the floor without bending your knees?
- ☐ Can you now (or could you ever) bend your thumb to touch your forearm?
- ☐ As a child, did you amuse your friends by contorting your body into strange shapes or could you do the splits?
- ☐ As a child or teenager, did your shoulder or kneecap dislocate on more than one occasion?
- ☐ Do you consider yourself "double jointed"?

CRITERION 2 – Two or more of the following features (A, B, or C) must be present

Feature A (five must be present)

- ☐ Unusually soft or velvety skin
- ☐ Mild skin hyperextensibility
- ☐ Unexplained striae distensae or rubae at the back, groins, thighs, breasts and/or abdomen in adolescents, men or pre-pubertal women without a history of significant gain or loss of body fat or weight
- ☐ Bilateral piezogenic papules of the heel
- ☐ Recurrent or multiple abdominal hernia(s)
- ☐ Atrophic scarring involving at least two sites and without the formation of truly papyraceous and/or hemosideric scars as seen in classical EDS
- ☐ Pelvic floor, rectal, and/or uterine prolapse in children, men or nulliparous women without a history of morbid obesity or other known predisposing medical condition
- ☐ Dental crowding and high or narrow palate
- ☐ Arachnodactyly, as defined in one or more of the following:
  (i) positive wrist sign (Walker sign) on both sides, (ii) positive thumb sign (Steinberg sign) on both sides
- ☐ Arm span-to-height ratio ≥1.05
- ☐ Mitral valve prolapse (MVP) mild or greater based on strict echocardiographic criteria
- ☐ Aortic root dilatation with Z-score >+2

Feature A total: _____/12

Feature B

- ☐ Positive family history; one or more first-degree relatives independently meeting the current criteria for hEDS

Feature C (must have at least one)

- ☐ Musculoskeletal pain in two or more limbs, recurring daily for at least 3 months
- ☐ Chronic, widespread pain for ≥3 months
- ☐ Recurrent joint dislocations or frank joint instability, in the absence of trauma

CRITERION 3 – All of the following prerequisites MUST be met

1. Absence of unusual skin fragility, which should prompt consideration of other types of EDS
2. Exclusion of other heritable and acquired connective tissue disorders, including autoimmune rheumatologic conditions. In patients with an acquired CTD (*e.g.* Lupus, Rheumatoid Arthritis, etc.), additional diagnosis of hEDS requires meeting both Features A and B of Criterion 2. Feature C of Criterion 2 (chronic pain and/or instability) cannot be counted toward a diagnosis of hEDS in this situation.
3. Exclusion of alternative diagnoses that may also include joint hypermobility by means of hypotonia and/or connective tissue laxity. Alternative diagnoses and diagnostic categories include, but are not limited to, neuromuscular disorders (*e.g.* Bethlem myopathy), other hereditary disorders of the connective tissue (e.g. other types of EDS, Loeys-Dietz syndrome, Marfan syndrome), and skeletal dysplasias (*e.g.* osteogenesis imperfecta). Exclusion of these considerations may be based upon history, physical examination, and/or molecular genetic testing, as indicated.

Diagnosis: _____

v9

Appendix B-Medical questionnaire

The Genetics of Hypermobile Ehlers-Danlos Syndrome- A local study

Details of Participant:

Assigned participant number:_____

Date of diagnosis:_____ Age:_____ Sex: M / F

Systemic symptoms:

Psychiatric symptoms

Do you suffer from anxiety or any other psychiatric conditions (depression, panic disorder, PTSD)? Yes / No

If yes, please specify the type of condition:

for how long?

Immunological symptoms

Do you suffer from asthma? Yes / No

Do you have any allergies /or eczema? Yes / No

Do you have normal immunoglobulin levels? Yes / No / Don't know

Cardiac symptoms

Do you have high / low blood pressure? Yes / No / Don't know

Do you sometimes feel lightheaded or dizzy when standing up that usually resolves by lying down? Yes / No / Don't know

Do you sometimes feel your heart beating faster than usual? Yes / No / Don't know

Dermatological symptoms

Have you experienced prolonged wound healing? Yes / No / Don't know

Do your wounds leave a flat scar when they heal? Yes / No / Don't know

Do you have thin, stretchy, elastic skin? Yes / No / Don't know

Is your skin very sensitive? Yes / No / Don't know

Do you bruise very easily? Yes / No / Don't know

Do you have small bumps on your heel and / or wrists (Piezogenic papules)? Yes / No

Gynaecological symptoms (only answer this section if you are female. If you are male, proceed to the next section)

Have your periods stopped permanently? Yes / No

If yes, Age at menopause: _____ years

Cycle length (time between periods): _____ days

Is your cycle regular? Yes / No / Don't know

Do you experience heavy or prolonged menstrual bleeding? Yes / No / Don't know

Do you experience menstrual cramps? Yes / No / Don't know

If yes, is the pain slight / moderate / excruciating

Have you ever been diagnosed with endometriosis? Yes / No / Don't know

Gastrointestinal symptoms

Do you frequently experience any of the following digestive problems: *(please tick ✓ where applicable)*	YES	NO
Acid reflux		
Constipation		
Abdominal pain/discomfort		
Diarrhoea		
Nausea and / or vomiting		

Musculoskeletal symptoms

Do you feel exhausted all the time or sometimes you can't seem to think clearly? Yes / No

If yes: does the fatigue start early in the morning or at the end of the day? *(underline correct answer)*

Does your jaw sometimes 'lock', making it difficult to open/ close your mouth? Yes / No

While chewing, do you sometimes experience pain or difficulty? Yes / No

Do you experience chronic pain, lasting more than 3 months? Yes/ No

(Only answer the next question if you replied yes to the previous question, please tick ✓ where applicable)

Site of pain	Pain Severity (0-no pain, 1-slight pain, 5-excruciating pain)					
Jaw	0 □	1 □	2 □	3 □	4 □	5 □
Neck	0 □	1 □	2 □	3 □	4 □	5 □
Back	0 □	1 □	2 □	3 □	4 □	5 □
Shoulders	0 □	1 □	2 □	3 □	4 □	5 □
Elbows	0 □	1 □	2 □	3 □	4 □	5 □
Wrists	0 □	1 □	2 □	3 □	4 □	5 □
Fingers	0 □	1 □	2 □	3 □	4 □	5 □
Hips	0 □	1 □	2 □	3 □	4 □	5 □
Knees	0 □	1 □	2 □	3 □	4 □	5 □

Ankles	0 □	1 □	2 □	3 □	4 □	5 □
Feet	0 □	1 □	2 □	3 □	4 □	5 □

Have you ever dislocated a joint? Yes / No

If yes, how many times _____

and where? _____

Do you have scoliosis/ kyphosis? Yes / No / Don't know

Have you ever had a bone fracture? Yes / No

Functional Limitations:

Do you have any difficulties in: (please tick ✓ where applicable)	YES	NO
Walking (do you stumble or trip frequently?)		
Climbing and descending stairs?		
Movement co-ordination and balance?		
Lifting / carrying things?		
Gripping objects?		
Twisting or bending things?		

General questions

Are you on any medication? Yes / No

If yes, please specify the type of medication you take and frequency of dosage:

_____

Have you ever had any surgical operations? Yes / No

If yes, where and at what age? _____

Have you ever been administered a local anesthetic? Yes / No

If yes, was the dose sufficient to completely eliminate pain / sensation? Yes / No

Lifestyle factors:

What is your occupation, if any? _____
For how long have you been working (both past and current employment)? _____
How many hours do you spend at work, per week? _____hours

Alcohol: (tick ✓ where applicable)

☐ Never drank alcohol

☐ Drink alcohol (Beer, Wine, Spirits) monthly or less ☐ Drink alcohol 3-4 times a week;
If yes: State if Beer / Wine / Spirits

☐ Drink alcohol every day
If yes: State if Beer / Wine / Spirits

During the past year, your alcohol consumption was **Regular / Moderate / Rare**

Smoking: (tick ✓ where applicable)

- ☐ Never smoked
- ☐ Current smoker (no. of cigarettes daily _____)
- ☐ Former smoker (started at age_____and stopped at age_____)
- ☐ Immediate family members/ workmates/ friends smoked indoors in your presence regularly

Physical activity: (tick ✓ where applicable)

☐ Sedentary lifestyle

☐ Regular light to moderate exercise
How many hours of exercise do you spend per week? _____ ☐ Regularly play a sport
What type of sport? _____
How many hours of exercise do you spend per week? _____

Diet:

Is your diet high in red meat or animal protein? Yes / No

Do you include dairy products frequently in your diet? Yes / No

Are green vegetables included in your diet? Yes / No

Do you drink more than 3 caffeinated drinks per day (coffee, tea, soda)? Yes / No

If yes: State which: ____

Supplements:

Do you take calcium, Vitamin D, Vitamin C or cod liver oil supplements? Yes / No

If yes: How many per day: __ How long have you been taking them for: __

Family history:

Does your family have a history of hypermobility? Yes / No

If yes, tick ✓ where applicable:

Mother	
Father	
Siblings	
Maternal grandmother	
Maternal grandfather	
Paternal grandmother	
Paternal grandfather	
Maternal aunts/ uncles	
Paternal aunts/ uncles	
Maternal cousins	
Paternal cousins	

Questions Related to Blood Testing:

Have you had a blood transfusion within the last three months? Yes / No / Don't know

Have you ever had a bone marrow transplant? Yes / No / Don't know When and at what time did you last eat?_____

Females: Are you currently menstruating? Yes / No

If no, when was your last day of menstruation? _____ days / Don't know

Thank you for your patience!

Date of Interview: _ _ / _ _ / _ _ _ _

Interviewer Name: _____

Interviewer Signature: _____

Appendix C-Protocols and volumes for preparation of buffers

1L 10x erythrocyte lysing buffer (ELB) pH 7.4:

Reagents	Final concentration	Molecular weight	Weight for 1L of 10x ELB
NH_4Cl	155mM	53.49 g/L	82.91 g
$KHCO_3$	10mM	100.12 g/L	10.01 g
Na_2EDTA	0.1mM	372.24 g/L	0.37 g

-Add 500 ml of deionized water, pH it with HCl or NaOH accordingly and top up to 1 L with deionized water. Autoclave.

Note: Solution is endothermic, let buffer warm up to room temperature before adjusting pH

-Dilute 100 ml of 10x ELB with 900ml of deionized water to get 1x ELB. Autoclave.

1L 10x SE buffer pH 8.0:

Reagents	Final concentration	Molecular weight	Weight for 1L of 10x SE
NaCl	75mM	58.44 g/L	43.83 g
Na_2EDTA	25mM	372.24 g/L	93.06 g

-Add 500 ml of deionized water, pH it with HCl or NaOH accordingly and top up to 1 L with deionized water. Autoclave.

Note: Buffer pH needs to be brought to 8 before the salts will dissolve; requires a substantial amount of concentrated NaOH (which is prepared as needed from pellets)

-Dilute 100 ml of 10x SE buffer with 900 ml of deionized water to get 1x SE buffer. Autoclave.

1L 10% SDS:

-Add 100g of SDS to 800mls of deionized water, mix and top up to 1L with deionized water.

Note: Cannot be autoclaved. Prepare in small volumes in sterile water.

1L of 1x TE buffer pH 7.4:

Reagents	Final concentration	Molecular weight	Weight for 1L of 1x TE
Tris-HCl	10mM	157.60 g/L	1.576 g
Na_2EDTA	1mM	372.24 g/L	0.372 g

-Add 500 ml of deionized water, pH it with HCl or NaOH accordingly and top up to 1L with deionized water. Autoclave the stock and prepare small volume aliquots which should also be autoclaved before use.

1L 6M NaCl:

Reagents	Final concentration	Molecular weight	Weight for 1L of 6M NaCl
NaCl	6M	56.44 g/L	350.64 g

-Top up to 1 L with deionized water. Divide into aliquots and autoclave.

Proteinase K (Promega) stock 100mg/L

Reagents	Final concentration	Molecular weight	Weight for 10ml reconstitution sol.
Tris-HCl	50mM	157.6 g/L	0.0788 g
$CaCl_2.2H_2O$	10mM	147.02 g/L	0.0147 g

-Reconstitute in 50 mM Tris-HCl, pH 8.0 and 10 mM $CaCl_2$ to a working solution of 20 mg/ml

-Top up to 10 ml with deionized water.

TotalOmics
Sequencing for Everyone

Appendix D-Theragen sample QC report

1. Service Information

C1	Customer	Kirsty Scicluna	Department	Applied Biomedical Science
Company or Institute		University of Malta		
Project		TBO211410		
S1	Service	Whole Genome Resequencing		
Experiment started on		15 November 2021	Experiment finished on	15 November 2021
Lab technician		****** Kim Signature	Lab supervisor	****** Lim Signature

2. Summary

	Service	Number of samples	Pass	Hold	Fail
S1	Whole Genome Resequencing	**3**	**1**	**2**	**0**

The experiment summary is shown above. Based on QC results, each sample is sorted into one of three categories: pass, hold, or fail. If a sample satisfies all criteria in Sample Requirements section, it is called *pass*ed, and if a sample satisfies no criterion at all, it is called *fail*ed. Otherwise it is said to be on *hold*. In Result Details section you may find the outcomes of your samples.

The passed samples proceed to the library construction stage automatically. For the on-hold or failed samples, however, human intervention is necessary. For on-hold samples your explicit permission is required to construct libraries with them. For failed samples please be advised to send replacement samples for library construction. In either case one of our staff members dedicated to this service will contact you to discuss the matter if it happens.

The rest of this report is organized as follows. *Result Details* section lists all the samples used in this experiment along with scores and values important to make a decision to proceed to the next stage. As a result of such decision, you may find out which samples are passed, failed, or on hold. In *Gel electrophoresis* or *Electropherograms* section, you can review a visualized in-depth quality of each sample. *QC Method Description* section briefly explains the methods used in our laboratory. In *Sample Requirements* section all the requirements each sample should meet are listed. You can find contact information of our institute and staff in the last section.

3. Result Details

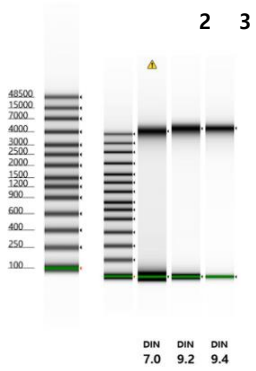
No	Sample or Kit	Result	Concentration (ng/μL)	Volume (μL)	Quantity (μg)	Purity		DIN Value
						260/280	260/230	
S1	**Whole Genome Resequencing — TruSeq DNA PCR-Free Sample Prep Kit (350 bp)**							
	Sample requirements		≥ 30	N/A	≥ 1	1.5~2.2	N/A	≥ 6
1	06-294CG Eukaryote TN2111D1131(DNA)	hold	8.10	65	0.526	1.44	0.63	7.0
2	03-486CZ Eukaryote TN2111D1132(DNA)	hold	15.20	41	0.623	1.56	0.73	9.2
3	SDT-005 Eukaryote TN2111D1133(DNA)	pass	40.30	70	2.821	1.70	1.14	9.4

l The values in *red*, if any, are out of the acceptable range specified in the sample requirements.

≥ The definition of pass, hold, and fail are in Summary section.

4. Genomic DNA ScreenTape

LDR1



No	Sample	DIN Value	No	Sample	DIN Value
1	06-294CG Eukaryote TN2111D1131	7	3	SDT-005 Eukaryote TN2111D1133	9.4
2	03-486CZ Eukaryote TN2111D1132	9.2			

⚠ : Sample concentration outside functional range for DIN and the assay (5~300ng/ul)
: Sample concentration outside recommended range (10~100ng/ul)

5. QC Method Description

Theragen Bio is checking sample quality using two methods; Spectrophotometer and Automated Electrophoresis. Please refer to detailed methods at below;

Spectrophotometer (Infinium F-200, NanoDrop) to assess sample purity

Absorbance measurement is simple, and requires commonly available laboratory equipment. To evaluate DNA purity, measure absorbance from 230 nm to 340 nm to detect other possible contaminants. The most common purity calculation is the ratio of the absorbance at 260 nm divided by the reading at 280 nm. Good-quality DNA will have an A260/A280 ratio of between 1.6 and 2.0. The A260/A230 is best if greater than 1.6.

Quality check with Automated Electrophoresis(4200 Tapestation) to assess sample integrity and concentration We

check DNA integrity using Agilent 4200 Tapestation with DNA Integrity Number (DIN) value greater than 6.

6. Customer Support

We, Theragen Bio, are committed to provide accurate, reliable, and timely services. As such we have dedicated staff members ready to answer your questions or inquiries regarding to the service we provide here. In the following you may find the contact information of these members assigned just for you. Please mention your project number **TBO211410** when you contact us so that your service record can be retrieved promptly.

A1	Name	Boryung Kim	Department	Genomic Division
Cell phone		+82-31-288-1222	Email	**boryung.kim@theragenbio.com**
A2	Name	Jiyeon Kim	Department	Genomic Division
Cell phone		+82-31-288-1222	Email	**jiyeon.kim@theragenbio.com**
Department		NGS Service Support		
Telephone		+82-31-888-9447	Email	**ngscs@theragenbio.com**

For general inquiries on TotalOmics services, please contact us using the following:

Home page	**www.theragenbio.com/en**
Genomic Division	**dept.ngsbusiness@theragenbio.com**
NGS Service Support	**ngscs@theragenbio.com**
Mailing address	Theragen Bio Co., Ltd. 4th Fl. Korea Bio Park Bldg. C, 700, Daewangpangyo-ro, Bundang-gu, Seongnam-si, Gyeonggi-do, Republic of Korea 13488
Business hours	9:00 AM to 6:00 PM on Monday thru Friday Closed on Saturday, Sunday, and holidays

Appendix E- TruSeq DNA PCR-free Protocol

Prepare Library | Sequence | Analyze Data

illumina®

TruSeq™ DNA PCR-Free

Simple, streamlined whole-genome sequencing library preparation that provides accurate and comprehensive coverage of complex genomes.

Highlights

- Streamlined Library Preparation
  PCR-free protocol accelerates the TruSeq DNA library preparation workflow
- Excellent Coverage Quality
  Very low library bias and fewer gaps in coverage enable deep insight into the genome
- High Flexibility
  Optimized PCR-free workflows support various read lengths and applications
- Inclusive Solution
  Reliable solution includes master-mixed reagents, size-selection beads, and up to 96 unique dual indexes (UDIs)

Introduction

TruSeq DNA PCR-Free offers numerous enhancements to the widely adopted TruSeq DNA workflow, providing optimized library preparation for whole-genome sequencing (WGS) applications. By eliminating PCR amplification steps, the PCR-free protocol significantly reduces typical PCR-induced bias and provides detailed sequence information for traditionally challenging regions of the genome. Low-throughput and high-throughput protocols are available to accommodate a range of study designs (Figure 1).

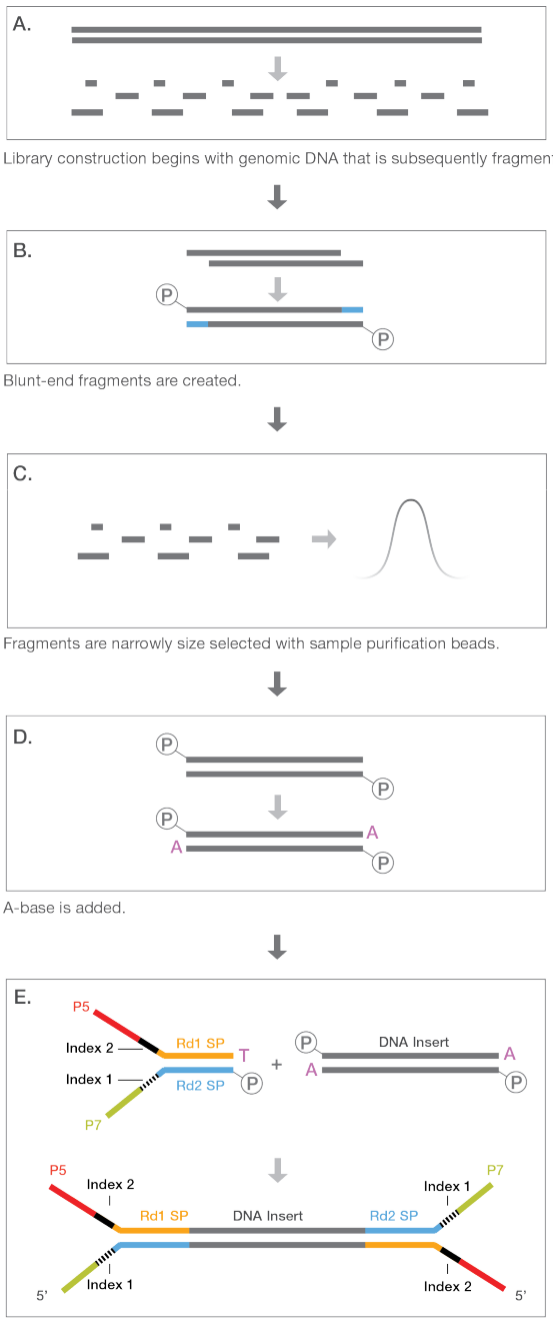
Accelerated Library Preparation

The TruSeq DNA library preparation workflow has been streamlined by removing the PCR amplification step and replacing gel-based size selection with bead-based selection (Figure 2). This workflow offers exceptional flexibility with two protocol options for generating either large (550 bp) or small (350 bp) insert sizes to support various applications. Master-mixed reagents, provided sample purification beads for clean up and size selection, robust TruSeq indexes, and optimized protocols contribute to the simplified library preparation workflow, requiring a low number of cleanup steps for processing large sample numbers. TruSeq DNA PCR-Free decreases library preparation time, enabling researchers to perform applications from microbial sequencing to human WGS.[2]

Figure 1: TruSeq DNA PCR-Free—TruSeq DNA PCR-Free offers an efficient solution for preparing and indexing sample libraries. TruSeq DNA PCR-Free accommodates up to 24 indexes for low-throughput studies, or up to either 96 dual indexes or 96 unique dual indexes (sold separately) for high-throughput studies.

Innovative Library Preparation Chemistry

TruSeq DNA PCR-Free can be used to prepare DNA libraries for single-read, paired-end, and indexed sequencing. TruSeq DNA PCR-Free supports shearing by Covaris ultrasonication, requiring 1 µg of input DNA for an average insert size of 350 bp or 2 µg DNA for an average insert size of 550 bp. Library construction begins with fragmented genomic DNA (gDNA) (Figure 2A). Blunt-end DNA fragments are generated using a combination of fill-in reactions and exonuclease activity (Figure 2B), and size selection is performed with provided sample purification beads (Figure 2C). An A-base is then added to the blunt ends of each strand, preparing them for ligation to the indexed adapters (Figure 2D). Each adapter contains a T-base overhang for ligating the adapter to the A-tailed fragmented DNA. These adapters contain the full complement of sequencing primer hybridization sites for single, paired-end, and indexed reads. With no need for additional PCR amplification, single or dual-index adapters are ligated to the fragments and products are ready for cluster generation (Figure 2E).



Dual-index adapters are ligated to the fragments* and final product is ready for cluster generation.

Figure 2: TruSeq DNA PCR-Free Workflow—The TruSeq DNA PCR-Free workflow features adapter ligation resulting in sequence-ready products without PCR amplification. *The TruSeq DNA PCR-Free LT indexing solution features a single-index adapter at Step E.

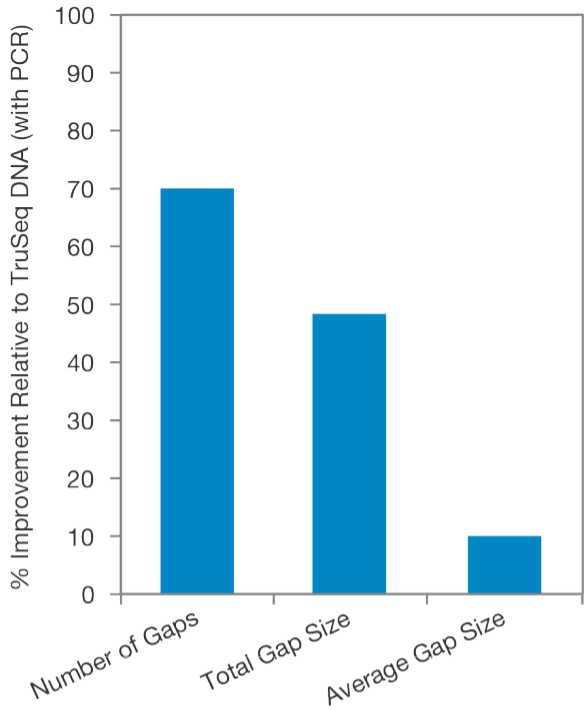


Figure 3: Fewer Gaps in Coverage—TruSeq DNA PCR-Free libraries show significant reduction in the number and total size of gaps when compared to libraries prepared using the TruSeq DNA (with PCR) protocol. A gap is defined as a region ≥ 10 bp in length, where an accurate genotype cannot be determined due to low depth, low alignment scores, or low base quality.

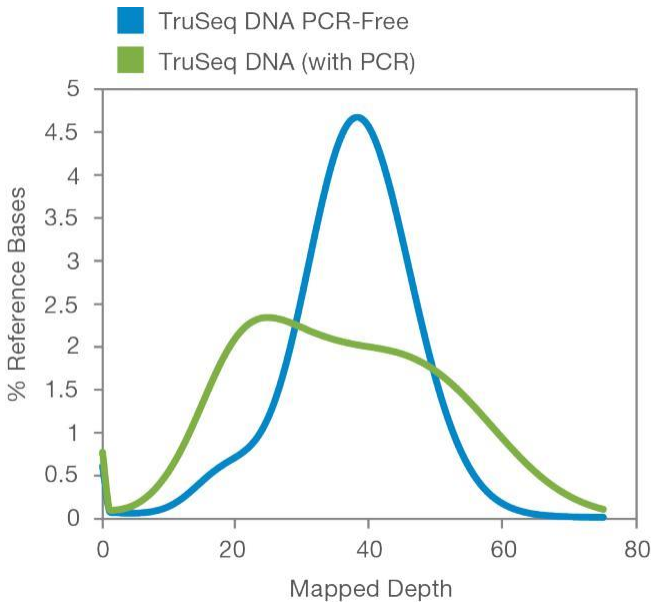


Figure 4: Greater Coverage Uniformity—TruSeq DNA PCR-Free libraries provide greater coverage uniformity across the genome when compared to those generated using the TruSeq DNA (with PCR) protocol.

Excellent Coverage Quality

TruSeq DNA PCR-Free reduces the number and average size of typical PCR-induced gaps in coverage (Figure 3), delivering exceptional data quality. The removal of PCR amplification from the TruSeq DNA PCR-Free workflow reduces library bias[2] and improves coverage uniformity across the genome (Figure 4). This workflow also provides excellent coverage of traditionally challenging genomic content, including GC-rich regions, promoters, and repetitive regions (Figure 5). High data quality delivers base-pair resolution, providing a detailed view of somatic and *de novo* mutations and supporting accurate identification of causative variants. TruSeq DNA PCR-Free provides a comprehensive view of the genome, including coding, regulatory, and intronic regions, enabling researchers to access more information from each sequencing run (Figure 6).

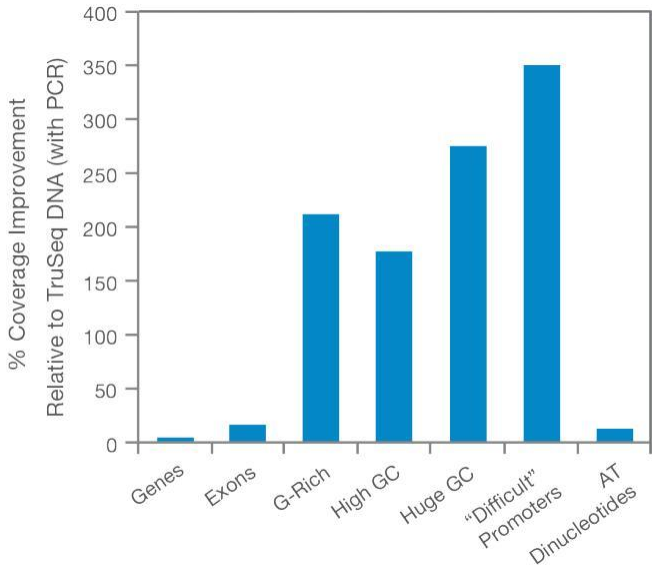


Figure 5: Increased Coverage of Challenging Regions—TruSeq DNA PCR-Free libraries demonstrate improved coverage of challenging genomic content. These regions include known human protein coding and nonprotein coding exons and genes defined in the RefSeq Genes track in the UCSC Genome Browser. G-Rich regions denote 30 bases with ≥ 80% G. High GC regions are defined as 100 bases with ≥ 75% GC content. Huge GC regions are defined as 100 bases with

≥ 85% GC content. "Difficult" Promoters denote the set of 100 promoter regions that are insufficiently covered, which have been empirically defined by the Broad Institute of MIT and Harvard. AT Dinucleotides indicate 30 bases of repeated AT dinucleotides.

Efficient Sample Multiplexing

Indexes are added to sample gDNA fragments using a simple PCR-free procedure to provide an innovative solution for sample multiplexing. For the greatest operational efficiency, up to 96 preplated, uniquely indexed samples can be pooled and sequenced together in a single flow cell lane on any Illumina sequencing platform. After sequencing, the indexes are used to demultiplex the data and accurately assign reads to the proper samples in the pool. TruSeq DNA PCR-Free can use a single indexing strategy or a dual-indexing strategy that uses a unique combination of two indexes to demultiplex. The unique dual index (UDI) adapters (available separately) were developed in a collaboration between Integrated DNA Technologies, Inc. (IDT) and Illumina to employ unique pairs to demultiplex.

The TruSeq DNA Single Indexes contain up to 24 indexes with two sets of 12 each, and the TruSeq DNA CD Indexes contain 96 indexes. Multisample studies can be conveniently managed using the Illumina Experiment Manager, a freely available software tool that provides easy reaction set up for plate-based processing. It allows researchers to configure the index sample sheet (ie, sample multiplexing matrix) for the instrument run, enabling automatic demultiplexing.

To learn more about Illumina Experiment Manager, visit www.illumina.com/informatics/research/experimental-design/illumina-experiment-manager.html.

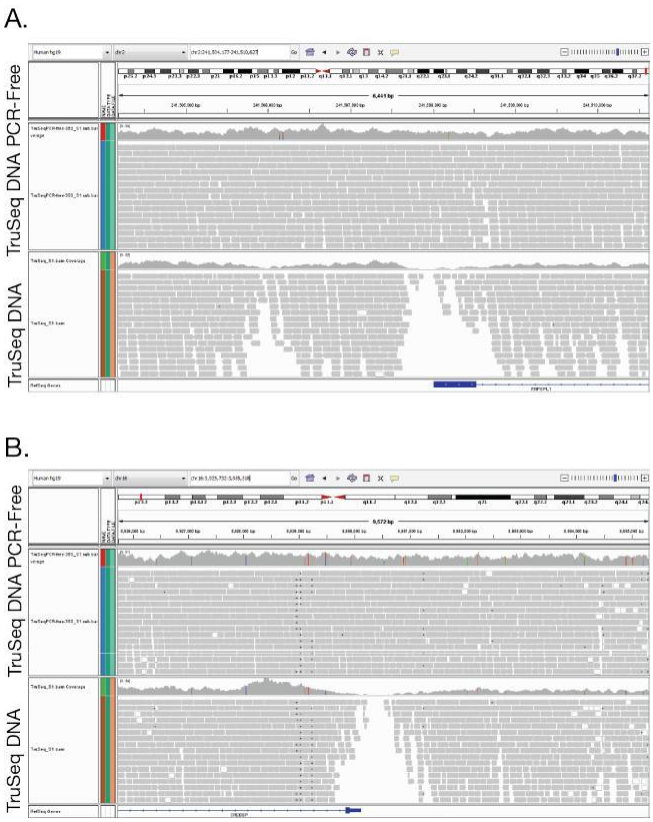


Figure 6: TruSeq DNA PCR-Free Reduces Number of Coverage Gaps—Increased coverage of TruSeq DNA PCR-Free libraries results in fewer coverage gaps, demonstrated here in the GC-rich coding regions of the (A) *RNPEPLI1* promoter and the (B) *CREBBP* promoter. Sequence information generated by TruSeq DNA PCR-Free is shown in the top panels of (A) and (B), while sequence data generated using TruSeq DNA protocol are shown in the lower panels.

Table 1: TruSeq DNA Library Preparation

Specification	TruSeq DNA Nano	TruSeq DNA PCR-Free	TruSeq DNA
Description	Based on widely adopted TruSeq library prep, with lower input and improved data quality	Excellent genomic coverage with radically reduced library bias and gaps	Original TruSeq next-generation sequencing library preparation method
Input Quantity	100–200 ng	1–2 µg	1 µg
Includes PCR	Yes	No	Yes
Assay Time	~ 6 hours	~ 5 hours	1–2 days
Hands-On Time	~ 5 hours	~ 4 hours	~ 8 hours
Target Insert Size	350 bp or 550 bp	350 bp or 550 bp	300 bp
Gel-Free	Yes	Yes	No
Number of Samples Supported	24 (LT) or 96 (HT)[a]	24 (LT) or 96 (HT)[a]	48 (LT) or 96 (HT)[a]
Supports Enrichment	No[b]	No[b]	Yes
Size-Selection Beads	Included	Included	Not Included
Applications	WGS, including whole-genome resequencing, *de novo* assembly, and metagenomics studies		
Sample Multiplexing	24 single indexes, 96 combinatorial dual indexes, 24 and 96 unique dual indexes (UDIs)		
Compatible Illumina Sequencing Systems	HiSeq™, HiScanSQ™, Genome Analyzer, and MiSeq™, and MiniSeq™ Systems		

a. LT, low-throughput; HT, high-throughput.

b. Nextera™ Rapid Capture products support various enrichment applications. For more information, visit www.illumina.com/NRC.

Flexible and Inclusive Library Preparation

The TruSeq family of library preparation solutions offers several options for sequencing applications, compatible with a range of research needs and study designs (Table 1). All TruSeq products support high- and low-throughput studies. These workflows provide numerous enhancements to the TruSeq DNA library preparation method, empowering various sequencing applications. Library prep reagents and sequencing indexes are now offered separately, enabling researchers to tailor these workflows to their experimental needs.

Streamlined Solution

TruSeq DNA PCR-Free contains library preparation reagents, sample purification beads, and robust TruSeq indexes for multiplexing, providing a comprehensive preparation method optimized for high performance on all Illumina sequencing platforms. TruSeq DNA PCR-Free offers the flexibility of two options, 24-sample and 96-sample, for scalable experimental design. With a simplified protocol and flexible multiplexing options, TruSeq DNA PCR-Free offers a streamlined library preparation method that delivers high-quality sequencing data.

Summary

TruSeq DNA PCR-Free optimizes the TruSeq workflow to deliver a streamlined library preparation method for any sequencing application. Low- and high-throughput options and varied insert sizes provide greater flexibility to support various applications and genomic studies. Workflow innovations reduce PCR-induced bias to facilitate detailed and accurate insight into the genome. By harnessing a faster workflow and enhanced data quality, TruSeq DNA PCR-Free provides a comprehensive sample preparation method for genome sequencing applications.

Learn More

To learn more about TruSeq DNA PCR-Free, visit www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/truseq-dna-pcr-free.html

References

1. Saunders CJ, Miller NA, Soden SE, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. *Sci Translational Med*. 2012;4(154):154ra135.

2. Aird D, Ross MG, Chen WS, et al. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol*. 2011;12:R18.

3. University of California, Santa Cruz (UCSC) Genome Browser. genome.ucsc.edu. Accessed July 2013.

4. The Broad Institute of MIT and Harvard. www.broadinstitute.org. Accessed July 2013.

Ordering Information

Product	Catalog No.
TruSeq DNA PCR-Free Library Prep 24 samples	20015962
TruSeq DNA PCR-Free Library Prep 96 samples	20015963
TruSeq DNA Single Indexes Set A	20015960
TruSeq DNA Single Indexes Set B	20015961
TruSeq DNA CD Indexes	20015949
IDT for Illumina–TruSeq DNA UD Indexes (24 indexes, 96 samples)	20020590
IDT for Illumina–TruSeq DNA UD Indexes (96 indexes, 96 samples)	20022370

Appendix F- TruSeq DNA Nano Protocol

TruSeq™ DNA Nano

illumina®

A low-input method that delivers a high-confidence, comprehensive view of the genome for virtually any sequencing application.

Highlights

- **Low sample input**
  Excellent data quality from as little as 100 ng input empowers interrogation of samples with limited available DNA
- **Excellent coverage quality**
  Very low library bias and fewer gaps in coverage enable deep insight into the genome
- **High flexibility**
  Streamlined workflow enables library preparation in less than one day, while supporting various read lengths
- **Inclusive solution**
  Reliable solution includes master-mixed reagents, size-selection beads, and up to 96 unique dual indexes (UDIs)

Figure 1: TruSeq DNA Nano—TruSeq DNA Nano offers a low-input solution for preparing and indexing sample libraries. TruSeq DNA Nano accommodates up to 24 indexes for low-throughput studies, or up to either 96 dual indexes or 96 unique dual indexes (sold separately) for high-throughput studies.

Introduction

By offering a low-input method based on the widely adopted TruSeq library preparation workflow, TruSeq DNA Nano enables efficient interrogation of samples that have limited available DNA. This workflow significantly reduces typical PCR-induced bias and provides detailed sequence information for traditionally challenging regions of the genome. Low-throughput and high-throughput protocols are available to accommodate a range of study designs (Figure 1).

Low Sample Input

TruSeq DNA Nano reduces the typical requirement for micrograms of DNA, enabling researchers to study samples with limited available DNA (eg, tumor samples) and supporting preservation of samples for use in future or alternate studies. This workflow offers the flexibility of two protocols for generating large (550 bp) or small (350 bp) insert sizes to support a diverse range of applications. In addition to accelerating the workflow, simple bead-based size selection avoids typical sample loss associated with gel-based selection. TruSeq DNA Nano is validated for high-quality genomic coverage for a wide range of Illumina whole-genome sequencing (WGS) applications.

Accelerated Library Preparation

The TruSeq DNA library preparation protocol has been streamlined by replacing gel-based size selection with bead-based selection (Figure 2), enabling researchers to prepare high-quality libraries in less than a day. TruSeq DNA Nano is optimized for various read lengths from 2 × 101 bp to 2 × 151 bp. Master-mixed reagents,

Innovative Library Preparation Chemistry

TruSeq DNA Nano can be used to prepare DNA libraries for single-read, paired-end, and indexed sequencing. TruSeq DNA Nano supports shearing by Covaris ultrasonication, requiring 100 ng of input DNA for an average insert size of 350 bp or 200 ng DNA for an average insert size of 550 bp. Library construction begins with fragmented genomic DNA (gDNA) (Figure 2A). Blunt-end DNA fragments are generated using a combination of fill-in reactions and exonuclease activity (Figure 2B), and size selection is performed with provided sample purification beads (Figure 2C). An A-base is then added to the blunt ends of each strand, preparing them for ligation to the indexed adapters (Figure 2D). Each adapter contains a T-base overhang for ligating the adapter to the A-tailed fragmented DNA. These adapters contain the full complement of sequencing primer hybridization sites for single, paired-end, and indexed reads. Single- or dual-index adapters are ligated to the fragments (Figure 2E) and the ligated products are amplified with reduced-bias PCR (Figure 2F)

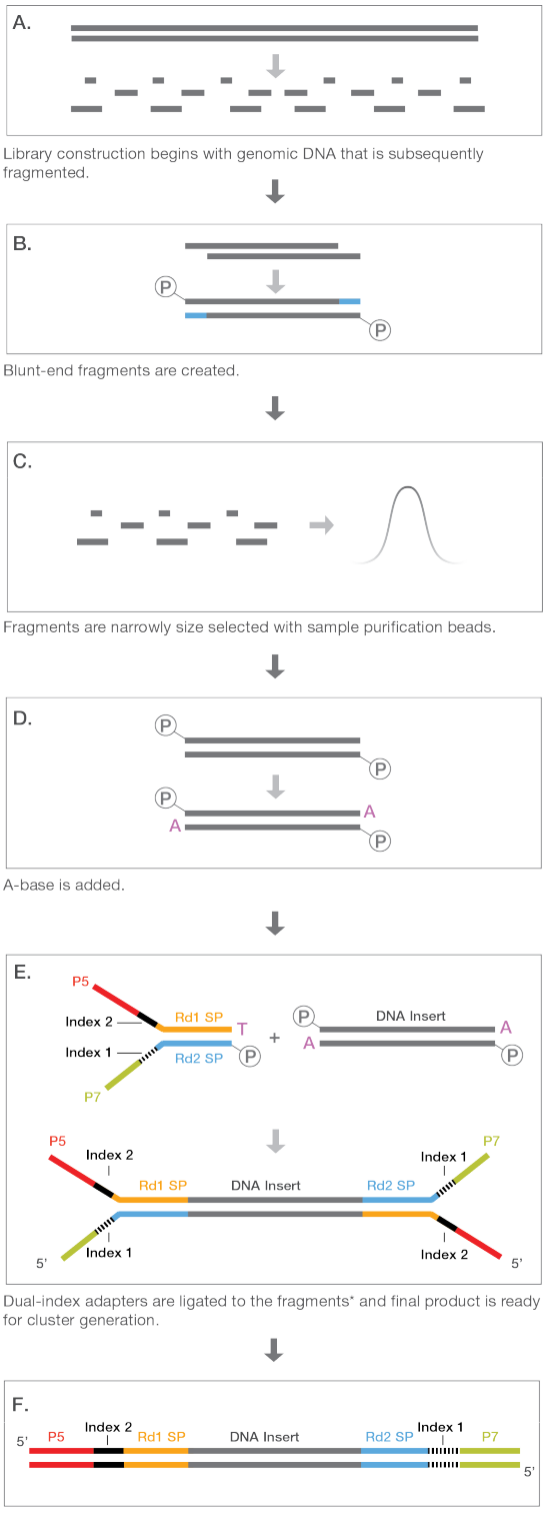


Figure 2: TruSeq DNA Nano Workflow—The TruSeq DNA Nano workflow features adapter ligation resulting in sequence-ready products with minimal PCR-induced bias. *The TruSeq DNA Nano LT indexing solution features a single-index adapter at Step E.

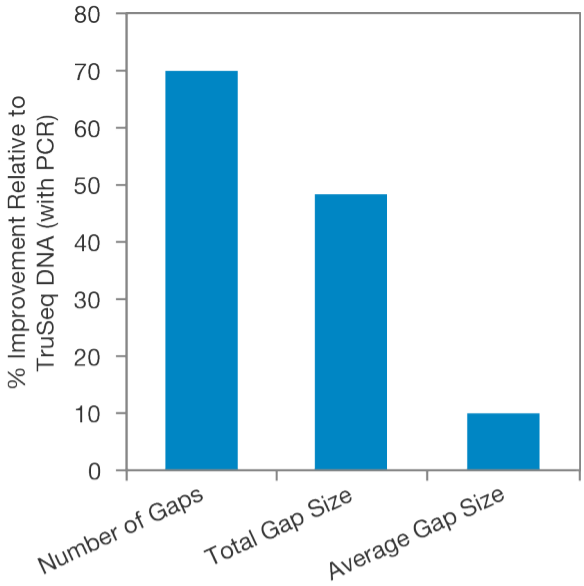


Figure 3: Fewer Gaps in Coverage—TruSeq DNA Nano libraries show significant reduction in the number and total size of gaps when compared to libraries prepared using the TruSeq DNA (with PCR) protocol. A gap is defined as a region ≥ 10 bp in length, where an accurate genotype cannot be determined due to low depth, low alignment scores, or low base quality.

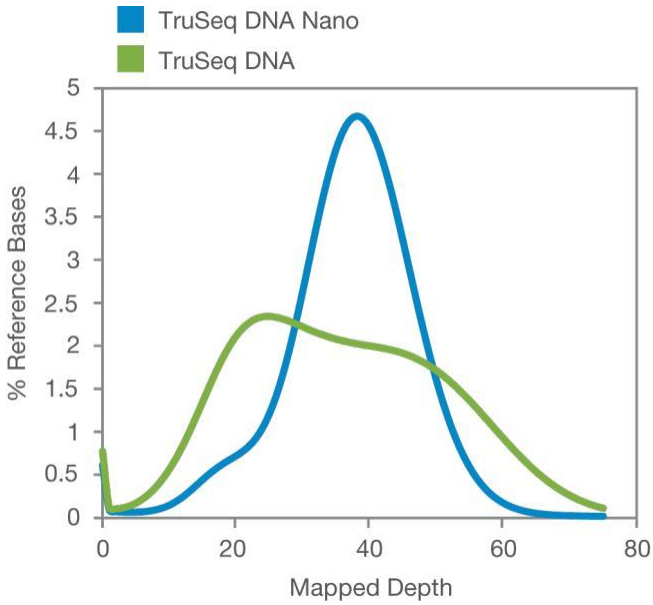


Figure 4: Greater Coverage Uniformity—TruSeq DNA Nano libraries provide greater coverage uniformity across the genome when compared to those generated using the TruSeq DNA (with PCR) protocol.

Table 1: TruSeq DNA Library Preparation

Specification	TruSeq DNA Nano	TruSeq DNA PCR-Free	Legacy TruSeq DNA
Description	Based on widely adopted TruSeq library prep, with lower input and improved data quality	Excellent genomic coverage with radically reduced library bias and gaps	Original TruSeq next-generation sequencing library preparation method
Input Quantity	100–200 ng	1–2 µg	1 µg
Includes PCR	Yes	No	Yes
Assay Time	~ 6 hours	~ 5 hours	1–2 days
Hands-On Time	~ 5 hours	~ 4 hours	~ 8 hours
Target Insert Size	350 bp or 550 bp	350 bp or 550 bp	300 bp
Gel-Free	Yes	Yes	No
Number of Samples Supported	24 (LT) or 96 (HT)[a]	24 (LT) or 96 (HT)[a]	48 (LT) or 96 (HT)[a]
Size-Selection Beads	Included	Included	Not Included
Applications	WGS applications, including whole-genome resequencing, *de novo* assembly, and metagenomics studies		
Sample Multiplexing	24 single indexes, 96 combinatorial dual indexes, 24 and 96 unique dual indexes (UDIs), (available soon)		
Compatible Illumina Sequencing Systems	HiSeq™ 1000, HiSeq 1500, HiSeq 2000, HiSeq 2500, HiSeq 3000, HiSeq 4000, HiSeq X Five. HiSeq X Ten, MiniSeq™, MiSeq™, MiSeqDx in research mode, NextSeq™ 500, NextSeq 550, NextSeq 1000, NextSeq 2000, and NovaSeq™ 6000 Systems		HiSeq, HiScanSQ, Genome Analyzer, and MiSeq, and MiniSeq Systems

a. LT, low-throughput; HT, high-throughput

Excellent Coverage Quality

TruSeq DNA Nano reduces the number and average size of typical PCR-induced gaps in coverage (Figure 3), delivering exceptional data quality. The enhanced workflow reduces library bias and improves coverage uniformity across the genome (Figure 4). It also provides excellent coverage of traditionally challenging genomic content, including GC-rich regions, promoters, and repetitive regions (Figure 5). High data quality delivers base-pair resolution, providing a detailed view of somatic and *de novo* mutations and supporting characterization of causative variants. TruSeq DNA Nano provides a comprehensive view of the genome, including coding, regulatory, and intronic regions, enabling researchers to access more information from each sequencing run (Figure 6).

Flexible and Inclusive Library Preparation

The TruSeq family of library preparation solutions offers several workflows for sequencing applications, compatible with a range of research needs and study designs (Table 1). All TruSeq workflows support high- and low-throughput studies. TruSeq DNA Nano supports WGS and is ideal for sequencing applications that require sparsely available DNA. These workflows provide numerous enhancements to the TruSeq DNA library preparation method, empowering various sequencing applications. Library prep reagents and sequencing indexes are now offered separately, enabling researchers to tailor these workflows to their experimental needs.

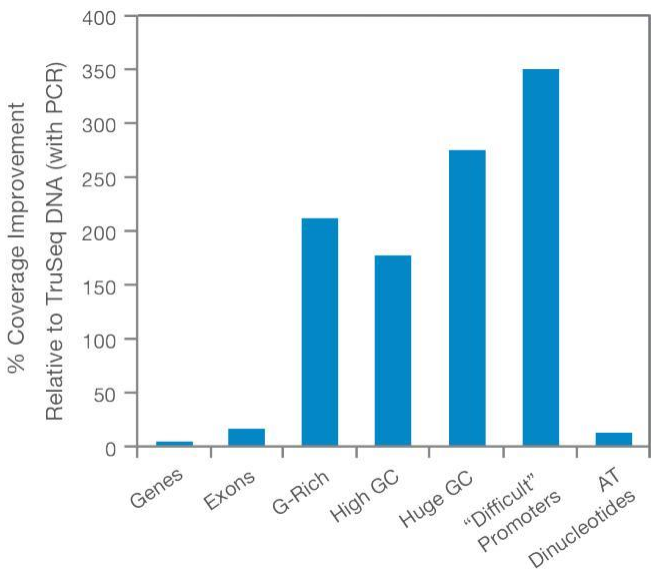


Figure 5: Increased Coverage of Challenging Regions—TruSeq DNA Nano libraries demonstrate improved coverage of challenging genomic content. These regions include known human protein coding and nonprotein coding exons and genes defined in the RefSeq Genes track in the UCSC Genome Browser.[1] G-Rich regions denote 30 bases with ≥ 80% G. High GC regions are defined as 100 bases with ≥ 75% GC content. Huge GC regions are defined as 100 bases with ≥ 85% GC content. "Difficult" Promoters denote the set of 100 promoter regions that are insufficiently covered, which have been empirically defined by the Broad Institute of MIT and Harvard.[2] AT Dinucleotides indicate 30 bases of repeated AT dinucleotides.

Efficient Sample Multiplexing

Using a simple procedure, indexes are added to sample gDNA fragments to provide an innovative solution for sample multiplexing. For the greatest operational efficiency, up to 96 preplated, uniquely indexed samples can be pooled and sequenced together in a single flow cell lane on any Illumina sequencing platform. After sequencing, the indexes are used to demultiplex the data and accurately assign reads to the proper samples in the pool.

TruSeq DNA Nano can use a single indexing strategy or a dual-indexing strategy that uses a unique combination of two indexes to demultiplex. The unique dual index (UDI) adapters were developed in a collaboration between Integrated DNA Technologies, Inc. (IDT) and Illumina to employ unique pairs of indexes to demultiplex. The newly introduced UDIs (24 and 96, available separately) offer increased plexity that enables accurate assignment of reads and efficient use of flow cells. Using UDI combinations is a best practice to make sure that reads with incorrect indexes do not impact variant calling.

Streamlined S olution

TruSeq DNA Nano contains library preparation reagents, sample purification beads, and robust TruSeq indexes for multiplexing, providing a comprehensive preparation method optimized for high performance on all Illumina sequencing platforms. TruSeq DNA Nano uses the flexibility of two options, 24-sample and 96-sample, for scalable experimental design. With a simplified protocol and flexible multiplexing options, TruSeq DNA Nano offers a streamlined library preparation method that delivers high-quality sequencing data.

Summary

TruSeq DNA Nano optimizes the TruSeq protocol to deliver a low-input library preparation method for virtually any sequencing application. Low- and high-throughput options and varied insert sizes provide greater flexibility to support various applications and genomic studies. Workflow innovations reduce PCR-induced bias to facilitate detailed and accurate insights into the genome. By harnessing a faster workflow and enhanced data quality, TruSeq DNA Nano provides a comprehensive sample preparation method for genome sequencing applications.

Learn More

To learn more about TruSeq DNA Nano, visit www.illumina.com/products/truseq-nano-dna-sample-prep-kits.html

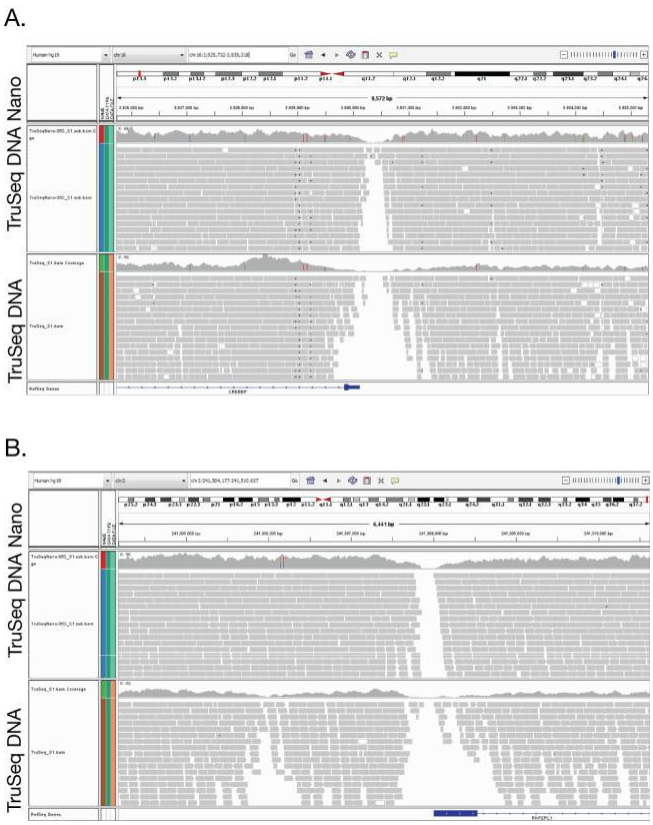


Figure 6: TruSeq DNA Nano Reduces Number of Coverage Gaps— Increased coverage of TruSeq DNA Nano libraries results in fewer coverage gaps, demonstrated here in the GC-rich coding regions of the (A) *RNPEPL1* promoter and (B) *ZBTB34* promoter. Sequence information generated by TruSeq DNA Nano prep is shown in the top panels of (A) and (B), while sequence data generated using TruSeq DNA protocol are shown in the lower panels.

Ordering Information

Product	Catalog No.
TruSeq DNA Nano Low Throughput Library Prep Kit (24 samples)	20015964
TruSeq DNA Nano High Throughput Library Prep Kit (96 samples)	20015965
TruSeq DNA Single Indexes Set A (12 indexes, 24 samples)	20015960
TruSeq DNA Single Indexes Set B (12 indexes, 24 samples)	20015961
TruSeq DNA CD Indexes 96 samples	20015949
IDT for Illumina–TruSeq DNA UD Indexes (24 indexes, 96 samples)	20020590
IDT for Illumina–TruSeq DNA UD Indexes (96 indexes, 96 samples)	20022370

References

1. University of California, Santa Cruz (UCSC) Genome Browser. genome.ucsc.edu. Accessed July 2013.
2. The Broad Institute of MIT and Harvard. www.broadinstitute.org. Accessed July 2013.

Illumina, Inc. • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

Appendix G-List of analysed genes in gene panels

Panel containing genes related to EDS/hEDS:

Gene	Gene Ensembl ID	Chromosome/scaffold name
ADAMTS2	ENSG00000283802	CHR_HG30_PATCH
ADAMTS2	ENSG00000087116	5
B3GALT6	ENSG00000176022	1
B4GALT7	ENSG00000027847	5
C1R	ENSG00000288512	CHR_HG1398_PATCH
C1R	ENSG00000159403	12
C1S	ENSG00000182326	12
CHST14	ENSG00000169105	15
COL12A1	ENSG00000111799	6
COL1A1	ENSG00000108821	17
COL1A2	ENSG00000164692	7
COL3A1	ENSG00000168542	2
COL5A1	ENSG00000130635	9
COL5A2	ENSG00000204262	2
DSE	ENSG00000111817	6
LZTS1	ENSG00000061337	8
PLOD1	ENSG00000083444	1
PRDM5	ENSG00000138738	4
SLC39A13	ENSG00000165915	11
TNXB	ENSG00000229341	CHR_HSCHR6_MHC_APD_CTG1
TNXB	ENSG00000236221	CHR_HSCHR6_MHC_MANN_CTG1
TNXB	ENSG00000229353	CHR_HSCHR6_MHC_QBL_CTG1
TNXB	ENSG00000231608	CHR_HSCHR6_MHC_DBB_CTG1
TNXB	ENSG00000206258	CHR_HSCHR6_MHC_MCF_CTG1
TNXB	ENSG00000233323	CHR_HSCHR6_MHC_COX_CTG1
TNXB	ENSG00000236236	CHR_HSCHR6_MHC_SSTO_CTG1
TNXB	ENSG00000168477	6
ZNF469	ENSG00000225614	16

Centoskin gene panel:

Gene name	Gene Ensembl ID	Chromosome/scaffold name
ABCA12	ENSG00000144452	2
ALDH18A1	ENSG00000059573	10
ALOX12B	ENSG00000179477	17
ALOXE3	ENSG00000179148	17
AP1S1	ENSG00000106367	7
APCDD1	ENSG00000154856	18
ATP6V0A2	ENSG00000185344	12
ATP7A	ENSG00000165240	X
CDSN	ENSG00000237165	CHR_HSCHR6_MHC_MCF_CTG1
CDSN	ENSG00000237123	CHR_HSCHR6_MHC_DBB_CTG1

Gene name	Gene Ensembl ID	Chromosome/scaffold name
CDSN	ENSG00000206460	CHR_HSCHR6_MHC_QBL_CTG1
CDSN	ENSG00000237114	CHR_HSCHR6_MHC_MANN_CTG1
CDSN	ENSG00000137197	CHR_HSCHR6_MHC_COX_CTG1
CDSN	ENSG00000204539	6
CERS3	ENSG00000154227	15
CHST8	ENSG00000124302	19
CLDN1	ENSG00000163347	3
COL17A1	ENSG00000065618	10
COL7A1	ENSG00000114270	3
CSTA	ENSG00000121552	3
CYP4F22	ENSG00000171954	19
DSG1	ENSG00000134760	18
DSG4	ENSG00000175065	18
DSP	ENSG00000096696	6
DST	ENSG00000151914	6
EBP	ENSG00000147155	X
EFEMP2	ENSG00000172638	11
ELN	ENSG00000049540	7
ERCC2	ENSG00000104884	19
ERCC3	ENSG00000163161	2
EXPH5	ENSG00000110723	11
FBLN5	ENSG00000140092	14
FERMT1	ENSG00000101311	20
FLG	ENSG00000143631	1
GJB2	ENSG00000165474	13
GJB3	ENSG00000188910	1
GJB4	ENSG00000189433	1
GTF2H5	ENSG00000272047	6
HR	ENSG00000168453	8
ITGA3	ENSG00000005884	17
ITGA6	ENSG00000091409	2
ITGB4	ENSG00000132470	17
JUP	ENSG00000173801	17
KRT1	ENSG00000167768	12
KRT10	ENSG00000186395	17
KRT14	ENSG00000186847	17
KRT2	ENSG00000172867	12
KRT5	ENSG00000186081	12
KRT71	ENSG00000139648	12
KRT74	ENSG00000170484	12
LAMA3	ENSG00000053747	18
LAMB3	ENSG00000196878	1
LAMC2	ENSG00000058085	1
LIPH	ENSG00000163898	3
LIPN	ENSG00000204020	10
LORICRIN	ENSG00000203782	1
LPAR6	ENSG00000139679	13
MMP1	ENSG00000196611	11

Gene name	Gene Ensembl ID	Chromosome/scaffold name
MPLKIP	ENSG00000168303	7
NIPAL4	ENSG00000172548	5
PEX7	ENSG00000112357	6
PHYH	ENSG00000107537	10
PKP1	ENSG00000081277	1
PLEC	ENSG00000178209	8
PNPLA1	ENSG00000180316	6
POMP	ENSG00000132963	13
PYCR1	ENSG00000183010	17
RPL21	ENSG00000122026	13
SLC27A4	ENSG00000167114	9
SNAP29	ENSG00000099940	22
SNRPE	ENSG00000182004	1
SPINK5	ENSG00000133710	5
ST14	ENSG00000149418	11
STS	ENSG00000101846	X
SUMF1	ENSG00000144455	3
TGM1	ENSG00000285348	CHR_HG1_PATCH
TGM1	ENSG00000092295	14
TGM5	ENSG00000104055	15

Skeletal dysplasias panel:

Gene name	Gene Ensembl ID	Chromosome/scaffold name
ACTB	ENSG00000075624	7
ACTG1	ENSG00000184009	17
ACVR1	ENSG00000115170	2
AIFM1	ENSG00000156709	X
ALPL	ENSG00000162551	1
ALX3	ENSG00000156150	1
ALX4	ENSG00000052850	11
ANKH	ENSG00000154122	5
ANKRD11	ENSG00000167522	16
ANO5	ENSG00000171714	11
ARHGAP31	ENSG00000031081	3
ARSB	ENSG00000113273	5
ARSL	ENSG00000157399	X
ATP6V0A2	ENSG00000185344	12
B3GALT6	ENSG00000176022	1
B3GAT3	ENSG00000149541	11
B4GALT7	ENSG00000027847	5
BCS1L	ENSG00000074582	2
BGN	ENSG00000182492	X
BMP1	ENSG00000168487	8
BMPR1B	ENSG00000138696	4
BRCA2	ENSG00000139618	13
BRIP1	ENSG00000136492	17
CA2	ENSG00000104267	8

Gene name	Gene Ensembl ID	Chromosome/scaffold name
CANT1	ENSG00000171302	17
CASR	ENSG00000036828	3
CBL	ENSG00000110395	11
CC2D2A	ENSG00000048342	4
CDH3	ENSG00000062038	16
CDKN1C	ENSG00000273707	CHR_HSCHR11_1_CTG7
CDKN1C	ENSG00000129757	11
CENPJ	ENSG00000151849	13
CEP152	ENSG00000103995	15
CEP290	ENSG00000198707	12
CHST14	ENSG00000169105	15
CHSY1	ENSG00000131873	15
CILK1	ENSG00000112144	6
CLCN5	ENSG00000171365	X
CLCN7	ENSG00000103249	16
COL10A1	ENSG00000123500	6
COL11A1	ENSG00000060718	1
COL11A2	ENSG00000235708	CHR_HSCHR6_MHC_MCF_CTG1
COL11A2	ENSG00000232541	CHR_HSCHR6_MHC_SSTO_CTG1
COL11A2	ENSG00000223699	CHR_HSCHR6_MHC_MANN_CTG1
COL11A2	ENSG00000206290	CHR_HSCHR6_MHC_QBL_CTG1
COL11A2	ENSG00000227801	CHR_HSCHR6_MHC_COX_CTG1
COL11A2	ENSG00000230930	CHR_HSCHR6_MHC_DBB_CTG1
COL11A2	ENSG00000204248	6
COL1A2	ENSG00000164692	7
COL2A1	ENSG00000139219	12
COL5A2	ENSG00000204262	2
COL9A1	ENSG00000112280	6
COL9A2	ENSG00000049089	1
COL9A3	ENSG00000092758	20
COMP	ENSG00000105664	19
CREB3L1	ENSG00000157613	11
CREBBP	ENSG00000005339	16
CRTAP	ENSG00000170275	3
CSPP1	ENSG00000104218	8
CUL7	ENSG00000044090	6
CYP27B1	ENSG00000111012	12
DDR2	ENSG00000162733	1
DHCR24	ENSG00000116133	1
DHCR7	ENSG00000172893	11
DHODH	ENSG00000102967	16
DMP1	ENSG00000152592	4
DYNC2H1	ENSG00000187240	11
EBP	ENSG00000147155	X
EFNB1	ENSG00000090776	X
EFTUD2	ENSG00000108883	17
EIF2AK3	ENSG00000172071	2
ENPP1	ENSG00000197594	6
EP300	ENSG00000100393	22
ERCC4	ENSG00000175595	16
ESCO2	ENSG00000171320	8

Gene name	Gene Ensembl ID	Chromosome/scaffold name
EVC	ENSG00000072840	4
EVC2	ENSG00000173040	4
FANCA	ENSG00000187741	16
FANCB	ENSG00000181544	X
FANCC	ENSG00000158169	9
FANCD2	ENSG00000144554	3
FANCE	ENSG00000112039	6
FANCF	ENSG00000183161	11
FANCG	ENSG00000221829	9
FANCI	ENSG00000140525	15
FANCL	ENSG00000115392	2
FANCM	ENSG00000187790	14
FBN1	ENSG00000166147	15
FBN2	ENSG00000138829	5
FGD1	ENSG00000102302	X
FGF10	ENSG00000070193	5
FGF23	ENSG00000118972	12
FGFR1	ENSG00000077782	8
FGFR2	ENSG00000066468	10
FGFR3	ENSG00000068078	4
FKBP10	ENSG00000141756	17
FLNA	ENSG00000196924	X
FLNB	ENSG00000136068	3
GHR	ENSG00000112964	5
GJA1	ENSG00000152661	6
GLI2	ENSG00000074047	2
GLI3	ENSG00000106571	7
GNAS	ENSG00000087460	20
GNPAT	ENSG00000116906	1
GORAB	ENSG00000120370	1
HDAC4	ENSG00000068024	2
HDAC8	ENSG00000147099	X
HESX1	ENSG00000163666	3
HPGD	ENSG00000164120	4
HRAS	ENSG00000276536	CHR_HSCHR11_1_CTG8
HRAS	ENSG00000174775	11
HSPG2	ENSG00000142798	1
IDH2	ENSG00000182054	15
IDS	ENSG00000010404	X
IFITM5	ENSG00000206013	11
IFT122	ENSG00000163913	3
IFT140	ENSG00000187535	16
IFT172	ENSG00000138002	2
IFT43	ENSG00000119650	14
IFT80	ENSG00000068885	3
IGF1	ENSG00000017427	12
IGF1R	ENSG00000140443	15
INPPL1	ENSG00000165458	11
INSR	ENSG00000171105	19
KAT6B	ENSG00000281813	CHR_HG2191_PATCH
KAT6B	ENSG00000156650	10

Gene name	Gene Ensembl ID	Chromosome/scaffold name
KIF7	ENSG00000166813	15
KMT2A	ENSG00000118058	11
KRAS	ENSG00000133703	12
LBR	ENSG00000143815	1
LHX3	ENSG00000107187	9
LHX4	ENSG00000121454	1
LIFR	ENSG00000113594	5
LMNA	ENSG00000160789	1
LMX1B	ENSG00000136944	9
LONP1	ENSG00000196365	19
LRP5	ENSG00000162337	11
LTBP2	ENSG00000119681	14
LZTR1	ENSG00000099949	22
MAP2K1	ENSG00000169032	15
MAP2K2	ENSG00000126934	19
MATN3	ENSG00000132031	2
MBTPS2	ENSG00000012174	X
MKS1	ENSG00000011143	17
MMP13	ENSG00000137745	11
MMP9	ENSG00000100985	20
MSX2	ENSG00000120149	5
MYCN	ENSG00000134323	2
NEK1	ENSG00000137601	4
NF1	ENSG00000196712	17
NFIX	ENSG00000008441	19
NIPBL	ENSG00000164190	5
NKX3-2	ENSG00000109705	4
NOG	ENSG00000183691	17
NOTCH1	ENSG00000148400	9
NOTCH2	ENSG00000134250	1
NRAS	ENSG00000213281	1
NSD1	ENSG00000165671	5
NSDHL	ENSG00000147383	X
ORC1	ENSG00000085840	1
ORC4	ENSG00000115947	2
OSTM1	ENSG00000081087	6
OTX2	ENSG00000165588	14
P3H1	ENSG00000117385	1
PALB2	ENSG00000083093	16
PCNT	ENSG00000160299	21
PEX14	ENSG00000142655	1
PEX19	ENSG00000162735	1
PEX7	ENSG00000112357	6
PHEX	ENSG00000102174	X
PIGV	ENSG00000060642	1
PIK3CA	ENSG00000121879	3
PITX2	ENSG00000164093	4
PLOD2	ENSG00000152952	3
PLS3	ENSG00000102024	X
POR	ENSG00000127948	7
POU1F1	ENSG00000064835	3

Gene name	Gene Ensembl ID	Chromosome/scaffold name
PPIB	ENSG00000166794	15
PRKAR1A	ENSG00000108946	17
PROP1	ENSG00000280635	CHR_HSCHR5_2_CTG5
PROP1	ENSG00000274382	CHR_HSCHR5_3_CTG5
PROP1	ENSG00000175325	5
PTDSS1	ENSG00000156471	8
PTH1R	ENSG00000160801	3
PTPN11	ENSG00000179295	12
PYCR1	ENSG00000183010	17
RAD21	ENSG00000164754	8
RAD51C	ENSG00000108384	17
RAF1	ENSG00000132155	3
RBBP8	ENSG00000101773	18
RBM8A	ENSG00000265241	1
RECQL4	ENSG00000160957	8
RIT1	ENSG00000143622	1
RMRP	ENSG00000269900	9
RMRP	ENSG00000277027	9
RNU4ATAC	ENSG00000264229	2
ROR2	ENSG00000169071	9
RPGRIP1L	ENSG00000103494	16
RUNX2	ENSG00000124813	6
SALL1	ENSG00000103449	16
SALL4	ENSG00000101115	20
SBDS	ENSG00000126524	7
SERPINF1	ENSG00000282307	CHR_HSCHR17_1_CTG2
SERPINF1	ENSG00000132386	17
SERPINH1	ENSG00000149257	11
SETBP1	ENSG00000152217	18
SHH	ENSG00000164690	7
SHOC2	ENSG00000108061	10
SKI	ENSG00000157933	1
SLC26A2	ENSG00000155850	5
SLC34A3	ENSG00000198569	9
SLC35D1	ENSG00000116704	1
SLC39A13	ENSG00000165915	11
SLCO2A1	ENSG00000174640	3
SLX4	ENSG00000188827	16
SMAD3	ENSG00000166949	15
SMAD4	ENSG00000141646	18
SMC1A	ENSG00000072501	X
SMC3	ENSG00000108055	10
SNX10	ENSG00000086300	7
SOS1	ENSG00000115904	2
SOST	ENSG00000167941	17
SOX2	ENSG00000181449	3
SOX3	ENSG00000134595	X
SOX9	ENSG00000125398	17
SP7	ENSG00000170374	12
STAMBP	ENSG00000124356	2
STAT5B	ENSG00000173757	17

Gene name	Gene Ensembl ID	Chromosome/scaffold name
TBX19	ENSG00000143178	1
TBX5	ENSG00000089225	12
TBXAS1	ENSG00000059377	7
TCF12	ENSG00000140262	15
TCIRG1	ENSG00000110719	11
TCTN3	ENSG00000119977	10
TGFB1	ENSG00000105329	19
TGFB2	ENSG00000092969	1
TGFB3	ENSG00000119699	14
TGFBR1	ENSG00000106799	9
TMEM216	ENSG00000187049	11
TMEM38B	ENSG00000095209	9
TMEM67	ENSG00000164953	8
TNFRSF11A	ENSG00000141655	18
TNFRSF11B	ENSG00000164761	8
TNFSF11	ENSG00000120659	13
TP63	ENSG00000073282	3
TREM2	ENSG00000095970	6
TRIP11	ENSG00000100815	14
TRPS1	ENSG00000104447	8
TRPV4	ENSG00000111199	12
TTC21B	ENSG00000123607	2
TWIST1	ENSG00000122691	7
TYROBP	ENSG00000011600	19
VDR	ENSG00000111424	12
VIPAS39	ENSG00000151445	14
WDR19	ENSG00000157796	4
WDR35	ENSG00000118965	2
WNT1	ENSG00000125084	12
WNT3	ENSG00000277641	CHR_HSCHR17_2_CTG5
WNT3	ENSG00000277626	CHR_HSCHR17_1_CTG5
WNT3	ENSG00000108379	17
WNT5A	ENSG00000114251	3
WNT7A	ENSG00000154764	3
XRCC2	ENSG00000196584	7

Appendix H-Sample concentration prior to Sanger

Sample no.	Patient	Variant	Concentration (ng/µl)
1	III.2	TNXB-617	110.0
2	IV.1		32.9
3	IV.2		7.6
4	III.5		7.3
5	IV.8		93.5
6	IV.9		5.84
7	Singleton		7.7
8	III.2	TNXB-140	14.5
9	IV.1		30.1
10	IV.2		5.6
11	III.5		14.5
12	IV.8		8.9
13	IV.9		12.5
14	Singleton		9.8
15	III.2	SMAD3	9.9
16	IV.1		17.2
17	IV.2		17.5
18	III.5		9.5
19	IV.8		8.9
20	IV.9		32.0
21	Singleton		7.6
22	III.2	FANCA	6.3
23	IV.1		7.8
24	IV.2		6.3
25	III.5		6.7
26	IV.8		6.9
27	IV.9		6.7
28	Singleton		7.7
29	III.2	PNPLA1	21.4
30	IV.1		14.6
31	IV.2		8.5

Sample no.	Patient	Variant	Concentration (ng/µl)
32	III.5		10.1
33	IV.8		6.1
34	IV.9		9.2
35	Singleton		8.5

Appendix I-Theragen WGS analysis report

Whole Genome Sequencing Analysis Report

Ver. 1.0

Theragen Bio

January 18, 2022

Address: (443-270) 2F, B-dong, 864-1, Iui-dong
Yeongtong-gu, Suwon-si, Gyeonnggi-do, Korea

Theragen
Theragen Etex Bio Institute

2nd fl. AICT bldg. B, 145 Gwanggyo-ro,
Yeongtong-gu, Suwon-si, Gyeonggi-do, Republic of Korea, 16229

General Information

1.1 Sequencing Data Statistics

Table 1.1: Sequencing statistics

Sample ID[TBI ID]	Total reads	Total yield(Gb)	N rate	Q30 Rate	Q20 Rate
06-294CG[TN2111D1131]	741040738	111.9 Gb	0.0%	90.94%	96.43%
03-486CZ[TN2111D1132]	663530846	100.19 Gb	0.0%	91.0%	96.39%
SDT-005[TN2111D1133]	640250862	96.68 Gb	0.0%	90.05%	95.95%

- Sample ID(TBI ID): Sample ID provided by customer(Temporary ID assigned by TheragenEtex)
- Total reads:The number of generat ed reads
- Total yield (Gb): Giga base pairs of produced sequence data for each sample
- N rate: The ratio of N, which is the base call instead of an A, C, G, or T
- Q30 Rate: The quality ratio satisfying Phred quality score greater than 30, which represents an error rate of 1 in 1000, with a corresponding call accuracy of 99.9%
- Q20 Rate: The quality ratio satisfying Phred quality score greater than 20, which represents an error rate of 1 in 100, with a corresponding call accuracy of 99%

1.2 Submitted data in FTP

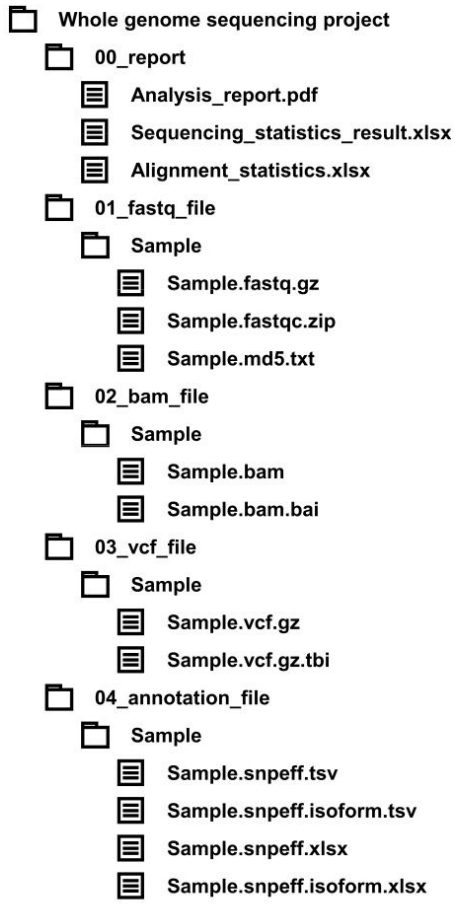


Figure 1.1: Directory structure

Data policy

- To locate analysis result easily on our FTP or HDD, please refer to the following directory structure. Result files can be either fastq, bam, vcf, or pdf with detail explanation. New customers will receive FTP account with randomly created ID and Password. Customers who already have an experience with our services will receive data on existing account without additional request.

- Additional cost will be charged for extra analysis, such as statistical analysis using different tools which are not in our analysis pipeline, drawing figures or writing phrases for the journal. Please contact sales representatives about detail services.

- The result will be stored for 3 months(until Jan-18-2022). For security reasons, data will be completely wiped out for data privacy of our customers after the expiration. None of our customer's data will exist on our server side. Please download the result immediately when you received the notification of completion of your priceless works. With confirmation from our customers, data will remain for extra 3 months. After the total 6 months, extra payment will be charged. For detail data storing services, please contact our sales representatives.

Workflow

2.1 Whole Genome Process

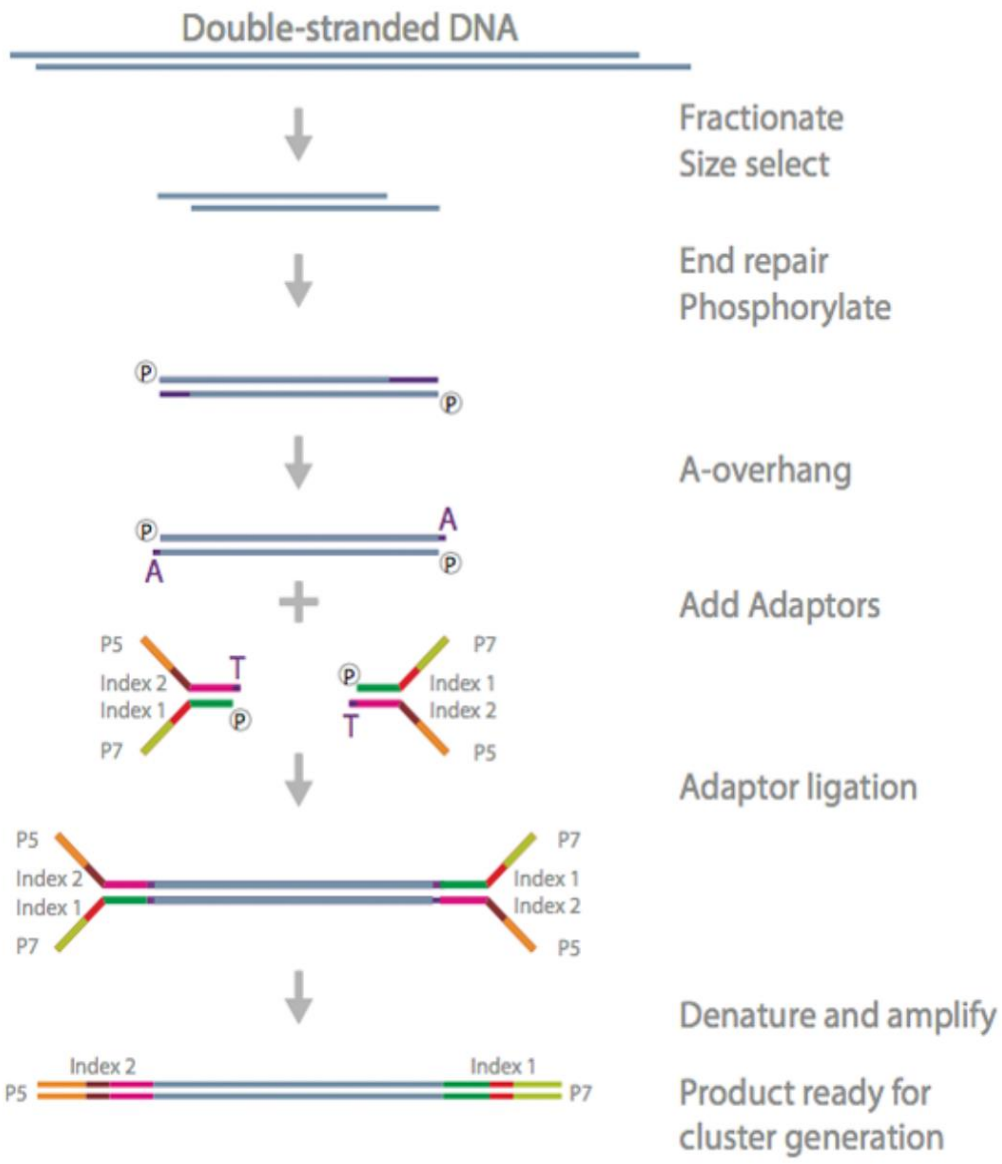
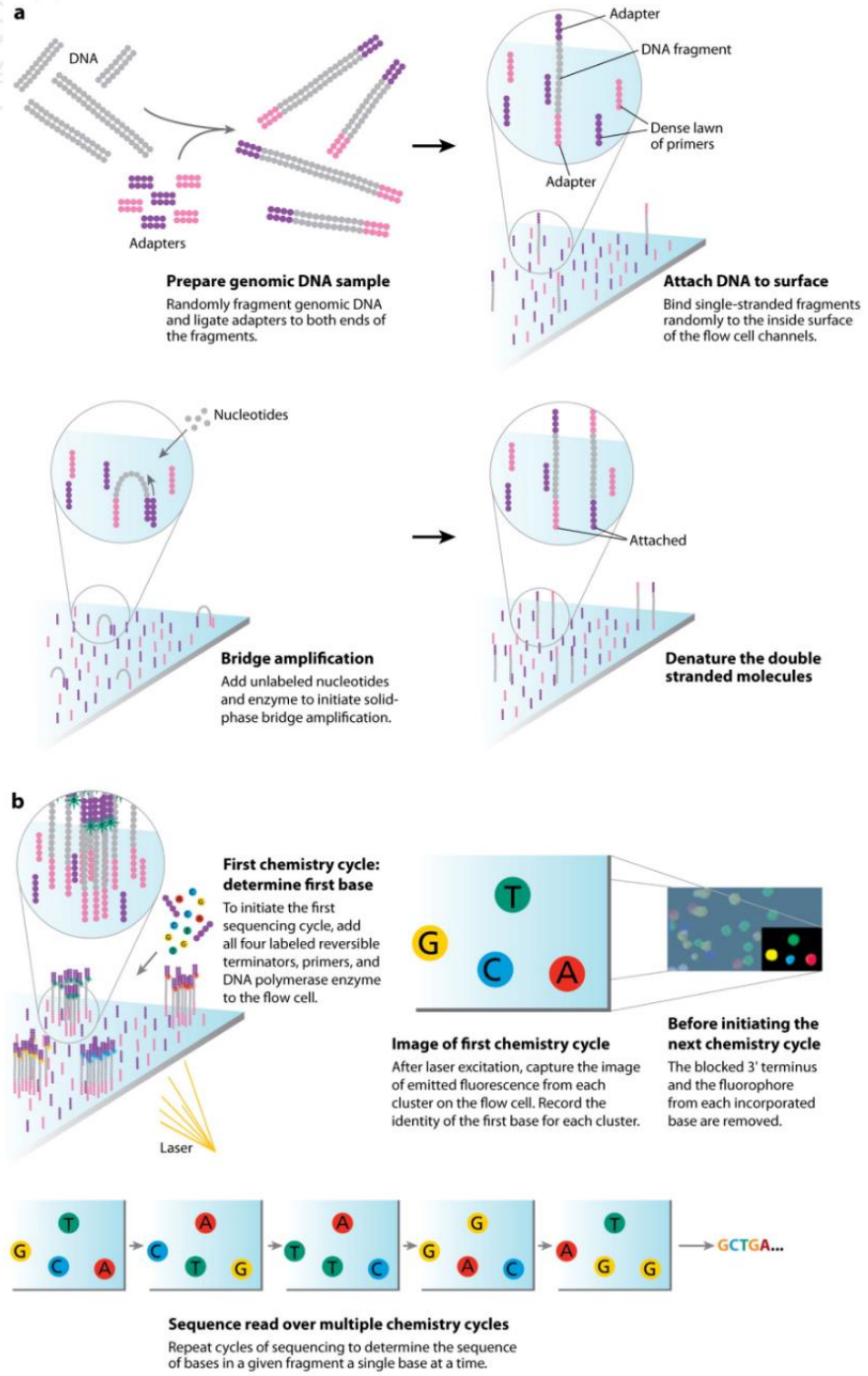


Figure 2.1: Whole genome process

Table 2.1: Whole genome process detail

Item	Description
Reference	UCSC hg19, Assembled chromosomes
Read length	150 x 2
Coverage uniformity (10X)	≥ 90%

2.2 Sequencing Method



Mardis ER. 2008.
Annu. Rev. Genomics Hum. Genet. 9:387–402

Figure 2.2: Sequencing method

2.3 Analysis Pipeline

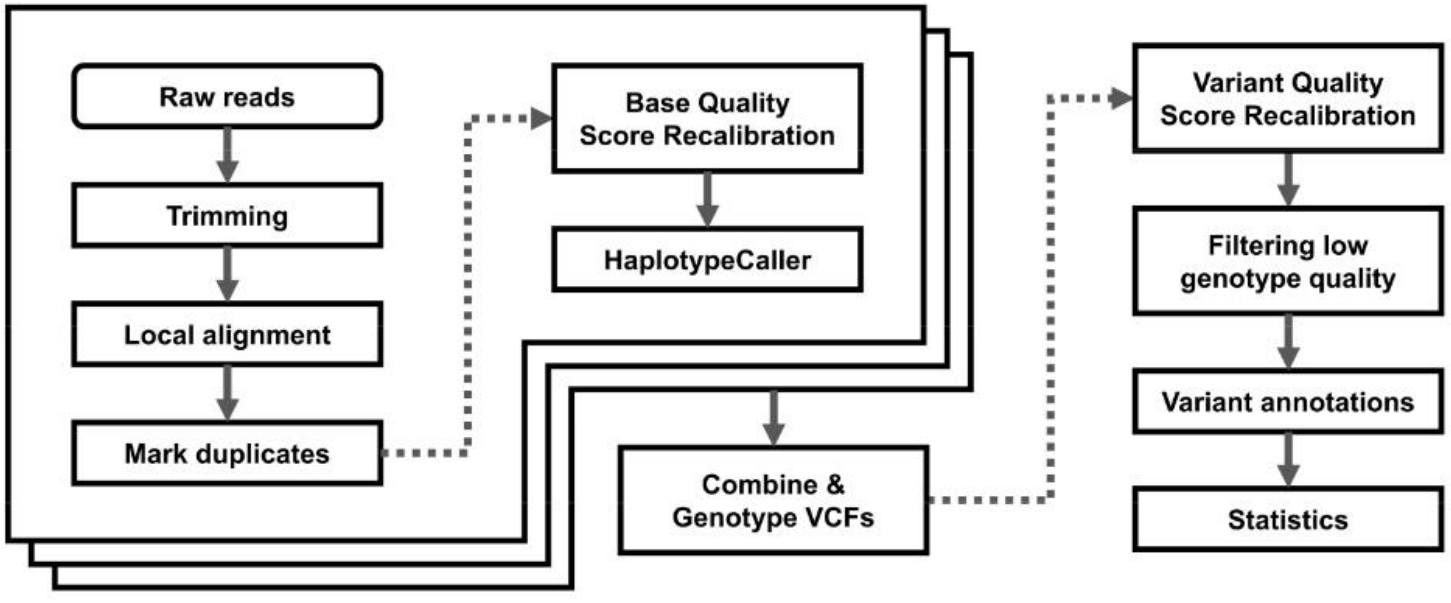


Figure 2.3: Analysis pipeline diagram

Pipeline details: Followed the GATK best practices recommendations [1]

c. FastQC (0.11.7): Quality control for sequencing data[2]

d. Trimmomatic (0.4.4): Trimming adapter sequences and low-quality sequences for fastq files[3]

e. BWA (0.7.17): Mapping and aligning sequencing reads to reference genome sequence using burrow-wheeler algorithem[4]

f. GATK (4.0.11.0): Improvement of alignments and genotype calling and refining with optimized parameters[5]

g. Samtools (1.8): Converting from SAM format to sorted, filtered, and indexed BAM files[6]

h. SnpEff (4.3t): Annotating genetic variants and predicting their effects[7]

Data Analysis Results

3.1 Basic Statistics

3.1.1 Mapping Summary for Produced Sequencing Data

Table 3.1: Mapping summary

Sample	Num of reads	Deduplication (%)	Mapping (%)	Unique (%)	Mean depth (std)
06-294CG	741,040,738	669,907,402(93.49)	654,049,360(97.63)	629,623,524(93.99)	31.413(132.2792)
03-486CZ	663,530,846	602,861,946(94.33)	585,725,031(97.16)	560,394,994(92.96)	28.1214(126.9666)
SDT-005	640,250,862	579,328,777(94.3)	564,960,556(97.52)	543,434,503(93.8)	27.0807(122.5262)

5. The mapping process is to compare the sequenced reads to a customer-designated reference gonome using string alignment algorithm.

6. Sample: ID provided from customer

7. Num of reads: Unfiltered sequence

8. Deduplicated (%): Discarded clean reads following PCR duplicate

9. Mapping (%): De-duplicated reads followed by mapping onto the reference genome

10. Unique (%): Reads with the same starting position on each end

11. Mean depth (std): Mean, standard deviation of coverage is calculated based on extraction of pair overlap-region

3.2 Annotated Variants

3.2.1 Number Variants By Type

Figure 3.1: A bar graph for the number of variations by type

Table 3.2: The number of variations by type

Sample ID	SNP	DEL	INS	TOTAL
06-294CG	4,255,633	629,232	598,653	5,483,518
03-486CZ	4,203,985	609,843	581,291	5,395,119
SDT-005	4,167,616	608,284	578,938	5,354,838

- Sample ID: ID provided from customer
- SNP: Number of Single-Nucleotide Polymophism
- INS: Number of Insertion mutation, which is the addition of one or more nucletide base pairs into a DNA sequence
- DEL: Number of Deletion mutation in which a part of a chromosome or sequence of DNA is missing

3.2.2 Ts/Tv and Hetero/Homo

Table 3.3: Statistics for variations of each sample

Sample ID	TS	TV	Ts/Tv	Hetero Variants	Homo Variants	Hetero/Homo
06-294CG	3,766,067	1,893,877	1.989	3,383,345	1,854,916	1.824
03-486CZ	3,753,706	1,885,573	1.991	3,269,997	1,885,075	1.735
SDT-005	3,711,702	1,865,854	1.989	3,262,048	1,852,944	1.76

- Sample ID: ID provided from customer
- TS (Transitions): Number of transitions, which are point mutations that changes a purine nucleotide to another purine or a pyrimidine nucleotide to another pyrimidine
- TV (Transversions): Number of transversions, which refer to the substitution of a purine for a pyrimidine or vice versa, in deoxyribonucleic acid (DNA)
- Ts/Tv ratio: The ratio of the number of transitions to the number of transversions for a pair of sequences
- Hetero Variants: Number of Hetero Variants
- Homo Variants: Number of Homo Variants
- Hetero/Homo ratio: The ratio of the number of hetero variants to the number of homo variants for a pair of sequences

3.2.3 Variants Categories

Figure 3.2: A bar graph for the variants categories

Table 3.4: The variants categories

Effect category	06-294CG	03-486CZ	SDT-005
DOWNSTREAM	522,467	513,647	504,283
EXON	91,929	92,058	89,250
INTERGENIC	3,218,801	3,164,385	3,156,685
INTRON	5,503,946	5,412,441	5,293,797
SPLICE SITE ACCEPTOR	391	377	355
SPLICE SITE DONOR	280	317	289
SPLICE SITE REGION	8,793	8,511	8,272
TRANSCRIPT	54,752	53,197	51,272
UPSTREAM	515,364	505,022	498,297
UTR 3 PRIME	89,160	88,816	85,042
UTR 5 PRIME	18,308	17,721	17,196
TOTAL	10,024,191	9,856,492	9,704,738

- If the variant is applicable to multiple regions, the variant will be counted repeatedly.
- DOWNSTREAM: Downstream of a gene (default length:5K bases)
- EXON: The variant hits a gene
- INTERGENIC: The variant is in an intergenic region
- INTRON: Variant hits and intron. Technically, hits no exon in the transcript
- SPLICE SITE ACCEPTOR: The variant hits a splice acceptor site
- SPLICE SITE DONOR: The variant hits a splice donor site
- SPLICE SITE REGION: The variant hits a splice region site
- TRANSCRIPT: The variant hits a transcript
- UPSTREAM: Upstream of a gene (default length: 5K bases)
- UTR 3 PRIME: Variant hits 3'UTR region
- UTR 5 PRIME: Variant hits 5'UTR region

3.2.4 Number of Effects by Functional Class

Figure 3.3: A bar graph for the variants effect

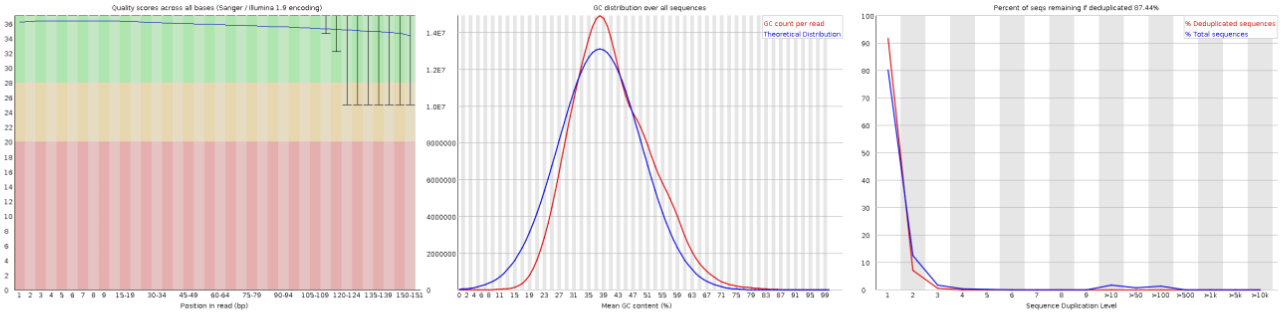
Table 3.5: the number of variations by type

Sample ID	Silent Mutation	Missense Mutation	Nonsense Mutation	TOTAL
06-294CG	25,779	23,790	239	49,808
03-486CZ	25,644	23,566	246	49,456
SDT-005	25,205	22,554	191	47,950

- Sample ID: ID provided from customer
- Silent Mutation: Number of Silent mutation, which is mutation in DNA tha do not significantly alter the phenotype of organism in which it occurs
- Missense Mutation: Number of Missense mutation, which is a point mutation in which a single nucleotide change results in a codon that codes for a different amino acid
- Nonsense Mutation: Number of Nonsense mutation, which is a point mutation in a sequence of DNA results in a premature stop codon

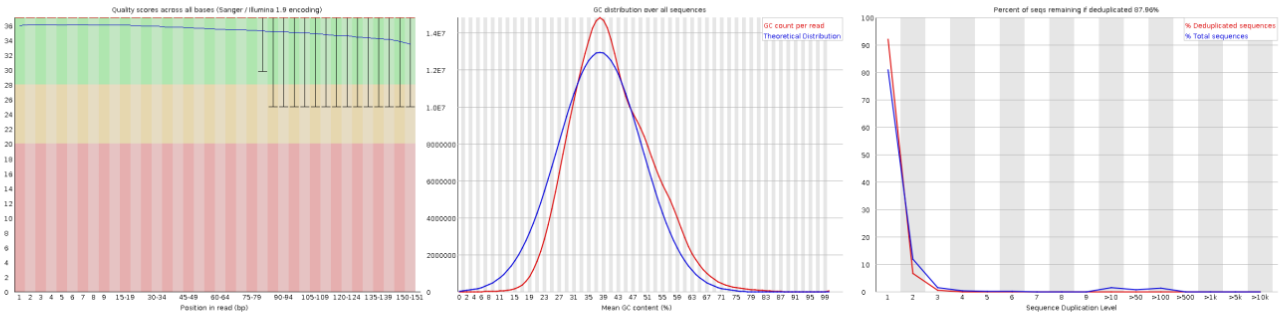
3.3 Quality Controls

3.3.1 FastQC Results



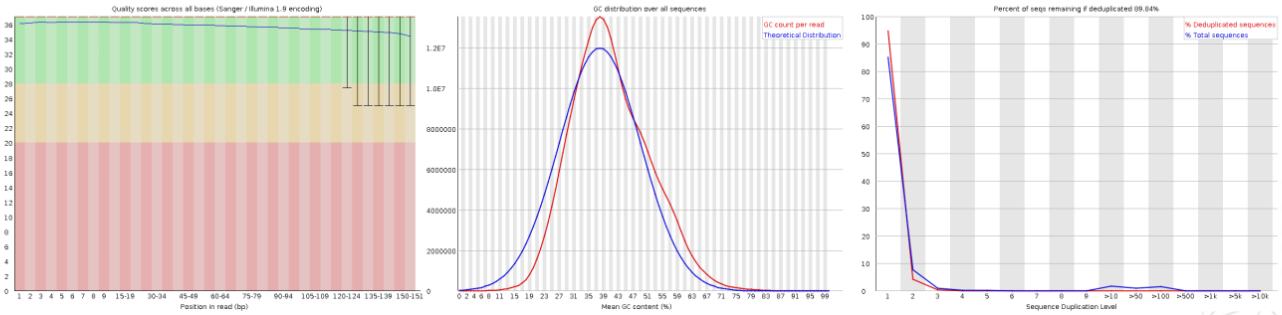
(a) Per base qaulity (b) GC contents (c) Duplication labels

Figure 3.4: 06-294CG_1_fastq.gz



(a) Per base qaulity (b) GC contents (c) Duplication labels

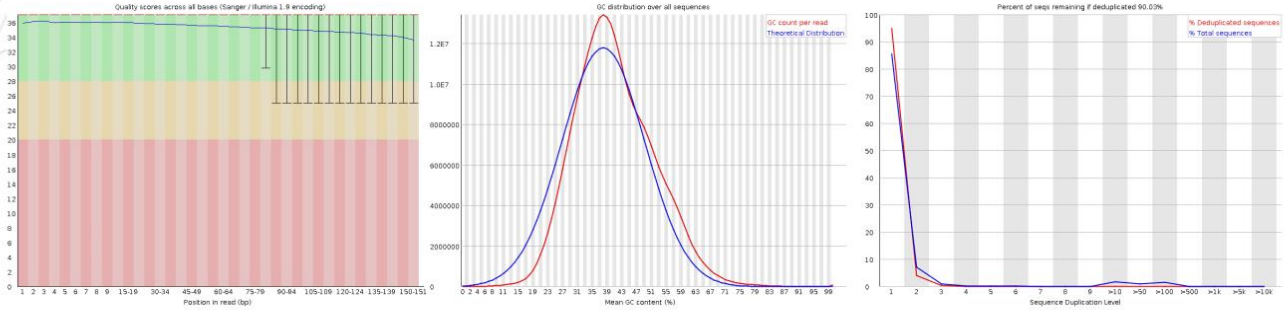
Figure 3.5: 06-294CG_2_fastq.gz



(a) Per base qaulity (b) GC contents (c) Duplication labels

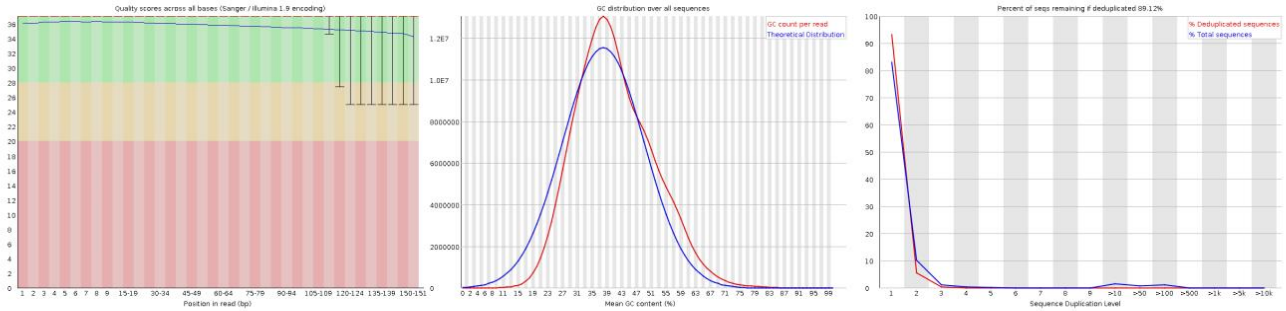
Figure 3.6: 03-486CZ_1_fastq.gz

- The QC results were generated using FastQC software
- Per base qual: Quality values across all bases at each position
- GC contents: GC content of each base position in a sample
- Duplication level: Proportion of sequence duplication level



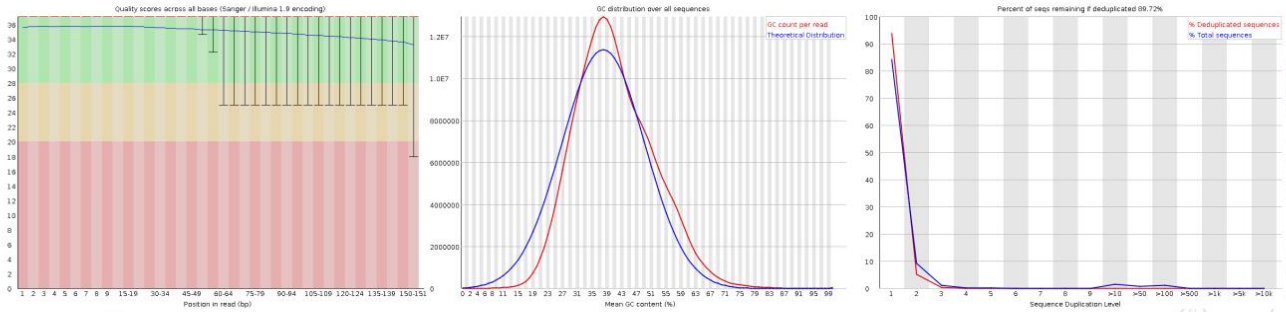
(a) Per base qaulity (b) GC contents (c) Duplication labels

Figure 3.7: 03-486CZ-2-fastq.gz



(a) Per base qaulity (b) GC contents (c) Duplication labels

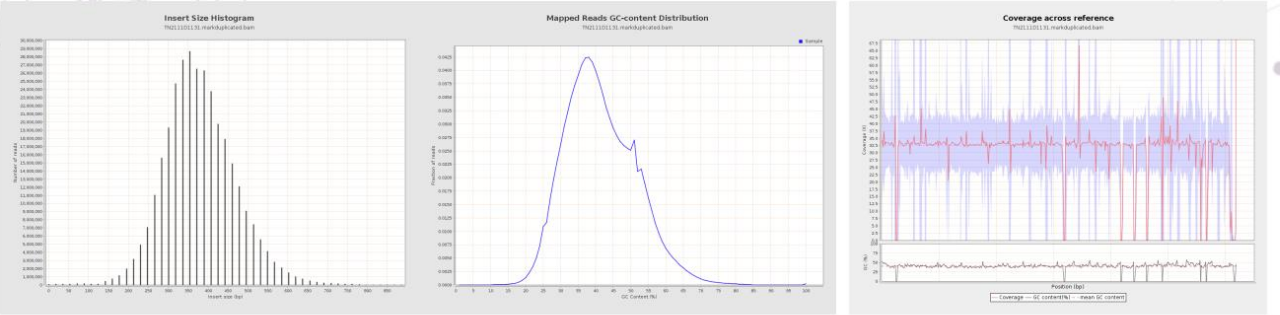
Figure 3.8: SDT-005-1-fastq.gz



(a) Per base qaulity (b) GC contents (c) Duplication labels

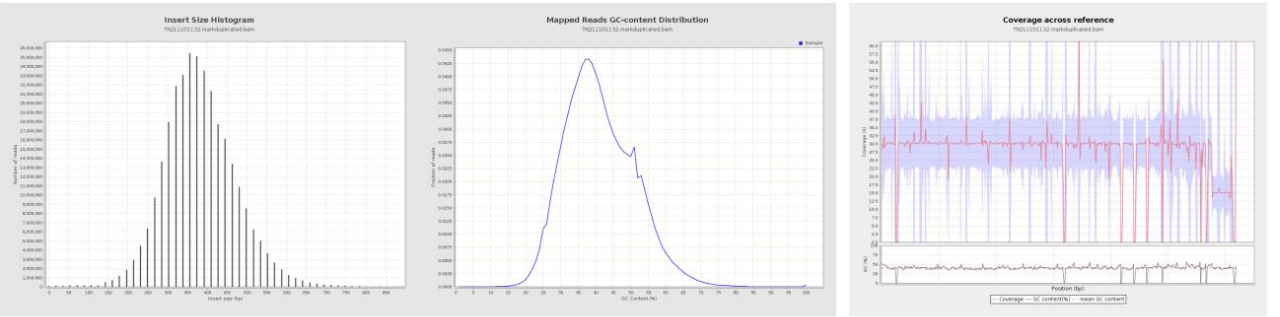
Figure 3.9: SDT-005-2-fastq.gz

3.3.2 Qualimap Results



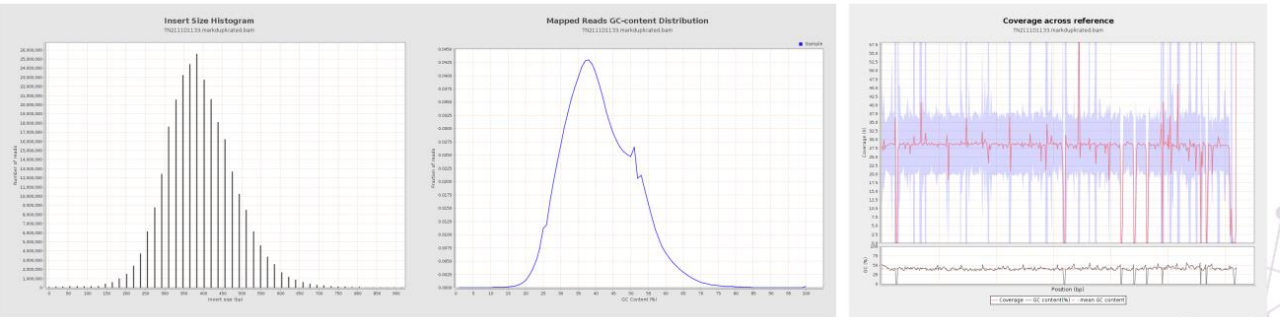
(a) Insert size (b) GC contents (c) Reference coverage

Figure 3.10: 06-294CG



(a) Insert size (b) GC contents (c) Reference coverage

Figure 3.11: 03-486CZ



(a) Insert size (b) GC contents (c) Reference coverage

Figure 3.12: SDT-005

- The QC results were generated using Qualimap software
- Insert Size: Histogram of insert size distribution of reads mapped on reference genome
- GC contents: GC content of each base position in a sample
- Reference coverage: Coverage per position(bp) of reference genome

Appendix

4.1 Figure Description

4.1.1 FastQC

Quality score across all bases

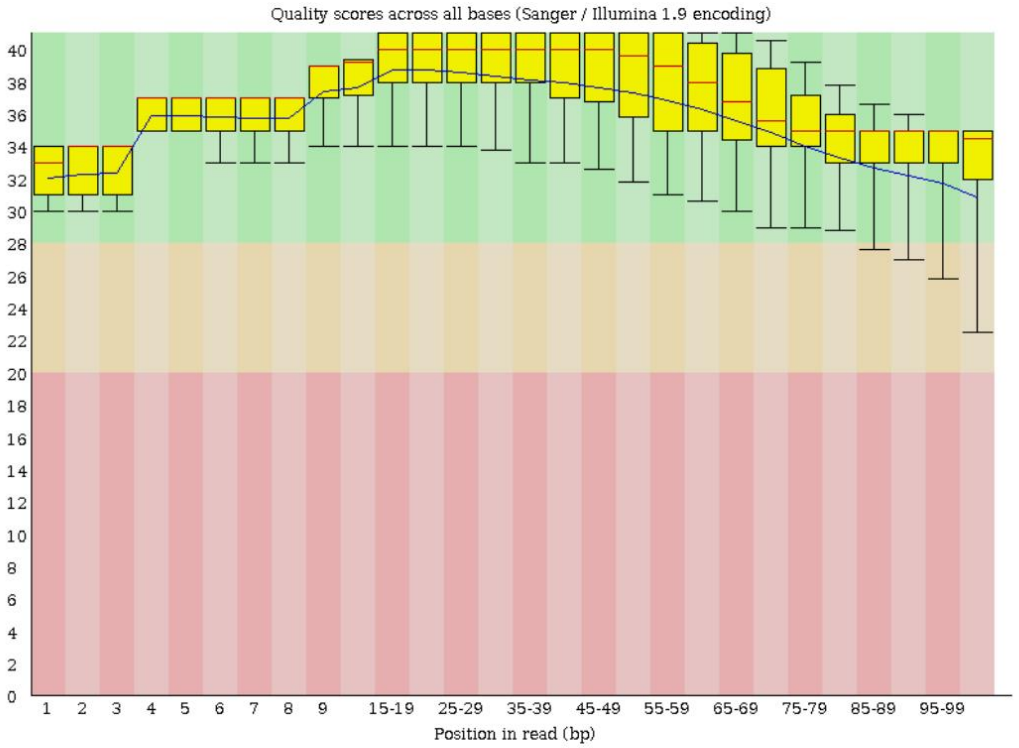


Figure 4.1: Quality score per base

This figure shows an overview of the range of quality values across all bases at each position in the fastq file.

- X-axis : Position in reads(bp)
- Y-axis : Quality score of sequence reads
- Green region : Very good quality calls
- Orange region: Calls of reasonable quality
- Red region : Calls of poor quality
- Central Red Line : Median Value
- Yellow Box: Inter-Quartile Range (25 75%)
- Upper and Lower whiskers : 10% and 90% points
- Blue Line : mean quality

GC content across all bases

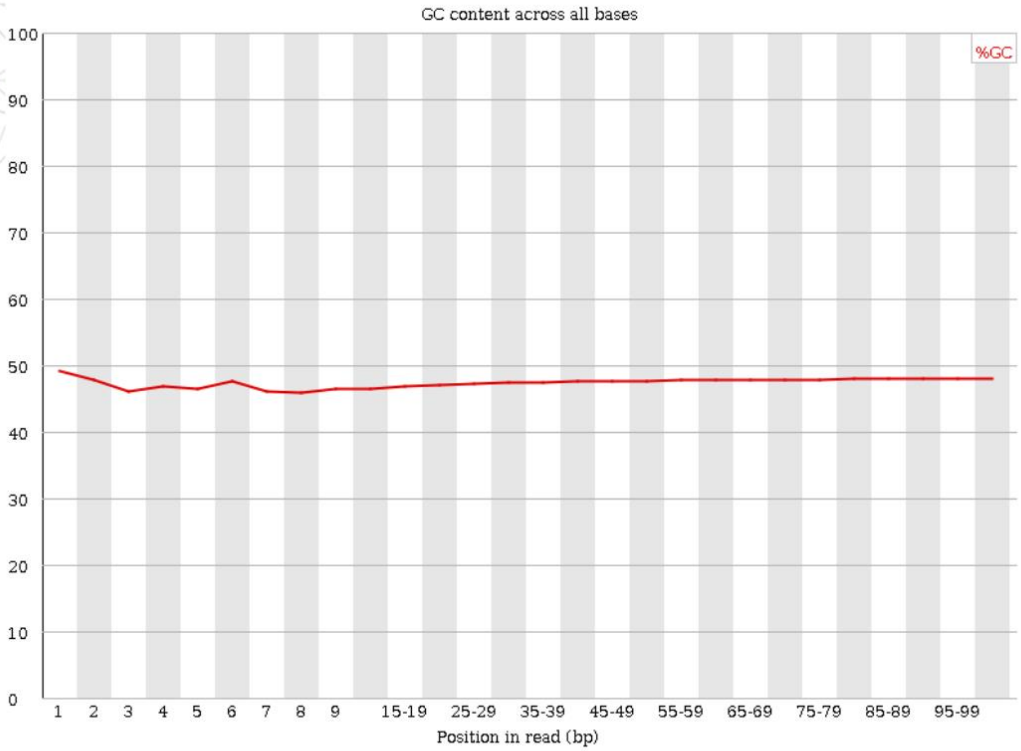


Figure 4.2: GC content per base

- This figure shows an overview of the GC content of each base position in a file.
- X-axis: Position in reads (bp)
- Y-axis: GC rate of each base position

4.1.2 Qualimap

Insert Size Histogram

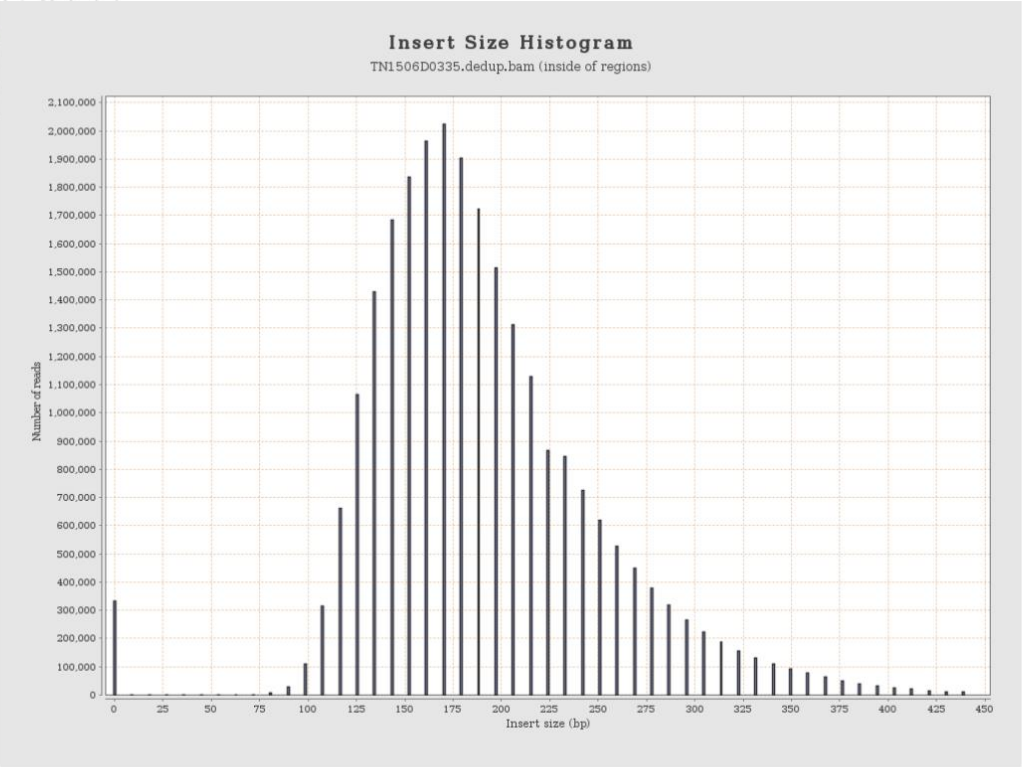


Figure 4.3: Insert size histogram

- The Distribution of insert size of reads mapped on reference genome.
- Software Information: Qualimap v.2.2.1
- X-axis: Insert size(bp)
- Y-axis: Number of reads

Mapped Reads GC-content Distribution

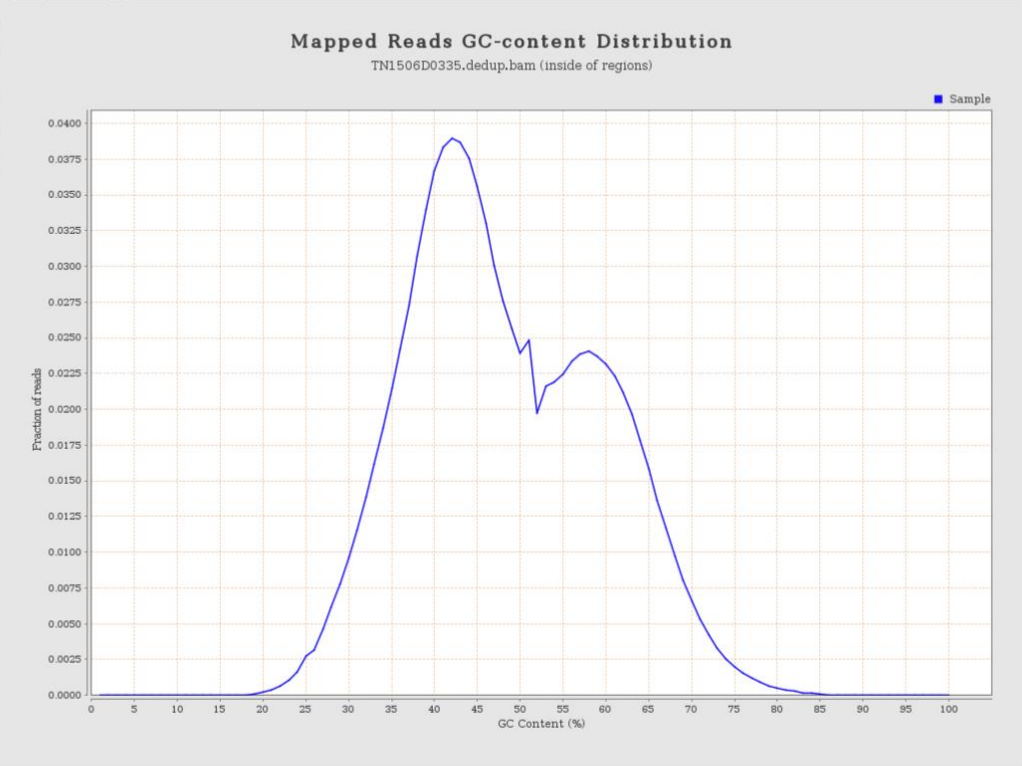


Figure 4.4: Distribution of mapped reads GC-content

- This graph shows the distribution of GC content per mapped read.
- X-axis: GC content (%)
- Y-axis: Fraction of reads

Mapped Reads GC-content Distribution

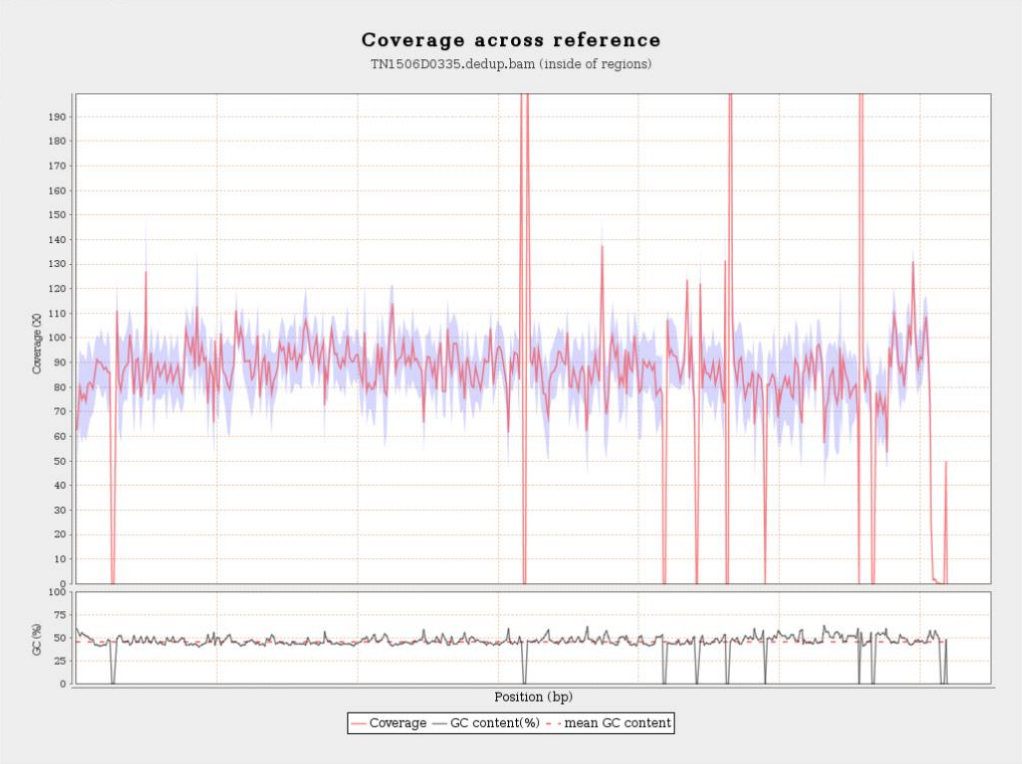


Figure 4.5: Coverage across reference

- This plot consists of two figures. The upper figure provides the coverage distribution (red line) and coverage deviation across the reference sequence. The coverage is measured in X. The lower figure shows GC content across reference (black line) together with its average value (red dotted line).
- X-axis: Position
- Y-axis of upper figure: Coverage (X)
- Y-axis of lower figure: GC (%)

4.2 File Format Description

4.2.1 Variant Calling Format (VCF)

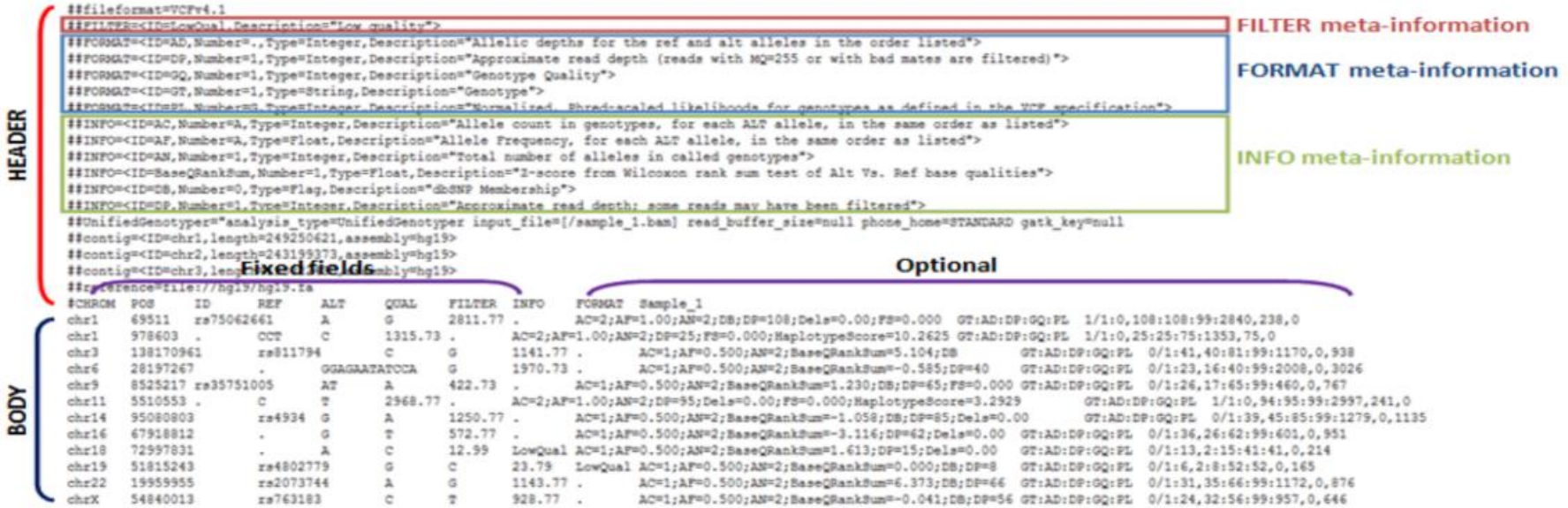


Figure 4.6: The VCF example form

- The HEADER contains meta-information lines that provide supplemental information about variants contained in the BODY of file.

- Body: Data lines are tab-delimited and list information about individual variants and associated genotypes across samples. The first 8 fields (Figure 1) are required to be listed in the VCF column header line. Some of these fields require non-null values (see Table 6) for each record. For the remaining fixed fields, even if the field does not have an associated value, it still needs to be specified with a missing value identifier (“.” in VCF 4.1). Subsequent fields are optional.

- Fixed/Optional fields: The field contains a colon-separated list of all pre-defined FORMAT sub-fields (same as ID in a FORMAT declaration) that are applicable to all samples that follow.

- INFO/FORMAT/Filter meta-Information: optional fields that have to be declared in the HEADER if they are being referred to in the BODY of file. Different keys that can be used to define them are described in Table 3. All three fields do not use the same set of keys. Please refer to individual field definitions for further details.

Table 4.1: The meta information such as INFO / FORMAT / filter in VCF file

Item	Description
CHROM	Chromosome
POS	Position
ID	Identifer
REF	Reference base(s)
ALT	Alteration base(s)
QUAL	Quality
FILTER	Filter status
INFO	Additional information
AC	Allele count in genotypes, for each ALT allele, in the same order as listed
AF	Allele frequency for each ALT allele in the same order as listed
AN	Total number of alleles in called genotypes
BaseQRankSum	Z-score from Wilcoxon rank sum test of Alt Vs. Ref base qualities
DB	dbSNP membership
DP	Combined depth across samples, e.g. DP=154
DS	Were any of the samples downsampled?
Dels	Fraction of Reads Containing Spanning Deletions
FS	Phred-scaled p-value using Fisher
HaplotypeScore	Consistency of the site with at most two segregating haplotypes
InbreedingCoeff	Inbreeding coefficient as estimated from the genotype likelihoods per-sample when compared against the Hardy-Weinberg expectation
MLEAC	Maximum likelihood expectation (MLE) for the allele counts (not necessarily the same as the AC)
MLEAF	Maximum likelihood expectation (MLE) for the allele frequency (not necessarily the same as the AF)
MQ	RMS mapping quality
MQ0	Total Mapping Quality Zero Reads
MQRankSum	Z-score From Wilcoxon rank sum test of Alt vs. Ref read mapping qualities
QD	Variant Confidence/Quality by Depth
RPA	Number of times tandem repeat unit is repeated, for each allele (including reference)
RU	Tandem repeat unit (bases)
ReadPosRankSum	Z-score from Wilcoxon rank sum test of Alt vs. Ref read position bias
STR	Variant is a short tandem repeat
AD	Allelic depths for the ref and alt alleles in the order listed
DP	Approximate read depth
GQ	Genotype quality
GT	Genotype
PL	Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification

4.2.2 Annotation Result Excel Format (xls)

#CHROM	POS	ID	REF	ALT	FILTER	VARTYPE	GEN[Sample_1].GT	GEN[Sample_1].AD	ANN[*].EFFECT	ANN[*].IMPACT	ANN[*].GENE	ANN[*].FEATURE	ANN[*].FEATUREID
chr1	69511	rs75062661	A	G	.	SNP	1/1	0,108	missense_variant	MODERATE	OR4F5	transcript	NM_001005484.1
chr1	897470	.	A	C	LowQual	SNP	0/1	19,5	upstream_gene_variant	MODIFIER	NOC2L	transcript	NM_015658.3
chr2	1442551	rs3036132;rs341	AT	A	.	DEL	0/1	48,42	intron_variant	MODIFIER	TPO	transcript	NM_000547.5
chr3	52868716	rs77032405	GC	G	.	DEL	1/1	2,95	intron_variant	MODIFIER	MUSTN1	transcript	NM_205853.3
chr4	88226322	rs78194430	C	G	.	SNP	1/1	0,40	3_prime_UTR_variant	MODIFIER	HSD17B13	transcript	NM_178135.4
chr4	88258483	rs9991501	T	C	.	SNP	1/1	0,17	missense_variant	MODERATE	HSD17B11	transcript	NM_016245.4
chr4	88533441	rs13131929	T	C	.	SNP	1/1	0,78	intron_variant	MODIFIER	DSPP	transcript	NM_014208.3
chr4	88534235	rs2736982	A	G	.	SNP	1/1	0,83	synonymous_variant	LOW	DSPP	transcript	NM_014208.3
chr6	80626375	rs3812153	T	C	.	SNP	0/1	27,29	missense_variant	MODERATE	ELOVL4	transcript	NM_022726.3
chr6	80715838	rs240226;rs2402	T	G	.	SNP	1/1	1,69	intron_variant	MODIFIER	TTK	transcript	NM_003318.4
chr6	80720487	rs111611266	G	GTA	.	INS	1/1	0,17	intron_variant	MODIFIER	TTK	transcript	NM_003318.4
chr6	80774236	rs7771228	C	T	.	SNP	1/1	0,115	intergenic_region	MODIFIER	TTK-BCKDHB	intergenic_region	TTK-BCKDHB
chr6	80777608	rs3967160	G	A	.	SNP	0/1	35,27	intergenic_region	MODIFIER	TTK-BCKDHB	intergenic_region	TTK-BCKDHB
chr9	131196336	rs2072397	T	A	.	SNP	1/1	1,69	intron_variant	MODIFIER	CERCAM	transcript	NM_016174.4
chr9	131196704	rs7259	G	A	.	SNP	1/1	0,50	synonymous_variant	LOW	CERCAM	transcript	NM_016174.4
chr9	131380225	rs7864187	A	C	.	SNP	0/1	20,19	intron_variant	MODIFIER	SPTAN1	transcript	NM_001130438.2
chr9	131392553	rs7866175;rs386	G	A	.	SNP	1/1	0,24	downstream_gene_variant	MODIFIER	WDR34	transcript	NM_052844.3
chr11	7723683	rs4373911	G	A	.	SNP	1/1	1,116	intron_variant	MODIFIER	OVCH2	transcript	NM_198185.4
chr11	7750622	rs4237761	T	C	.	SNP	1/1	0,122	intergenic_region	MODIFIER	OVCH2-OR5P2	intergenic_region	OVCH2-OR5P2
chr12	56494991	rs2271189	G	A	.	SNP	0/1	44,43	upstream_gene_variant	MODIFIER	PA2G4	transcript	NM_006191.2
chr12	56537026	.	T	A	.	SNP	0/1	20,24	missense_variant	MODERATE	ESYT1	transcript	NM_001184796.1
chr12	56537027	.	C	A	.	SNP	0/1	20,24	synonymous_variant	LOW	ESYT1	transcript	NM_001184796.1
chr12	56551339	rs28365930;rs28	C	CAA	.	INS	0/1	21,18	upstream_gene_variant	MODIFIER	MYL6	transcript	NM_021019.4
chr12	56604416	rs10532115	GAT	G	.	DEL	1/1	0,16	intron_variant	MODIFIER	RNF41	transcript	NM_001242826.1
chr16	2317778	rs11642797	T	G	LowQual	SNP	1/1	0,1	upstream_gene_variant	MODIFIER	MIR940	transcript	NR_030636.1

Figure 4.7: The snpEff annotation example form

- Refer to 2nd, 3rd sheet of the excel file for the detail explanation.

Bibliography

[1] Geraldine A. Van der Auwera, Mauricio O. Carneiro, Chris Hartl, Ryan Poplin, Guillermo del Angel, Ami Levy-Moonshine, Tadeusz Jordan, Khalid Shakir, David Roazen, Joel Thibault, Eric Banks, Kiran V. Garimella, David Altshuler, Stacey Gabriel, and Mark A. DePristo. From fastq data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics, 11(1110):11.10.1–11.10.33, Oct 2013.

[2] Simon Andrews. Fastqc: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[3] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics, 30(15):2114–2120, 2014.

[4] Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics, 25(14):1754–1760, July 2009.

[5] Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, and Mark A DePristo. The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. Genome Research, 20(9):1297–303, Sep 2010.

[6] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard Durbin. The sequence alignment/map format and samtools. Bioinformatics, 25(16):2078–2079, August 2009.

[7] P. Cingolani, A. Platts, M. Coon, T. Nguyen, L. Wang, S.J. Land, X. Lu, and D.M. Ruden. A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: Snps in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2):80–92, 2012.

[8] Konstantin Okonechnikov, Ana Conesa, and Fernando Garc´ıa-Alcalde. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics, 32(2):292–294, 2016

Appendix J-Medical questionnaire answer list

	III.2	IV.1	IV.2	III.5	IV.8	IV.9	SDT-005
General information							
Gender	F	M	F	F	F	F	F
Age	57	36	30	44	21	27	48
Prescriptions daily use	Yes, various	No	Yes, various	Yes, paracetamol	No	Yes, oral contraceptive pill	Yes, various
Surgical interventions	Yes	Yes	Yes	Yes	No	No	Yes
If yes, site of intervention	Lumpectomy, hysterectomy	Removal of excess cartilage in right knee	Caesarean sections, tonsillectomy, MVP plug	Grommets, Tonsillectomy, Adenoidectomy, Arterial stent	/	/	Wisdom tooth extraction, cholecystectomy, left-shoulder alignment
Past experience with anaesthetic	Yes	Yes	Yes	Yes	No	Yes	Yes
Dose sufficient for analgesia	No	No	No	No, Occasional	/	No	No
Physical involvement							
Occupation	Housewife	Supervisor	Customer service	Clerk	Cashier	Brand Manager	Teacher
Duration of employment	>20 years	21 years	15 years	12 years	3 years	9 years	25 years
Work hours per week	21 hours	40+ hours	20 hours	40+ hours	40 hours	40+ hours	40+ hours
Physical activity (Sedentary, light, active/sports)	Sedentary	Active-Basketball & shooting	Light	Sedentary	Light	Light	Active-swimming, running
Nutritional factors							
Diet high in animal protein	No	Yes	Yes	Yes	Yes	No	No
Dairy products	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Green vegetables	Yes	Yes	Yes	Yes	Yes	Yes	Yes
>3 Caffeinated drinks daily?	Yes	Yes	Yes	Yes	No	Yes	No
Alcohol	Rarely	Weekly	Non-	Rarely	Rarely	Weekly	Non-

consumption			drinker				drinker
Smoker?	Yes	Yes	No	Yes	Yes	Yes	Non-smoker
If yes, no. of cigarettes daily	6 units daily	10 units daily	/	20 units daily	1 unit daily	10 units daily	/
Daily supplement use	Yes	No	No	No	Yes	Yes	Yes
If yes, which & frequency	Vitamin D, Tumeric	/	/	/	Vitamin D	Multivitamin pills inc. Vit D	Electrolytes, Vit D, fish oil, etc

Symptomatology	III.2	IV.1	IV.2	III.5	IV.8	IV.9	SDT-005
Psychiatric							
Anxiety/Depression/PTSD	No	No	Yes	Yes	Yes	Yes	Yes
Immunological							
Asthma	Yes	No	Yes	Yes	Yes	No	Yes
Allergies	No	No	No	No	Yes	Yes	Yes
Cardiac							
Hyper/hypotension	No	No	No	Low BP	Low BP	No	Low BP
Postural hypotension	Yes	No	Yes	Yes	Yes	No	Yes
Palpitations	Yes	No	Yes	Yes	No	No	Yes
Dermatological							
Prolonged wound healing	Yes	No	Yes	Yes	No	No	Yes
Atrophic scarring	Yes	No	Yes	Yes	Yes	Yes	Yes
Thin stretchy skin	Yes	No	Yes	Yes	Yes	Yes	Yes
Easy bruising	Yes	No	Yes	Yes	Yes	Yes	Yes
Piezogenic papules	Yes	No	Yes	Yes	No	Yes	Yes
Gynaecological							
Regular cycle	Yes (Past)	N/A	Yes	Yes	No	Yes	Yes (Past)
Menorrhagia	Yes	N/A	Yes	Yes	Yes	Yes	Yes
Menstrual cramp severity	High	N/A	Moderate/ High	High	High	Moderate	High
Endometriosis?	Yes	N/A	Yes	No	No	No	No
Gastrointestinal							
Acid reflux	Yes	No	No	Yes	No	Yes	Yes
Diarrhoea/ constipation	No	No	No	No	No	No	No
Abdominal pain/ discomfort	Yes	No	Yes	Yes	No	No	Yes
Nausea/vomiting	Yes	No	Yes	Yes	Post-prandial	No	No

Symptomatology	III.2	IV.1	IV.2	III.5	IV.8	IV.9	SDT-005
Musculoskeletal							
Fatigue/ brain fog	Yes	No	Yes	Yes	No	No	Yes
If yes, time of onset	Early	/	Early, chronic	Early	/	/	Early
Jaw lock	Yes	No	Yes	Yes	No	Yes	Yes
Pain upon mastication	Yes	No	No	Yes	No	No	No
Chronic pain	Yes	No	No	Yes	Yes	No	Yes
Site of pain, if any	Widespread, knees	Shoulders	Knees	Widespead	Neck	/	Widespread
History of joint dislocation	No	Yes	Yes	No	No	Yes	Yes
History of fractures	No	Yes	No	No	No	No	Yes
Scoliosis/Kyphosis	No	No	No	No	Slight scoliosis	No	No
Functional limitations							
Walking	No	No	Yes	Yes	No	No	Yes
Climbing/descending stairs	No	No	Yes	No	No	No	Yes
Balance & coordination	No	Yes	Yes	Yes	Yes	No	Yes
Lifting/carrying/twisting	No	No	No	Yes	No	No	Yes
Gripping objects	Yes	No	No	Yes	No	No	Yes

Appendix K-Filtered variants in extended coding regions

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr1	19570175	.	G	A	A	stop_gained	HIGH	*EMC1*	ENST00000375199.3	protein_coding
chr1	19570175	.	G	A	A	stop_gained	HIGH	*EMC1*	ENST00000375208.3	protein_coding
chr1	19570175	.	G	A	A	3_prime_UTR_variant & NMD_transcript_variant	MODIFIER	*EMC1*	ENST00000475079.1	nonsense_mediated_decay
chr1	19570175	.	G	A	A	stop_gained	HIGH	*EMC1*	ENST00000477853.1	protein_coding
chr1	20440717	rs138535421	C	T	T	missense_variant	MODERATE	*PLA2G2D*	ENST00000375105.3	protein_coding
chr1	210415135	rs138093906	G	A	A	missense_variant	MODERATE	*SERTAD4*	ENST00000367012.3	protein_coding
chr1	210415135	rs138093906	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*SERTAD4*	ENST00000490620.1	processed_transcript
chr2	12858444	rs55813198	A	G	G	missense_variant	MODERATE	*TRIB2*	ENST00000155926.4	protein_coding
chr2	12858444	rs55813198	A	G	G	missense_variant	MODERATE	*TRIB2*	ENST00000405331.3	protein_coding
chr2	96521107	rs115953176	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*ANKRD36C*	ENST00000488721.1	retained_intron
chr3	33868210	.	GT	GTT, G	TT	non_coding_transcript_exon_variant	MODIFIER	*PDCD6IP*	ENST00000459659.1	retained_intron
chr3	33868210	.	GT	GTT, G	-	non_coding_transcript_exon_variant	MODIFIER	*PDCD6IP*	ENST00000459659.1	retained_intron
chr3	75779389	rs145333558	C	A	A	3_prime_UTR_variant	MODIFIER	*ZNF717*	ENST00000477374.1	protein_coding
chr3	75779540	rs77298160	A	G	G	3_prime_UTR_variant	MODIFIER	*ZNF717*	ENST00000477374.1	protein_coding
chr3	195594250	rs373134167	T	C	C	non_coding_transcript_exon_variant	MODIFIER	*TNK2*	ENST00000420716.2	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr3	195595745	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TNK2*	ENST00000420716.2	retained_intron
chr3	195595745	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TNK2*	ENST00000464041.1	retained_intron
chr3	195595745	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TNK2*	ENST00000481865.1	retained_intron
chr3	196612576	.	A	G	G	missense_variant	MODERATE	*SENP5*	ENST00000323460.5	protein_coding
chr3	196612576	.	A	G	G	missense_variant	MODERATE	*SENP5*	ENST00000445299.2	protein_coding
chr5	34925312	.	CT	CTT,C	-	non_coding_transcript_exon_variant	MODIFIER	*BRIX1*	ENST00000515798.1	retained_intron
chr5	99871359	rs139985000	G	A	A	missense_variant	MODERATE	*FAM174A*	ENST00000312637.4	protein_coding
chr5	99871359	rs139985000	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*FAM174A*	ENST00000509040.1	processed_transcript
chr5	107716450	.	G	C	C	5_prime_UTR_variant	MODIFIER	*FBXL17*	ENST00000359660.5	protein_coding
chr5	107716450	.	G	C	C	non_coding_transcript_exon_variant	MODIFIER	*FBXL17*	ENST00000518486.1	retained_intron
chr5	107716450	.	G	C	C	missense_variant	MODERATE	*FBXL17*	ENST00000542267.1	protein_coding
chr5	166711862	rs115605129	G	A	A	missense_variant	MODERATE	*TENM2*	ENST00000518659.1	protein_coding
chr5	166711862	rs115605129	G	A	A	missense_variant	MODERATE	*TENM2*	ENST00000545108.1	protein_coding
chr5	167921700	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*RARS*	ENST00000519346.1	retained_intron
chr5	167921700	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*RARS*	ENST00000521939.1	retained_intron
chr5	169361686	.	G	T	T	missense_variant	MODERATE	*DOCK2*	ENST00000519628.1	protein_coding
chr5	169678744	rs139974334	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*C5orf58*	ENST00000517575.1	processed_transcript

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr5	169678744	rs139974334	G	A	A	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*C5orf58*	ENST00000524171.1	nonsense_mediated_decay
chr6	24145896	rs11544636	C	T	T	missense_variant	MODERATE	*NRSN1*	ENST00000378477.2	protein_coding
chr6	24145896	rs11544636	C	T	T	missense_variant	MODERATE	*NRSN1*	ENST00000378491.4	protein_coding
chr6	24798959	rs75474054	CCTT	C	-	inframe_deletion	MODERATE	*C6orf229*	ENST00000565469.1	protein_coding
chr6	24861193	rs35730227	A	G	G	splice_region_variant&intron_variant	LOW	*FAM65B*	ENST00000259698.4	protein_coding
chr6	24861193	rs35730227	A	G	G	splice_region_variant&intron_variant	LOW	*FAM65B*	ENST00000378023.4	protein_coding
chr6	24861193	rs35730227	A	G	G	splice_region_variant&intron_variant	LOW	*FAM65B*	ENST00000510784.2	protein_coding
chr6	24861193	rs35730227	A	G	G	splice_region_variant&intron_variant	LOW	*FAM65B*	ENST00000538035.1	protein_coding
chr6	24861193	rs35730227	A	G	G	splice_region_variant&intron_variant	LOW	*FAM65B*	ENST00000540914.1	protein_coding
chr6	29430661	rs180989090	G	T	T	3_prime_UTR_variant	MODIFIER	*OR2H1*	ENST00000377133.1	protein_coding
chr6	29430661	rs180989090	G	T	T	3_prime_UTR_variant	MODIFIER	*OR2H1*	ENST00000377136.1	protein_coding
chr6	29430661	rs180989090	G	T	T	3_prime_UTR_variant	MODIFIER	*OR2H1*	ENST00000442615.1	protein_coding
chr6	29911925	rs149288080	C	T	T	missense_variant	MODERATE	*HLA-A*	ENST00000376802.2	protein_coding
chr6	29911925	rs149288080	C	T	T	missense_variant	MODERATE	*HLA-A*	ENST00000376806.5	protein_coding
chr6	29911925	rs149288080	C	T	T	missense_variant	MODERATE	*HLA-A*	ENST00000376809.5	protein_coding
chr6	29911925	rs149288080	C	T	T	missense_variant	MODERATE	*HLA-A*	ENST00000396634.1	protein_coding
chr6	29911925	rs149288080	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*HLA-A*	ENST00000461903.1	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr6	29911925	rs149288080	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*HLA-A*	ENST00000479320.1	retained_intron
chr6	29911925	rs149288080	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*HLA-A*	ENST00000495183.1	retained_intron
chr6	29911925	rs149288080	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*HLA-A*	ENST00000496081.1	retained_intron
chr6	30314060	rs17194859	T	G	G	non_coding_transcript_exon_variant	MODIFIER	*RPP21*	ENST00000498414.1	retained_intron
chr6	30529605	rs114406793	C	T	T	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant	LOW	*PRR3*	ENST00000376557.3	protein_coding
chr6	30529605	rs114406793	C	T	T	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant	LOW	*PRR3*	ENST00000376560.3	protein_coding
chr6	30529605	rs114406793	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*PRR3*	ENST00000470703.1	retained_intron
chr6	30529605	rs114406793	C	T	T	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant&non_coding_transcript_variant	LOW	*PRR3*	ENST00000481741.1	processed_transcript
chr6	30529605	rs114406793	C	T	T	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant&non_coding_transcript_variant	LOW	*PRR3*	ENST00000498336.1	processed_transcript
chr6	30671229	rs28994875	C	T	T	missense_variant	MODERATE	*MDC1*	ENST00000376405.2	protein_coding
chr6	30671229	rs28994875	C	T	T	missense_variant	MODERATE	*MDC1*	ENST00000376406.3	protein_coding
chr6	30671229	rs28994875	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*MDC1*	ENST00000489540.1	retained_intron
chr6	31499229	rs74367810	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*DDX39B*	ENST00000474961.1	retained_intron
chr6	31499229	rs74367810	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*DDX39B*	ENST00000478365.1	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr6	31514280	.	A	T	T	non_coding_transcript_exon_variant	MODIFIER	*ATP6V1G2-DDX39B*	ENST00000475917.1	processed_transcript
chr6	31514280	.	A	T	T	non_coding_transcript_exon_variant	MODIFIER	*ATP6V1G2*	ENST00000481998.1	retained_intron
chr6	31540693	rs3093542	G	C	C	non_coding_transcript_exon_variant	MODIFIER	*LTA*	ENST00000471842.1	retained_intron
chr6	31540693	rs3093542	G	C	C	non_coding_transcript_exon_variant	MODIFIER	*LTA*	ENST00000489638.1	retained_intron
chr6	31604670	rs377006285	G	A	A	missense_variant	MODERATE	*PRRC2A*	ENST00000376007.4	protein_coding
chr6	31604670	rs377006285	G	A	A	missense_variant	MODERATE	*PRRC2A*	ENST00000376033.2	protein_coding
chr6	31604670	rs377006285	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*PRRC2A*	ENST00000462617.1	retained_intron
chr6	31604670	rs377006285	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*PRRC2A*	ENST00000482441.1	retained_intron
chr6	31604670	rs377006285	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*PRRC2A*	ENST00000492691.1	retained_intron
chr6	31692387	.	G	T	T	missense_variant	MODERATE	*C6orf25*	ENST00000375805.2	protein_coding
chr6	31692387	.	G	T	T	missense_variant	MODERATE	*C6orf25*	ENST00000375806.2	protein_coding
chr6	31692387	.	G	T	T	missense_variant	MODERATE	*C6orf25*	ENST00000375809.3	protein_coding
chr6	31692387	.	G	T	T	missense_variant	MODERATE	*C6orf25*	ENST00000375810.4	protein_coding
chr6	31692387	.	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*C6orf25*	ENST00000460663.1	retained_intron
chr6	31692387	.	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*C6orf25*	ENST00000471545.1	retained_intron
chr6	31692387	.	G	T	T	missense_variant	MODERATE	*C6orf25*	ENST00000480039.1	protein_coding
chr6	31778953	.	C	T	T	missense_variant	MODERATE	*HSPA1L*	ENST00000375654.4	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr6	31778953	.	C	T	T	missense_variant	MODERATE	*HSPA1L*	ENST00000417199.3	protein_coding
chr6	31795619	rs11576012	C	T	T	5_prime_UTR_variant	MODIFIER	*HSPA1B*	ENST00000375650.3	protein_coding
chr6	31917291	rs2072634	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*CFB*	ENST00000452035.2	retained_intron
chr6	31917291	rs2072634	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*CFB*	ENST00000461483.1	retained_intron
chr6	31917291	rs2072634	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*CFB*	ENST00000465750.1	retained_intron
chr6	31917291	rs2072634	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*CFB*	ENST00000467150.1	retained_intron
chr6	31917291	rs2072634	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*CFB*	ENST00000497841.1	processed_transcript
chr6	31947349	rs141843362	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*STK19*	ENST00000463823.1	processed_transcript
chr6	31947349	rs141843362	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*STK19*	ENST00000471028.1	retained_intron
chr6	31947349	rs141843362	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*STK19*	ENST00000473983.1	retained_intron
chr6	31947349	rs141843362	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*STK19*	ENST00000478486.1	retained_intron
chr6	31947349	rs141843362	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*STK19*	ENST00000484540.1	retained_intron
chr6	32030129	rs140304758	C	T	T	missense_variant	MODERATE	*TNXB*	ENST00000375244.3	protein_coding
chr6	32030129	rs140304758	C	T	T	missense_variant	MODERATE	*TNXB*	ENST00000375247.2	protein_coding
chr6	32065113	rs61746206	C	T	T	missense_variant	MODERATE	*TNXB*	ENST00000375244.3	protein_coding
chr6	32065113	rs61746206	C	T	T	missense_variant	MODERATE	*TNXB*	ENST00000375247.2	protein_coding
chr6	32065113	rs61746206	C	T	T	missense_variant	MODERATE	*TNXB*	ENST00000479795.1	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr6	32065113	rs61746206	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TNXB*	ENST00000486148.1	retained_intron
chr6	32135181	rs138835391	G	T	T	missense_variant	MODERATE	*EGFL8*	ENST00000333845.6	protein_coding
chr6	32135181	rs138835391	G	T	T	missense_variant	MODERATE	*EGFL8*	ENST00000395512.1	protein_coding
chr6	32135181	rs138835391	G	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*PPT2-EGFL8*	ENST00000422437.1	nonsense_mediated_decay
chr6	32135181	rs138835391	G	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*PPT2-EGFL8*	ENST00000428388.2	nonsense_mediated_decay
chr6	32135181	rs138835391	G	T	T	missense_variant	MODERATE	*EGFL8*	ENST00000432129.1	protein_coding
chr6	32135181	rs138835391	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*EGFL8*	ENST00000466239.1	retained_intron
chr6	32374865	rs41395046	T	C	C	5_prime_UTR_variant	MODIFIER	*BTNL2*	ENST00000544175.1	protein_coding
chr6	32628130	rs375217182	A	G	G	splice_acceptor_variant&non_coding_transcript_variant	HIGH	*HLA-DQB1-AS1*	ENST00000419852.1	antisense
chr6	32709219	rs147432782	A	G	G	5_prime_UTR_variant	MODIFIER	*HLA-DQA2*	ENST00000374940.3	protein_coding
chr6	33384572	rs200107088	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*CUTA*	ENST00000487637.1	retained_intron
chr6	34214525	rs141610597	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*C6orf1*	ENST00000463083.1	retained_intron
chr6	36263171	rs45524833	G	A	A	missense_variant	MODERATE	*PNPLA1*	ENST00000312917.5	protein_coding
chr6	36263171	rs45524833	G	A	A	missense_variant	MODERATE	*PNPLA1*	ENST00000388715.3	protein_coding
chr6	36263171	rs45524833	G	A	A	missense_variant	MODERATE	*PNPLA1*	ENST00000394571.2	protein_coding
chr6	36263171	rs45524833	G	A	A	missense_variant	MODERATE	*PNPLA1*	ENST00000457797.1	protein_coding
chr6	36954044	.	G	T	T	5_prime_UTR_variant	MODIFIER	*MTCH1*	ENST00000373616.5	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr6	36954044	.	G	T	T	5_prime_UTR_variant	MODIFIER	*MTCH1*	ENST00000373627.5	protein_coding
chr7	1588948	rs144522592	T	TTGGACCTTCGCAGGGTCCAAGGGTGAGCTGTG	TGGACCTTCGCAGGGGTCCAAGGGGTGAGCTGTG	non_coding_transcript_exon_variant	MODIFIER	*TMEM184A*	ENST00000468535.1	retained_intron
chr7	21934511	rs72658825	G	A	A	missense_variant	MODERATE	*DNAH11*	ENST00000328843.6	protein_coding
chr7	21934511	rs72658825	G	A	A	missense_variant	MODERATE	*DNAH11*	ENST00000409508.3	protein_coding
chr7	21934511	rs72658825	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*DNAH11*	ENST00000479878.1	retained_intron
chr7	21942676	.	A	T	T	missense_variant	MODERATE	*CDCA7L*	ENST00000435717.1	protein_coding
chr7	21942676	.	A	T	T	non_coding_transcript_exon_variant	MODIFIER	*CDCA7L*	ENST00000488845.1	retained_intron
chr7	29927689	rs61750795	G	A	A	missense_variant	MODERATE	*WIPF3*	ENST00000242140.5	protein_coding
chr7	29927689	rs61750795	G	A	A	missense_variant	MODERATE	*WIPF3*	ENST00000409123.1	protein_coding
chr7	29927689	rs61750795	G	A	A	missense_variant	MODERATE	*WIPF3*	ENST00000409290.1	protein_coding
chr7	29954275	rs35346638	C	G	G	3_prime_UTR_variant	MODIFIER	*WIPF3*	ENST00000409290.1	protein_coding
chr7	30486489	.	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*NOD1*	ENST00000489614.1	retained_intron
chr7	30693024	rs8192493	A	T	T	3_prime_UTR_variant	MODIFIER	*CRHR2*	ENST00000341843.4	protein_coding
chr7	30693024	rs8192493	A	T	T	3_prime_UTR_variant	MODIFIER	*CRHR2*	ENST00000348 43	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
									8.4	
chr7	30693024	rs8192493	A	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*CRHR2*	ENST00000452278.1	nonsense_mediated_decay
chr7	30693024	rs8192493	A	T	T	3_prime_UTR_variant	MODIFIER	*CRHR2*	ENST00000471646.1	protein_coding
chr7	30963192	rs144238098	A	G	G	missense_variant	MODERATE	*AQP1*	ENST00000311813.4	protein_coding
chr7	30963192	rs144238098	A	G	G	missense_variant	MODERATE	*AQP1*	ENST00000409611.1	protein_coding
chr7	30963192	rs144238098	A	G	G	missense_variant	MODERATE	*AQP1*	ENST00000409899.1	protein_coding
chr7	30963192	rs144238098	A	G	G	missense_variant	MODERATE	*AQP1*	ENST00000434909.2	protein_coding
chr7	30963192	rs144238098	A	G	G	missense_variant	MODERATE	*AQP1*	ENST00000441328.2	protein_coding
chr7	30963192	rs144238098	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*AQP1*	ENST00000482461.1	processed_transcript
chr7	30963192	rs144238098	A	G	G	missense_variant	MODERATE	*AQP1*	ENST00000509504.1	protein_coding
chr7	93540255	rs77925896	A	C	C	3_prime_UTR_variant	MODIFIER	*GNGT1*	ENST00000248572.5	protein_coding
chr7	93540255	rs77925896	A	C	C	3_prime_UTR_variant	MODIFIER	*GNGT1*	ENST00000429473.1	protein_coding
chr7	93540255	rs77925896	A	C	C	3_prime_UTR_variant	MODIFIER	*GNGT1*	ENST00000455502.1	protein_coding
chr7	94164787	rs146953436	C	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*CASD1*	ENST00000443644.1	nonsense_mediated_decay
chr7	97862665	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TECPR1*	ENST00000490842.1	retained_intron
chr7	99055969	rs144185727	C	T	T	missense_variant	MODERATE	*ATP5J2*	ENST00000292475.3	protein_coding
chr7	99055969	rs144185727	C	T	T	missense_variant	MODERATE	*ATP5J2*	ENST00000359832.4	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr7	99055969	rs144185727	C	T	T	missense_variant	MODERATE	*ATP5J2*	ENST00000394186.3	protein_coding
chr7	99055969	rs144185727	C	T	T	missense_variant&NMD_transcript_variant	MODERATE	*ATP5J2*	ENST00000414062.1	nonsense_mediated_decay
chr7	99055969	rs144185727	C	T	T	missense_variant	MODERATE	*ATP5J2*	ENST00000449683.1	protein_coding
chr7	99055969	rs144185727	C	T	T	missense_variant	MODERATE	*ATP5J2*	ENST00000488775.1	protein_coding
chr7	99055969	rs144185727	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*ATP5J2*	ENST00000491560.1	retained_intron
chr7	99055969	rs144185727	C	T	T	missense_variant	MODERATE	*ATP5J2*	ENST00000544611.1	protein_coding
chr7	99689283	rs139971845	C	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*COPS6*	ENST00000426712.1	nonsense_mediated_decay
chr7	99689283	rs139971845	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*COPS6*	ENST00000474823.1	retained_intron
chr7	99705258	rs375667816	G	A	A	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*AP4M1*	ENST00000416938.1	nonsense_mediated_decay
chr8	69381053	rs61736081	G	A	A	missense_variant&splice_region_variant	MODERATE	*C8orf34*	ENST00000337103.4	protein_coding
chr8	69381053	rs61736081	G	A	A	missense_variant&splice_region_variant	MODERATE	*C8orf34*	ENST00000348340.2	protein_coding
chr8	69381053	rs61736081	G	A	A	missense_variant&splice_region_variant	MODERATE	*C8orf34*	ENST00000518698.1	protein_coding
chr8	69381053	rs61736081	G	A	A	missense_variant&splice_region_variant&NMD_transcript_variant	MODERATE	*C8orf34*	ENST00000521406.1	nonsense_mediated_decay
chr8	69381053	rs61736081	G	A	A	missense_variant&splice_region_variant	MODERATE	*C8orf34*	ENST00000539993.1	protein_coding
chr8	104342201	rs199954064	G	A	A	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*FZD6*	ENST00000519011.1	nonsense_mediated_decay
chr8	104342201	rs199954064	G	A	A	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*FZD6*	ENST00000521195.1	nonsense_mediated_decay
chr8	104342201	rs199954064	G	A	A	3_prime_UTR_variant&NM	MODIFIER	*FZD6*	ENST00000522482	nonsense_mediate

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
						D_transcript_variant			4.1	d_decay
chr8	109240531	.	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*EIF3E*	ENST00000519517.1	processed_transcript
chr8	109240531	.	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*EIF3E*	ENST00000521614.1	retained_intron
chr8	109240531	.	G	A	A	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*EIF3E*	ENST00000522445.1	nonsense_mediated_decay
chr8	109240531	.	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*EIF3E*	ENST00000523646.1	retained_intron
chr8	110505963	rs118053060	A	G	G	missense_variant	MODERATE	*PKHD1L1*	ENST00000378402.5	protein_coding
chr8	110505963	rs118053060	A	G	G	missense_variant	MODERATE	*PKHD1L1*	ENST00000526472.1	protein_coding
chr9	27197486	rs35030851	G	T	T	missense_variant	MODERATE	*TEK*	ENST00000380036.4	protein_coding
chr9	27197486	rs35030851	G	T	T	missense_variant	MODERATE	*TEK*	ENST00000406359.4	protein_coding
chr9	27197486	rs35030851	G	T	T	missense_variant	MODERATE	*TEK*	ENST00000519080.1	protein_coding
chr9	27197486	rs35030851	G	T	T	missense_variant	MODERATE	*TEK*	ENST00000519097.1	protein_coding
chr9	33948436	.	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*UBAP2*	ENST00000462799.1	processed_transcript
chr9	35042296	rs61742471	G	A	A	missense_variant	MODERATE	*C9orf131*	ENST00000312292.5	protein_coding
chr9	35042296	rs61742471	G	A	A	missense_variant	MODERATE	*C9orf131*	ENST00000378745.3	protein_coding
chr9	95782662	.	G	A	A	missense_variant	MODERATE	*FGD3*	ENST00000337352.6	protein_coding
chr9	95782662	.	G	A	A	missense_variant	MODERATE	*FGD3*	ENST00000375482.3	protein_coding
chr9	95782662	.	G	A	A	missense_variant	MODERATE	*FGD3*	ENST00000416701.2	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr9	95782662	.	G	A	A	missense_variant&NMD_transcript_variant	MODERATE	*FGD3*	ENST00000467786.1	nonsense_mediated_decay
chr9	95782662	.	G	A	A	missense_variant	MODERATE	*FGD3*	ENST00000538555.1	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000297954.4	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000349097.3	protein_coding
chr9	96051100	rs41296061	C	T	T	5_prime_UTR_variant	MODIFIER	*WNK2*	ENST00000356055.3	protein_coding
chr9	96051100	rs41296061	C	T	T	3_prime_UTR_variant	MODIFIER	*WNK2*	ENST00000395475.2	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000395477.2	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000411624.1	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000427277.2	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000432730.1	protein_coding
chr9	96051100	rs41296061	C	T	T	missense_variant	MODERATE	*WNK2*	ENST00000448251.1	protein_coding
chr9	98734844	rs41281212	A	G	G	5_prime_UTR_variant	MODIFIER	*ERCC6L2*	ENST00000407474.3	protein_coding
chr9	98734844	rs41281212	A	G	G	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*ERCC6L2*	ENST00000479391.2	nonsense_mediated_decay
chr9	99699577	rs201094122	G	A	A	missense_variant	MODERATE	*NUTM2G*	ENST00000354649.3	protein_coding
chr9	99699577	rs201094122	G	A	A	missense_variant	MODERATE	*NUTM2G*	ENST00000372322.3	protein_coding
chr9	132637691	rs35071307	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*USP20*	ENST00000474895.2	processed_transcript
chr9	132652630	rs149605597	G	A	A	3_prime_UTR_variant	MODIFIER	*FNBP1*	ENST00000420781.1	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr9	132652630	rs149605597	G	A	A	3_prime_UTR_variant	MODIFIER	*FNBP1*	ENST00000446176.2	protein_coding
chr11	16339873	rs72869972	C	A	A	non_coding_transcript_exon_variant	MODIFIER	*SOX6*	ENST00000533658.1	processed_transcript
chr11	19955194	rs16937251	G	C	C	missense_variant	MODERATE	*NAV2*	ENST00000349880.4	protein_coding
chr11	19955194	rs16937251	G	C	C	missense_variant	MODERATE	*NAV2*	ENST00000360655.4	protein_coding
chr11	19955194	rs16937251	G	C	C	missense_variant	MODERATE	*NAV2*	ENST00000396085.1	protein_coding
chr11	19955194	rs16937251	G	C	C	missense_variant	MODERATE	*NAV2*	ENST00000396087.3	protein_coding
chr11	19955194	rs16937251	G	C	C	missense_variant	MODERATE	*NAV2*	ENST00000527559.2	protein_coding
chr11	19955194	rs16937251	G	C	C	non_coding_transcript_exon_variant	MODIFIER	*NAV2*	ENST00000528008.1	processed_transcript
chr11	19955194	rs16937251	G	C	C	missense_variant	MODERATE	*NAV2*	ENST00000540292.1	protein_coding
chr11	25098911	rs116821457	C	T	T	missense_variant	MODERATE	*LUZP2*	ENST00000336930.6	protein_coding
chr11	25098911	rs116821457	C	T	T	missense_variant	MODERATE	*LUZP2*	ENST00000533227.1	protein_coding
chr11	82895826	rs118001085	A	G	G	missense_variant	MODERATE	*PCF11*	ENST00000298281.4	protein_coding
chr11	82895826	rs118001085	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*RP11-727A23.5*	ENST00000602381.1	lincRNA
chr11	111789512	rs139772777	T	C	C	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*HSPB2-C11orf52*	ENST00000527616.1	nonsense_mediated_decay
chr11	111789512	rs139772777	T	C	C	5_prime_UTR_variant	MODIFIER	*C11orf52*	ENST00000529342.1	protein_coding
chr11	111789512	rs139772777	T	C	C	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*HSPB2-C11orf52*	ENST00000534100.1	nonsense_mediated_decay
chr11	111789512	rs139772777	T	C	C	3_prime_UTR_variant	MODIFIER	*HSPB2*	ENST00000537382.1	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr12	75900396	rs35071517	CAAAAA	CAAAA,CAAAAAA,CAAAAAAAAAAAA,C	-	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant	LOW	*KRR1*	ENST00000229214.4	protein_coding
chr12	75900396	rs35071517	CAAAAA	CAAAA,CAAAAAA,CAAAAAAAAAAAA,C	-	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant	LOW	*KRR1*	ENST00000438169.2	protein_coding
chr12	75900396	rs35071517	CAAAAA	CAAAA,CAAAAAA,CAAAAAAAAAAAA,C	-	splice_region_variant&splice_polypyrimidine_tract_variant&intron_variant&non_coding_transcript_variant	LOW	*KRR1*	ENST00000550023.1	retained_intron
chr12	75900396	rs35071517	CAAAAA	CAAAA,CAAAAAA,CAAAAAAAAAAAA,C	AAAA	non_coding_transcript_exon_variant	MODIFIER	*KRR1*	ENST00000550898.1	retained_intron
chr12	75900396	rs35071517	CAAAAA	CAAAA,CA	AAAAAA	non_coding_transcript_exon_variant	MODIFIER	*KRR1*	ENST00000550898.1	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
				AAAAA,CAAAAAAAAAAAA,C						
chr12	75900396	rs35071517	CAAAAA	CAAAA,CAAAAA,CAAAAAAAAAAAA,C	AAAAAAAAAAAAA	non_coding_transcript_exon_variant	MODIFIER	*KRR1*	ENST00000550898.1	retained_intron
chr12	75900396	rs35071517	CAAAAA	CAAAA,CAAAAA,CAAAAAAAAAAAA,C	-	non_coding_transcript_exon_variant	MODIFIER	*KRR1*	ENST00000550898.1	retained_intron
chr12	111785515	rs61745424	G	A	A	missense_variant	MODERATE	*CUX2*	ENST00000261726.6	protein_coding
chr12	112369505	rs147550432	G	T	T	missense_variant	MODERATE	*TMEM116*	ENST00000354825.3	protein_coding
chr12	112369505	rs147550432	G	T	T	missense_variant	MODERATE	*TMEM116*	ENST00000355445.3	protein_coding
chr12	112369505	rs147550432	G	T	T	missense_variant	MODERATE	*TMEM116*	ENST00000437003.2	protein_coding
chr12	112369505	rs147550432	G	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*TMEM116*	ENST00000547878.2	nonsense_mediated_decay
chr12	112369505	rs147550432	G	T	T	missense_variant	MODERATE	*TMEM116*	ENST00000549537.2	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr12	112369505	rs147550432	G	T	T	missense_variant	MODERATE	*TMEM116*	ENST00000550831.3	protein_coding
chr12	112369505	rs147550432	G	T	T	missense_variant	MODERATE	*TMEM116*	ENST00000552374.2	protein_coding
chr12	112369505	rs147550432	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*TMEM116*	ENST00000552801.2	retained_intron
chr12	113346789	.	T	A	A	3_prime_UTR_variant	MODIFIER	*OAS1*	ENST00000553185.1	protein_coding
chr14	81743565	rs36017951	T	C	C	missense_variant	MODERATE	*STON2*	ENST00000267540.2	protein_coding
chr14	81743565	rs36017951	T	C	C	non_coding_transcript_exon_variant	MODIFIER	*STON2*	ENST00000555284.1	retained_intron
chr14	81743565	rs36017951	T	C	C	missense_variant	MODERATE	*STON2*	ENST00000555447.1	protein_coding
chr15	64463979	rs181304652	C	G	G	3_prime_UTR_variant	MODIFIER	*CSNK1G1*	ENST00000303052.7	protein_coding
chr15	64463979	rs181304652	C	G	G	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*CSNK1G1*	ENST00000606225.1	nonsense_mediated_decay
chr15	65903547	.	G	A	A	5_prime_UTR_variant	MODIFIER	*VWA9*	ENST00000313182.2	protein_coding
chr15	65903547	.	G	A	A	5_prime_UTR_variant	MODIFIER	*VWA9*	ENST00000420799.2	protein_coding
chr15	65903547	.	G	A	A	5_prime_UTR_variant	MODIFIER	*VWA9*	ENST00000431261.2	protein_coding
chr15	65903547	.	G	A	A	5_prime_UTR_variant	MODIFIER	*VWA9*	ENST00000442903.3	protein_coding
chr15	65903547	.	G	A	A	5_prime_UTR_variant	MODIFIER	*VWA9*	ENST00000567923.1	protein_coding
chr15	67390985	rs189286879	G	A	A	5_prime_UTR_variant	MODIFIER	*SMAD3*	ENST00000559092.1	protein_coding
chr15	75641682	rs5745908	T	C	C	splice_donor_variant	HIGH	*NEIL1*	ENST00000355059.4	protein_coding
chr15	75641682	rs5745908	T	C	C	splice_donor_variant&non_coding_transcript_variant	HIGH	*NEIL1*	ENST00000561643.1	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr15	75641682	rs5745908	T	C	C	splice_donor_variant&non_coding_transcript_variant	HIGH	*NEIL1*	ENST00000564738.1	retained_intron
chr15	75641682	rs5745908	T	C	C	splice_donor_variant	HIGH	*NEIL1*	ENST00000564784.1	protein_coding
chr15	75641682	rs5745908	T	C	C	non_coding_transcript_exon_variant	MODIFIER	*NEIL1*	ENST00000564951.1	retained_intron
chr15	75641682	rs5745908	T	C	C	splice_donor_variant	HIGH	*NEIL1*	ENST00000565051.1	protein_coding
chr15	75641682	rs5745908	T	C	C	splice_donor_variant	HIGH	*NEIL1*	ENST00000567005.1	protein_coding
chr15	75641682	rs5745908	T	C	C	splice_donor_variant	HIGH	*NEIL1*	ENST00000567657.1	protein_coding
chr15	75641682	rs5745908	T	C	C	splice_donor_variant&NMD_transcript_variant	HIGH	*NEIL1*	ENST00000568519.1	nonsense_mediated_decay
chr15	75641682	rs5745908	T	C	C	splice_donor_variant	HIGH	*NEIL1*	ENST00000569035.1	protein_coding
chr15	78880752	rs2229961	G	A	A	missense_variant	MODERATE	*CHRNA5*	ENST00000299565.5	protein_coding
chr15	78880752	rs2229961	G	A	A	missense_variant	MODERATE	*CHRNA5*	ENST00000394802.4	protein_coding
chr15	78880752	rs2229961	G	A	A	missense_variant	MODERATE	*CHRNA5*	ENST00000559554.1	protein_coding
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000262510.6	protein_coding
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000308149.7	protein_coding
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000399221.3	protein_coding
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000436936.1	protein_coding
chr16	57093436	rs117950337	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*NLRC5*	ENST00000535284.1	retained_intron
chr16	57093436	rs117950337	G	T	T	missense_variant&NMD_transcript_variant	MODERATE	*NLRC5*	ENST00000537056.1	nonsense_mediated_decay

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000538110.1	protein_coding
chr16	57093436	rs117950337	G	T	T	missense_variant&NMD_transcript_variant	MODERATE	*NLRC5*	ENST00000538453.1	nonsense_mediated_decay
chr16	57093436	rs117950337	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*NLRC5*	ENST00000538778.1	retained_intron
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000538805.1	protein_coding
chr16	57093436	rs117950337	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*NLRC5*	ENST00000538930.1	retained_intron
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000539144.1	protein_coding
chr16	57093436	rs117950337	G	T	T	missense_variant&NMD_transcript_variant	MODERATE	*NLRC5*	ENST00000540182.1	nonsense_mediated_decay
chr16	57093436	rs117950337	G	T	T	missense_variant	MODERATE	*NLRC5*	ENST00000543030.1	protein_coding
chr16	57093436	rs117950337	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*NLRC5*	ENST00000543402.1	retained_intron
chr16	57093436	rs117950337	G	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*NLRC5*	ENST00000545081.1	nonsense_mediated_decay
chr16	74664698	rs78796563	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*RFWD3*	ENST00000575154.1	retained_intron
chr16	74706385	rs76652394	G	A	A	3_prime_UTR_variant	MODIFIER	*MLKL*	ENST00000306247.7	protein_coding
chr16	74706385	rs76652394	G	A	A	3_prime_UTR_variant	MODIFIER	*MLKL*	ENST00000308807.7	protein_coding
chr16	74706385	rs76652394	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*MLKL*	ENST00000570846.1	retained_intron
chr16	74706385	rs76652394	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*MLKL*	ENST00000575695.1	retained_intron
chr16	74706385	rs76652394	G	A	A	3_prime_UTR_variant	MODIFIER	*MLKL*	ENST00000576529.1	protein_coding
chr16	74709233	rs144526386	C	A	A	missense_variant	MODERATE	*MLKL*	ENST00000306247.7	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr16	74709233	rs144526386	C	A	A	non_coding_transcript_exon_variant	MODIFIER	*MLKL*	ENST00000575695.1	retained_intron
chr16	74709233	rs144526386	C	A	A	missense_variant	MODERATE	*MLKL*	ENST00000576529.1	protein_coding
chr16	86585743	rs34005514	G	C	C	missense_variant	MODERATE	*MTHFSD*	ENST00000322911.6	protein_coding
chr16	86585743	rs34005514	G	C	C	missense_variant	MODERATE	*MTHFSD*	ENST00000360900.6	protein_coding
chr16	86585743	rs34005514	G	C	C	5_prime_UTR_variant	MODIFIER	*MTHFSD*	ENST00000561522.1	protein_coding
chr16	86585743	rs34005514	G	C	C	non_coding_transcript_exon_variant	MODIFIER	*MTHFSD*	ENST00000561848.1	processed_transcript
chr16	86585743	rs34005514	G	C	C	missense_variant	MODERATE	*MTHFSD*	ENST00000561989.1	protein_coding
chr16	86585743	rs34005514	G	C	C	missense_variant	MODERATE	*MTHFSD*	ENST00000562940.2	protein_coding
chr16	86585743	rs34005514	G	C	C	5_prime_UTR_variant	MODIFIER	*MTHFSD*	ENST00000562994.1	protein_coding
chr16	86585743	rs34005514	G	C	C	missense_variant	MODERATE	*MTHFSD*	ENST00000565482.1	protein_coding
chr16	86585743	rs34005514	G	C	C	missense_variant&NMD_transcript_variant	MODERATE	*MTHFSD*	ENST00000566050.1	nonsense_mediated_decay
chr16	86585743	rs34005514	G	C	C	missense_variant	MODERATE	*MTHFSD*	ENST00000566469.1	protein_coding
chr16	89804166	rs201316239	G	A	A	3_prime_UTR_variant	MODIFIER	*FANCA*	ENST00000389301.3	protein_coding
chr16	89804166	rs201316239	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*ZNF276*	ENST00000569901.1	retained_intron
chr17	15584256	rs181088462	C	T	T	5_prime_UTR_variant	MODIFIER	*TRIM16*	ENST00000336708.7	protein_coding
chr17	15584256	rs181088462	C	T	T	missense_variant	MODERATE	*TRIM16*	ENST00000416464.2	protein_coding
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000460728.1	processed_transcript

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000473540.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000473659.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	5_prime_UTR_variant	MODIFIER	*TRIM16*	ENST00000578237.1	protein_coding
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000579272.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000579535.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000580388.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	missense_variant	MODERATE	*TRIM16*	ENST00000581200.1	protein_coding
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000581224.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000582182.2	processed_transcript
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000582708.1	processed_transcript
chr17	15584256	rs181088462	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TRIM16*	ENST00000584142.1	processed_transcript
chr17	34964937	rs141422978	C	G	G	3_prime_UTR_variant	MODIFIER	*MRM1*	ENST00000250156.7	protein_coding
chr17	34964937	rs141422978	C	G	G	3_prime_UTR_variant	MODIFIER	*MRM1*	ENST00000585770.1	protein_coding
chr18	59854641	rs142041492	GGCTGTGGAGGTGC	G	-	5_prime_UTR_variant	MODIFIER	*KIAA1468*	ENST00000256858.6	protein_coding
chr18	59854641	rs142041492	GGCTGTGGAGGTGC	G	-	5_prime_UTR_variant	MODIFIER	*KIAA1468*	ENST00000398130.2	protein_coding

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr18	59854641	rs142041492	GGCTGTGGAGGTGC	G	-	5_prime_UTR_variant	MODIFIER	*KIAA1468*	ENST00000587725.1	non_stop_decay
chr18	59899576	rs3764493	T	C	C	non_coding_transcript_exon_variant	MODIFIER	*KIAA1468*	ENST00000591227.1	processed_transcript
chr18	59899576	rs3764493	T	C	C	non_coding_transcript_exon_variant	MODIFIER	*KIAA1468*	ENST00000592479.1	processed_transcript
chr18	59925286	rs146220305	A	G	G	missense_variant	MODERATE	*KIAA1468*	ENST00000256858.6	protein_coding
chr18	59925286	rs146220305	A	G	G	missense_variant	MODERATE	*KIAA1468*	ENST00000398130.2	protein_coding
chr18	59925286	rs146220305	A	G	G	missense_variant	MODERATE	*KIAA1468*	ENST00000587725.1	non_stop_decay
chr19	505024	.	T	A	A	3_prime_UTR_variant	MODIFIER	*MADCAM1*	ENST00000215637.3	protein_coding
chr19	505024	.	T	A	A	3_prime_UTR_variant	MODIFIER	*MADCAM1*	ENST00000346144.4	protein_coding
chr19	505024	.	T	A	A	3_prime_UTR_variant	MODIFIER	*MADCAM1*	ENST00000587541.1	protein_coding
chr19	810573	rs11549883	C	A	A	non_coding_transcript_exon_variant	MODIFIER	*PTBP1*	ENST00000585856.1	retained_intron
chr19	810573	rs11549883	C	A	A	non_coding_transcript_exon_variant	MODIFIER	*PTBP1*	ENST00000586944.1	retained_intron
chr19	929481	.	C	CGTG	GTG	5_prime_UTR_variant	MODIFIER	*ARID3A*	ENST00000263620.3	protein_coding
chr19	929481	.	C	CGTG	GTG	non_coding_transcript_exon_variant	MODIFIER	*ARID3A*	ENST00000585895.1	processed_transcript
chr19	929481	.	C	CGTG	GTG	non_coding_transcript_exon_variant	MODIFIER	*ARID3A*	ENST00000592916.1	processed_transcript
chr19	1088580	.	C	T	T	3_prime_UTR_variant	MODIFIER	*POLR2E*	ENST00000215587.7	protein_coding
chr19	1088580	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*POLR2E*	ENST00000586215.1	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr19	1088580	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*POLR2E*	ENST00000586817.1	retained_intron
chr19	1088580	.	C	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*POLR2E*	ENST00000591767.1	nonsense_mediated_decay
chr19	1236045	rs199935954	CAGTGGT	C	-	inframe_deletion	MODERATE	*C19orf26*	ENST00000215376.6	protein_coding
chr19	1236045	rs199935954	CAGTGGT	C	-	inframe_deletion	MODERATE	*C19orf26*	ENST00000382477.2	protein_coding
chr19	1236045	rs199935954	CAGTGGT	C	-	inframe_deletion&NMD_transcript_variant	MODERATE	*C19orf26*	ENST00000589260.1	nonsense_mediated_decay
chr19	1236045	rs199935954	CAGTGGT	C	-	inframe_deletion	MODERATE	*C19orf26*	ENST00000590083.1	protein_coding
chr19	1806667	rs45574836	C	T	T	missense_variant	MODERATE	*ATP8B3*	ENST00000310127.6	protein_coding
chr19	1806667	rs45574836	C	T	T	missense_variant	MODERATE	*ATP8B3*	ENST00000525591.1	protein_coding
chr19	1806667	rs45574836	C	T	T	missense_variant	MODERATE	*ATP8B3*	ENST00000526092.2	protein_coding
chr19	1806667	rs45574836	C	T	T	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*ATP8B3*	ENST00000531925.1	nonsense_mediated_decay
chr19	1806667	rs45574836	C	T	T	missense_variant	MODERATE	*ATP8B3*	ENST00000533993.2	protein_coding
chr19	1806667	rs45574836	C	T	T	missense_variant	MODERATE	*ATP8B3*	ENST00000539485.1	protein_coding
chr19	2098207	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*IZUMO4*	ENST00000478879.1	retained_intron
chr19	2098207	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*IZUMO4*	ENST00000497445.1	retained_intron
chr19	2098207	.	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*IZUMO4*	ENST00000590985.1	retained_intron
chr19	2109834	rs758286	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*AP3D1*	ENST00000585652.1	retained_intron
chr19	2110989	rs115359236	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*AP3D1*	ENST00000585652.1	retained_intron

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr19	2110989	rs115359236	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*AP3D1*	ENST00000589223.1	retained_intron
chr19	2111131	rs76369653	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*AP3D1*	ENST00000585652.1	retained_intron
chr19	2111131	rs76369653	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*AP3D1*	ENST00000589223.1	retained_intron
chr19	2113284	rs372908018	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*AP3D1*	ENST00000592488.1	retained_intron
chr19	2214311	rs76972872	G	T	T	non_coding_transcript_exon_variant	MODIFIER	*AC004490.1*	ENST00000585593.1	antisense
chr19	2236097	rs191615952	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*PLEKHJ1*	ENST00000585423.1	retained_intron
chr19	2236097	rs191615952	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*PLEKHJ1*	ENST00000586497.1	retained_intron
chr19	2249583	rs61736572	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*AMH*	ENST00000609455.1	retained_intron
chr19	2249634	rs61736575	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*AMH*	ENST00000609455.1	retained_intron
chr19	2250236	rs77671243	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*AMH*	ENST00000589313.2	retained_intron
chr19	2399071	rs140728595	C	T	T	missense_variant	MODERATE	*TMPRSS9*	ENST00000332578.3	protein_coding
chr19	2399071	rs140728595	C	T	T	non_coding_transcript_exon_variant	MODIFIER	*TMPRSS9*	ENST00000395264.3	retained_intron
chr19	2427575	.	GTAAC	G	-	5_prime_UTR_variant	MODIFIER	*TIMM13*	ENST00000215570.3	protein_coding
chr19	2717164	rs146665806	G	A	A	3_prime_UTR_variant	MODIFIER	*DIRAS1*	ENST00000323469.4	protein_coding
chr19	2717164	rs146665806	G	A	A	3_prime_UTR_variant	MODIFIER	*DIRAS1*	ENST00000585334.1	protein_coding
chr19	2833692	rs36084282	A	G	G	non_coding_transcript_exon_variant	MODIFIER	*ZNF554*	ENST00000588534.1	retained_intron
chr19	2833692	rs36084282	A	G	G	3_prime_UTR_variant&NMD_transcript_variant	MODIFIER	*ZNF554*	ENST00000590116.1	nonsense_mediated_decay

Chr	Position	Rs	RA	AA	Allele	Consequence	Impact	Gene	Feature	Biotype
chr19	2833692	rs36084282	A	G	G	3_prime_UTR_variant	MODIFIER	*ZNF554*	ENST00000591265.1	protein_coding
chr19	30935704	rs61744158	C	A	A	missense_variant	MODERATE	*ZNF536*	ENST00000355537.3	protein_coding
chr19	30935704	rs61744158	C	A	A	missense_variant	MODERATE	*ZNF536*	ENST00000585628.1	protein_coding
chr19	31039669	rs77238711	C	T	T	missense_variant	MODERATE	*ZNF536*	ENST00000355537.3	protein_coding
chr19	41383115	rs111390860	G	A	A	missense_variant	MODERATE	*CYP2A7*	ENST00000291764.3	protein_coding
chr19	41383115	rs111390860	G	A	A	missense_variant	MODERATE	*CYP2A7*	ENST00000301146.4	protein_coding
chr19	41383115	rs111390860	G	A	A	non_coding_transcript_exon_variant	MODIFIER	*CYP2A7*	ENST00000598264.1	retained_intron
chr20	33432915	rs79530197	T	TCAC, TCACCAC	CAC	3_prime_UTR_variant	MODIFIER	*GGT7*	ENST00000336431.5	protein_coding
chr20	33432915	rs79530197	T	TCAC, TCACCAC	CACCAC	3_prime_UTR_variant	MODIFIER	*GGT7*	ENST00000336431.5	protein_coding
chr20	33432915	rs79530197	T	TCAC, TCACCAC	CAC	non_coding_transcript_exon_variant	MODIFIER	*GGT7*	ENST00000469018.1	processed_transcript
chr20	33432915	rs79530197	T	TCAC, TCACCAC	CACCAC	non_coding_transcript_exon_variant	MODIFIER	*GGT7*	ENST00000469018.1	processed_transcript
chr22	17265194	rs114989947	G	A	A	missense_variant	MODERATE	*XKR3*	ENST00000331428.5	protein_coding

Appendix L-Annotated gene list

Gene	Gene Function	Disorder/phenotype caused
AC004490.1	Putative uncharacterized protein ENSP00000394456 Fragment	/
AMH	Plays an important role in several reproductive functions inc sex differentiation and regression of Mullerian ducts in the male embryo	Persistent Mullerian Duct Syndrome, PCOS, Mixed Gonadal Dysgenesis
ANKRD36C	Ion channel inhibitor activity.	Immune-Mediated Thrombotic Thrombocytopenic Purpura 1
AP3D1	The AP-3 complex facilitates the budding of vesicles from the Golgi membrane, and may be directly involved in trafficking to lysosomes	Hermansky-Pudlak Syndrome, Albinism, Ocular
AP4M1	Involved in the recognition and sorting of intracellular cargo proteins	Spastic Paraplegia 50, Autosomal Recessive, Congenital Nervous System Abnormality, Non-Syndromic X-Linked Intellectual Disability 46
AQP1	Protein permits passive transport of water along an osmotic gradient.	Obstructive Hydrocephalus, Central Pontine Myelinolysis
ARID3A	Has in embryonic patterning, cell lineage gene regulation, cell cycle control, transcriptional regulation, and possibly in chromatin structure modification.	Flinders island spotted fever
ATP5J2	Produces subunit of mitochondrial ALT synthase- catalyzes ATP synthesis	/
ATP6V1G2	Signaling by Receptor Tyrosine Kinases and Insulin receptor recycling.	Distal Renal Tubular Acidosis , dilated cardiomyopathy
ATP6V1G2-DDX39B	Atpase-coupled transmembrane transporter activity	Hypertrophic and dilated Cardiomyopathy
ATP8B3	P4-atpase flippase which catalyzes the hydrolysis of ATP coupled to the transport of aminophospholipids from the outer to the inner leaflet of various membranes and ensures the maintenance of asymmetric distribution of phospholipids.	Superficial Keratitis and Cerebellar Ataxia, Mental Retardation, And Dysequilibrium Syndrome 1.
BRIX1	Enables RNA binding activity.	/

BTNL2	Involved in immune surveillance-decreasing T-cell proliferation and cytokine release.	Sarcoidosis 2, Ulcerative colitis
C11orf52	Uncharacterised protein	/
C19orf26	Predicted to be involved in negative regulation of calcium ion-dependent exocytosis and negative regulation of voltage-gated calcium channel activity	Sleeping Sickness
C5orf58	/	/
C6orf1	Predicted to be integral component of membrane.	/
C6orf229	/	/
C6orf25	Response to elevated platelet cytosolic Ca2+ and Microglia pathogen phagocytosis pathway.	Thrombocytopenia, Anemia, And Myelofibrosis, Myelofibrosis , Thrombocytopenia
C8orf34	Encodes a protein that is related to the cyclic AMP dependent protein kinase regulators	Familial Renal Papillary Carcinoma, Branchiootorenal Syndrome
C9orf131	/	/
CASD1	Involved in carbohydrate metabolic process.	Myoclonus-Dystonia
CDCA7L	Expressor that inhibits monoamine oxidase A (MAOA) activity and gene expression	Ciliary Dyskinesia, Dextrocardia With Situs Inversus , medulloblastoma
CFB	Serine-type endopeptidase activity and complement binding.	Hemolytic uremic syndrome, complement factor b deficiency, macular degeneration
CHRNA5	Mediate fast signal transmission at synapses.	Smoking as a quantitative trait locus, substance dependance,
COPS6	It is involved in the regulation of cell cycle	Xeroderma Pigmentosum, Complementation Group E
CRHR2	Peptide hormone metabolism and ADORA2B mediated anti-inflammatory cytokines production.	Eating Disorder , irritable bowel syndrome, anxiety
CSNK1G1	The encoded kinase plays a role in cell cycle checkpoint arrest in response to stalled replication forks	Aortic Valve Prolapse, Gangliosidosis
CUTA	Enables enzyme binding activity. Involved in protein localization	Optic Atrophy 3, Autosomal Dominant , Intestinal Tuberculosis
CUX2	Transcription factor involved in the control of neuronal proliferation and differentiation in the brain.	Developmental And Epileptic Encephalopathy, thrombocytopenia
CYP2A7	Catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.	17-hydroxysteroid dehydrogenase deficiency, 18-hydroxylase deficiency, various enzyme deficiencies

DDX39B	Gene expression (Transcription) and Transport of Mature Transcript to Cytoplasm.	Malaria and autoimmune diseases , RA
DIRAS1	Involved in GTP binding and mitogen-activated protein kinase binding.	Sciatic Neuropathy, photosensitive epilepsy
DNAH11	Reported to be involved in the movement of respiratory cilia	Ciliary dyskinesia, situs inversus
DOCK2	Remodeling of the actin cytoskeleton required for lymphocyte migration	Immunodeficiency and lymphopenia
EGFL8	Predicted to enable signaling receptor binding activity and to be involved in anatomical structure development.	Ludwig'S Angina , Bardet-Biedl Syndrome 11
EIF3E	Regulation of gene expression; and translational initiation.	/
EMC1	Part of ER membrane protein complex (EMC)	Cerebellar atrophy, visual impairment, and psychomotor retardation
ERCC6L2	May play a role in DNA repair and mitochondrial function.	Bone Marrow Failure Syndrome, AML, pancytopenia
FAM174A	Predicted to be integral component of membrane.	Prognostic marker in glioma (unfavorable) and renal cancer (favorable)
FAM65B	Involved in myotube formation and in regulation of cell adhesion, polarization, and migration.	Hearing loss
FANCA	Assembly into a common nuclear protein complex- Fanconi anemia complementation group A (FANC)	Fanconi Anemia, Pituitary Stalk Interruption Syndrome
FBXL17	The differentiation and survival of neural crest and neuronal cells, mediates ubiquitination	Proprotein convertase 1/3 deficiency.
FGD3	Filopodium assembly; regulation of gtpase activity; and regulation of cell shape.	Aarskog-Scott Syndrome , Charcot-Marie-Tooth Disease,
FNBP1	May act as a link between RND2 signaling and regulation of the actin cytoskeleton (By similarity).	Familial Hypocalciuric Hypercalcemia, Persistent Hyperplastic Primary Vitreous
FZD6	Inhibiting the processes that trigger oncogenic transformation, cell proliferation, and inhibition of apoptosis.	Neural tube defects
GGT7	Encodes enzymes involved in both the metabolism of glutathione and in the transpeptidation of amino acids.	Glutathionuria
GNGT1	Modulator or transducer in various transmembrane signaling systems.	Citrullinemia, night blindness, retinitis pigmentosa
HLA-A	RNA binding and peptide antigen binding.	Severe Cutaneous Adverse Reaction, Birdshot Chorioretinopathy, Chronic

		Pyelonephritis, Severe Acute Respiratory Syndrome
HLA-DQA2	Plays a central role in the peptide loading of MHC class II molecules	Penis Carcinoma In Situ, Podoconiosis , narcolepsy
HLA-DQB1-AS1	Promotes cell proliferation and inhibits apoptosis in hepatocellular carcinoma	Hepatocellular carcinoma
HSPA1B	Encodes a 70kda heat shock protein, stabilizes existing proteins against aggregation	Mulibrey Nanism
HSPA1L	Encodes a 70kda heat shock protein, stabilizes existing proteins against aggregation	Paranoid Schizophrenia , phototoxic dermatitis
HSPB2	Plays an important role in maintenance of muscle structure and function.	Malignant fibrous histiocytoma, charcot-marie-tooth disease, neuromuscular disease
HSPB2-C11orf52	Candidate for nonsense-mediated mrna decay (NMD), unlikely to produce a protein product	/
IZUMO4	Members of the IZUMO family, such as IZUMO4, are expressed in sperm and have potential roles in sperm-egg fusion	/
KIAA1468	Involved in intracellular cholesterol transport.	Chronic Recurrent Multifocal Osteomyelitis, Familial Expansile Osteolysis
KRR1	Predicted to be involved in rrna processing.	Duodenum Adenoma, Colorectal Cancer, Hereditary Nonpolyposis
LTA	Signaling receptor binding and tumor necrosis factor receptor binding.	Psoriatic arthritis
LUZP2	Leucine zipper protein.	Low grade glioma, albinism
MADCAM1	Endothelial cell adhesion molecule to direct leukocytes into mucosal and inflamed tissues.	Ileitis, sclerosing cholangitis
MDC1	Homology Directed Repair and DNA IR-double strand breaks and cellular response via ATM.	Nijmegen Breakage Syndrome and Cornelia De Lange Syndrome.
MLKL	Plays a critical role in tumor necrosis factor (TNF)-induced necroptosis,	Maturity-Onset Diabetes Of The Young, Inflammatory bowel disease, Caspase 8 Deficiency
MRM1	Predicted to be involved in rrna 2'-O-methylation	Non-Syndromic X-Linked Intellectual Disability, 3-Methylcrotonyl-Coa Carboxylase 1 Deficiency
MTCH1	Induces apoptosis independent of the proapoptotic proteins Bax and Bak	Developmental And Epileptic Encephalopathy
MTHFSD	Enables RNA binding activity.	Microphthalmia, Syndromic, Bardet-Biedl Syndrome

NAV2	May play a role in cellular growth and migration	Neuroblastoma
NEIL1	Participates in the DNA repair pathway by initiating base excision repair by removing damaged bases, primarily oxidized pyrimidines.	Congenital Disorder Of Deglycosylation , Cockayne Syndrome, Werner syndrome
NLRC5	Plays a role in cytokine response and antiviral immunity through its inhibition of NF-kappa-B activation and negative regulation of type I interferon signaling pathways.	Familial Pityriasis Rubra Pilaris, Bare Lymphocyte Syndrome
NOD1	Plays a role in innate immunity by initiating inflammation. This protein has also been implicated in the immune response to viral and parasitic infection.	Inflammatory bowel disease, blau syndrome , gastritis
NRSN1	Predicted to be involved in nervous system development.	Reading Disorder ,Dyscalculia,Hirschsprung Disease
NUTM2G	? Uncharacterised protein. Expressed significantly in testis	/
OAS1	Plays a key role in innate cellular antiviral response, and has been implicated in other cellular processes like cell growth and apoptosis.	immunodeficiency 100 with pulmonary alveolar proteinosis and hypogammaglobulinemia
OR2H1	Olfactory Signaling Pathway and Class A/1 (Rhodopsin-like receptors).	/
PCF11	Necessary for efficient Pol II transcription termination and may be involved in degradation of the 3' product of polyA site cleavage.	Amyotrophic Lateral Sclerosis 4, Juvenile , Oculopharyngeal Muscular Dystrophy
PDCD6IP	Endocytosis, multivesicular body biogenesis, membrane repair, cytokinesis, apoptosis and maintenance of tight junction integrity.	Adult neuronal ceroid lipofuscinosis 1, microcephaly
PKHD1L1	Predicted to act upstream of or within sensory perception of sound.	Polycystic Kidney Disease
PLA2G2D	Involved in inflammation and immune response	Surfactant dysfunction and gas gangrene
PLEKHJ1	Predicted to be involved in endosome organization; receptor recycling; and retrograde transport, endosome to Golgi.	/
PNPLA1	Has a role in glycerophospholipid metabolism in the cutaneous barrier.	Ichthyosis
POLR2E	Encodes the fifth largest subunit of RNA polymerase II, the	Esophageal Cancer

	polymerase responsible for synthesizing messenger RNA in eukaryotes.	
PPT2-EGFL8	A candidate for nonsense-mediated mrna decay -unlikely to produce a protein product	/
PRR3	Enables RNA binding activity.	/
PRRC2A	RNA binding.	Phonagnosia, Trichohepatoenteric Syndrome 2, RA, DM
PTBP1	Plays a role in pre-mrna splicing and in the regulation of alternative splicing events.	Human T-Cell Leukemia Virus, Congenital Myasthenic Syndrome
RARS	Catalyze the aminoacylation of trna	Leukodystrophy, Hypomyelinating , Embryonal Carcinoma
RFWD3	Involved in DNA metabolic process; regulation of cell cycle phase transition; and response to ionizing radiation.	Fanconi anaemia
RP11-727A23.5	Pseudogene	/
RPP21	Rrna processing in the nucleus and cytosol and trna processing	Anauxetic Dysplasia , Cartilage-Hair Hypoplasia
SENP5	Metabolism of proteins and Transport of the SLBP independent Mature mrna	3-Methylglutaconic Aciduria, Type Iii 64
SERTAD4	Predicted to be located in nucleus, function unknown	/
SMAD3	Forms a complex with other SMAD proteins and binds DNA, functioning both as a transcription factor and tumor suppressor.	Loeys-Dietz Syndrome, Aortic aneurysm, Connective Tissue Disease
SOX6	Transcriptional activator that is required for normal development of the central nervous system, chondrogenesis and maintenance of cardiac and skeletal muscle cells	Tolchin-le caignec syndrome, osteochondroma, craniosynostosis
STK19	Possible that phosphorylation of this protein is involved in transcriptional regulation.	Skin squamous cell carcinoma, melanoma
STON2	Involved in regulating endocytotic complexes.been shown to participate in synaptic vesicle recycling	/
TECPR1	Involved in targeting protein aggregates, damaged mitochondria, and bacterial pathogens for autophagy	/
TEK	Mediates a signaling pathway that functions in embryonic	Venous malformations, multiple cutaneous and mucosal, genuine diffuse

	vascular development.	phlebectasia, glaucoma
TENM2	Enables cell adhesion molecule binding activity and signaling receptor binding activity.	Usher Syndrome, Type Iic
TIMM13	Mitochondrial intermembrane chaperone that participates in the import and insertion of some multi-pass transmembrane proteins into the mitochondrial inner membrane.	Mohr-Tranebjaerg Syndrome, Visual Cortex Disease, Charcot-Marie-Tooth Disease, X-Linked Dominant,
TMEM116	Predicted to be integral component of membrane.	/
TMEM184A	Predicted to enable heparin binding activity, Acts as a heparin receptor in vascular cells	/
TMPRSS9	This gene enhances the invasive capability of pancreatic cancer cells and may be involved in cancer progression	Pancreatic cancer
TNK2	Directly linked to a tyrosine phosphorylation signal transduction pathway.	Infantile-onset mesial temporal lobe epilepsy with severe cognitive regression 99, parkinsoms, aml
TNXB	Function in matrix maturation during wound healing	ehlers-danlos syndrome, classic-like,adrenal hyperplasia, congenital, due to 21-hydroxylase deficiency, vesicoureteral reflux 8,
TRIB2	Apoptosis of cells mainly of the hematopoietic origin	Causes AML
TRIM16	Ncoded protein can homodimerize or heterodimerize with other TRIM proteins and has E3 ubiquitin ligase activity. This gene is also a tumor suppressor and is involved in secretory autophagy.	Laurence-Moon Syndrome, Lice infestation
UBAP2	Function in the ubiquitination pathway	/
USP20	Play a role in angiogenesis, glucose metabolism, cell proliferation and metastasis.	/
VWA9	Predicted to be involved in snrna 3'-end processing.	/
WIPF3	Predicted to be involved in actin filament-based movement, May be a regulator of cytoskeletal organization	Azoospermia
WNK2	Regulation of electrolyte homeostasis, cell signaling survival, and proliferation	Agenesis of the corpus callosum with peripheral neuropathy, hypomagnesemia, meningioma
XKR3	Homologs of the Kell blood group precursor XK	/
ZNF276	Predicted to be involved in regulation of transcription by RNA polymerase II	Fanconi Anemia, Complementation Group A

ZNF536	Highly conserved zinc finger protein. The encoded protein is most abundant in brain, where it negatively regulates neuronal differentiation	/
ZNF554	May be involved in transcriptional regulation.	/
ZNF717	Transcriptional regulator-cell proliferation, differentiation and apoptosis	/

Received: 30 March 2021 | Revised: 8 June 2021 | Accepted: 2 July 2021
DOI: 10.1111/cge.14026

CLINICAL GENETICS WILEY

Appendix M-Publication by author

Hypermobile Ehlers–Danlos syndrome: A review and a critical appraisal of published genetic research to date

Kirsty Scicluna[1] | Melissa M. Formosa[1,2] | Rosienne Farrugia[1,2] | Isabella Borg[2,3,4]

[1]Department of Applied Biomedical Science, Faculty of Health Sciences, University of Malta, Msida, Malta

[2]Centre for Molecular Medicine and Biobanking, University of Malta, Msida, Malta

[3]Medical Genetics Unit, Department of Pathology, Mater Dei Hospital, Msida, Malta

[4]Department of Pathology, Faculty of Medicine and Surgery, University of Malta, Msida, Malta

Correspondence
Isabella Borg, Department of Applied Biomedical Science, Faculty of Health Sciences, University of Malta, Msida, Malta. Email: isabella.borg@um.edu.mt

Abstract

The Ehlers–Danlos syndromes (EDS) are a collection of rare hereditary connective tissue disorders with heterogeneous phenotypes, usually diagnosed following clinical examination and confirmatory genetic testing. Diagnosis of the commonest subtype, hypermobile Ehlers–Danlos Syndrome (hEDS), relies solely on a clinical diagnosis since its molecular aetiology remains unknown. We performed an up-to-date litera-ture search and selected 11 out of 304 publications according to a set of established criteria. Studies reporting variants affecting collagen proteins were found to be hin-dered by cohort misclassification and subsequent lack of reproducibility of these genetic findings. The role of the described variants affecting Tenascin-X and LZTS1 is yet to be demonstrated in the majority of hEDS cases, while the functional implica-tion of associated signaling pathways and genes requires further elucidation. The available literature on the genetics of hEDS is scant, dispersed and conflicting due to out-dated nosology terminology. Recent literature has suggested the role of several promising candidate mechanisms which may be linked to the underlying molecular aetiology. Knowledge of the molecular genetic basis of hEDS is expected to increase in the near future through the mainstream use of high-throughput sequencing com-bined with the updated classification of EDS, and the upcoming Hypermobile Ehlers– Danlos Genetic Evaluation (HEDGE) study.

KEYWORDS
connective tissue diseases/genetics, Ehlers–Danlos syndrome, Ehlers–Danlos syndrome type 3, hypermobile EDS

1 | INTRODUCTION

The Ehlers–Danlos syndromes (EDS) are a group of 13 rare hereditary connective tissue disorders distinguished by substantial phenotypic and genetic heterogeneity which often obscures their diagnosis.[1] Most EDS subtypes are associated with genes primarily affecting the synthesis, metabolism and deposition of collagen with a minority affecting other extracellular matrix (ECM) proteins.[1,2] The most promi-nent features observed in EDS include a triad of joint hypermobility, skin hyper-extensibility and generalized fragility of soft tissue. How-ever, the extent and severity of symptoms vary significantly between EDS subtypes.[3] Due to the increasing use of high-throughput sequencing (HTS) and discovery of variants in genes which are corre-lated with distinct phenotypes, the 2017 International classification of the EDS was established.[1] This classification increased the number of EDS cases having a definite molecular diagnosis,[4] with the exclusion of the elusive hypermobile Ehlers–Danlos syndrome (hEDS) subtype.

Hypermobile EDS (hEDS) is considered the commonest hereditary connective tissue disorder.[5] Despite numerous research efforts, no gene or gene variant has so far been associated with its development, thus remaining the least understood EDS subtype in terms of its

underlying molecular aetiology.[6,7] This is further illustrated by the fact that although it appears to be inherited in an autosomal dominant manner,[8,9] clinical practice as well as research studies have demon-strated that 90% of cases are female.[10–12] From a clinical perspective, the hEDS phenotype is seemingly milder than the classical and vascu-lar EDS subtypes, mainly due to the absence of severe cardiovascular events.[9] Nonetheless, hEDS is associated with a high disability poten-tial[13] and significant psychosocial impact.[14] Patients report an early emergence of generalized joint hypermobility, consequently leading to joint instability and recurrent dislocations or subluxations upon mini-mal trauma. This leads to complaints of chronic pain, chronic fatigue and autonomic dysfunction, which significantly affect the patients' quality of life.[15–17] Dermal involvement in hEDS is attenuated com-pared to other subtypes,[9,18] but a pronounced involvement of gastro-intestinal disturbance is observed.[15] Furthermore, psychological distress such as anxiety and depression are found in high prevalence in hEDS patients.[19] There is no standardized treatment for hEDS; clin-ical management is mainly preventative and tailored to the individual's clinical manifestations.[20]

The 2017 international EDS classification sought to delineate the hEDS criteria in order to improve phenotyping and clinical diagnoses, as well as to refine and homogenize cohorts from a clinical and research perspective[1] with the aim of facilitating the discovery of a genetic mechanism.[9] Due to the absence of known associated genes and a confirmatory clinical molecular test, diagnosis of hEDS relies exclusively on clinical examination and family history.[1] Inter- and intra-familial phenotypic variability, unawareness of the disorders amongst physicians, as well as clinical overlap with other connective tissue disorders further confound its diagnosis.[21,22] Therefore, research focused on identifying genes and genetic variants associated with hEDS would be instrumental in establishing its genetic aetiology and in the development of a diagnostic test.[23]

hEDS is closely associated with phenotypically similar disorders such as EDS type III and HT-EDS, defined using the former Berlin and Villefranche classification systems respectively.[24,25] Since the new established criteria for hEDS are limited only to individuals showing widespread manifestations or clear Mendelian inheritance,[26]this results in the exclusion of affected patients some of which were previ-ously diagnosed as EDS type III or HT-EDS.[9,27] The HT-EDS pheno-type significantly overlaps and is now thought to be clinically indistinguishable from joint hypermobility syndrome (JHS) and benign joint hypermobility syndrome (BJHS). These two disorders were previ-ously defined with their respective criteria[26] until expert opinion con-cluded that HT-EDS and these disorders can be deemed as one[28] and were proven to co-segregate within the same family.[21,28] In order to resolve this conundrum, a continuous spectrum of hypermobility was proposed. Patients who present with some mandatory features included in the hEDS criteria and those diagnosed with these pheno-typically similar disorders can be classified as having "Generalised Hypermobility Spectrum Disorder" (G-HSD),[29] one of the "Hyper-mobility Spectrum Disorders" (HSD) which reside in the middle of the spectrum. Individuals within the same pedigree may show different hypermobility phenotypes described in the spectrum: hEDS patients may have family members with isolated JH and HSD as well as hEDS.[29] For the scope of this review, while in full understanding that EDS type III, HT-EDS, JHS, and BJHS may not display complete phenotypic and genotypic overlap with hEDS, we will take into consideration studies that also investigate these cohorts as well.

Literature pertaining to the topic is often complicated by the dif-ferent nosology terms adapted over time and by the existence of these phenotypically similar disorders, consequently dispersing research according to the search term used. Most articles on hEDS focus on the clinical, nosology and rehabilitative aspects. While these are essential for a stricter delineation of the condition, it is also crucial to map out and explore the lacunae present in genetic research con-ducted thus far. Identification of study limitations, methodologies used and potential areas of interest may accelerate and substantiate any studies on the subject conducted in the near future. As opposed to the highly structured, evidence-based nature of systematic reviews, conducting a scoping review was determined to be more appropriate for reviewing hEDS in order to assemble the heterogeneous and scant evidence available for this condition.[30,31]

2 | METHODS

2.1 | Databases and search strategy

In order to search the available literature and identify articles perti-nent to this review, PubMed was primarily used to access the MEDLINE database. The search terms used in each string are delin-eated in Table 1. Previous terminology from now obsolete criteria were also used in the search strings as using terminology from the 2017 classification only would limit the retrieval of relevant literature. Other combinations using these terms were also implemented in order to exhaust all possibilities. One search query resulted in a high number of retrieved articles that were mostly related to the classical, vascular and kyphoscoliotic EDS subtypes. This search string was then modified by excluding these subtypes and respective acronyms, as given in Table 1. All queries were searched from 1966, the earliest published literature featured in MEDLINE, to 21 January 2021.

In addition to using a traditional database, the search using the term "hypermobile Ehlers–Danlos Syndrome" was also extended to other sources of information such as Online Mendelian Inheritance in Man (OMIM),[32] National Organisation for Rare Disorders (NORD),[33] OrphaNet,[34] Genetic Testing Registry (GTR),[35] The Genetic and Rare Diseases Information Centre (GARD),[36] and SNPedia.[37] Recent comprehensive reviews on hEDS[9,23,38] were also included.

2.2 | Inclusion and exclusion criteria

Database-specific filters were applied in order to display articles writ-ten in the English language and only concerning human subjects. No time restrictions were applied. Articles were included if (a) the condition(s) investigated included hEDS or its associated synonyms,

TABLE 1 Search strategy for PubMed

Search term	No. of results	Excluded articles	Selected articles	Studies
("Ehlers–Danlos Syndrome/genetics"[Mesh] AND "Ehlers–Danlos syndrome type 3" [supplementary concept])	11	9	2	Chiarelli et al., 2016 Yamada et al., 2019
("Ehlers–Danlos syndrome/genetics"[MeSh]) AND ("hypermobile Ehlers–Danlos syndrome" OR "hEDS "OR "joint hypermobility syndrome" OR "JHS "OR "benign joint hypermobility syndrome" OR "EDS type III" OR "Ehlers–Danlos syndrome hypermobility type" OR "EDS-HT" OR "hypermobility Ehlers–Danlos"[Title/Abstract]) AND ((humans[Filter]) AND (english[Filter])) NOT (classical OR vascular OR cEDS OR vEDS OR VI OR kyphoscoliotic)	161	152	9	Chiarelli et al, 2016 Hoffman et al., 2008 Narcisi et al, 1994 Syx et al., 2015 Yamada et al, 2019 Zoppi et al., 2018 Zweers et al., 2003 Zweers et al., 2005a Zweers et al., 2005b
"Mutation/genetics"[MeSH] AND ("hypermobile Ehlers–Danlos syndrome" OR "hEDS "OR "joint hypermobility syndrome" OR "JHS "OR "benign joint hypermobility syndrome" OR "EDS type III" OR "Ehlers–Danlos syndrome hypermobility type" OR "EDS-HT" OR "hypermobility Ehlers–Danlos"[Title/Abstract])	56	54	2	Henney et al, 1992 Hoffman et al, 2008
("hypermobility-type" OR "hypermobile Ehlers–Danlos syndrome" OR "joint hypermobility syndrome"[Title/Abstract]) AND "Ehlers–Danlos Syndrome/genetics"[MAJR]	23	19	4	Chiarelli et al,2016, Syx et al, 2015 Zweers et al, 2003 Zweers et al, 2005
("Ehlers–Danlos syndrome/genetics"[MAJR]) AND "Mutation/genetics"[MeSH]	53	51	2	Hoffman et al, 2008 Weerakkody et al., 2016

Note: Articles highlighted in bold are duplicates found across several search strings and were included in the duplicate count.

and (b) if the condition was investigated using transcriptomic, geno-mic, or molecular genetics methods. Exclusion criteria included (a) sole investigation of other heritable connective tissue disorder or the other 12 EDS subtypes, (b) investigation of overlapping syndromes, and

l absence of molecular genetic testing. These criteria were applied to optimize the search strategy and thus narrow down the number of articles included in this review.

2.3 | Selection process

After applying the MESH terms and filters into the database, duplicate articles found in multiple search strings were discarded. Article titles and abstracts were briefly screened in order to apply the inclusion and exclusion criteria. When the abstract was unavailable, the full-text was accessed to determine the relevance of the article to the subject. Ultimately, the articles satisfying the criteria outlined above were selected and retained for qualitative analysis.

2.4 | Data extraction

From the selected articles, the following categories of data were extracted: (a) names of authors and year of publication, (b) principal aims of the study, (c) size and gender of the cohort, (d) attributed diagnosis synonym, (e) diagnostic criteria used, (f) sample type, (g) method of genetic analysis, and (h) outcome of study. This data were extracted and compiled for this review.

2.5 | Analysis

A qualitative examination of each of the selected articles was performed so as to explore the different criteria and methodologies utilized for the studies and thus determine the relevance of outcomes from these studies for future genetic research on hEDS.

≥ | RESULTS

3.1 | Findings

Figure 1 outlines the results of the literature search and the quantita-tive selection process using the online PubMed search engine. Online resources commonly used for genetic research were then utilized. Four of these online resources, OMIM (OMIM 130020), NORD, Genetic Testing Registry and SNPedia referred to studies already included in the analysis.[2,39–41] Additional searches on GARD and

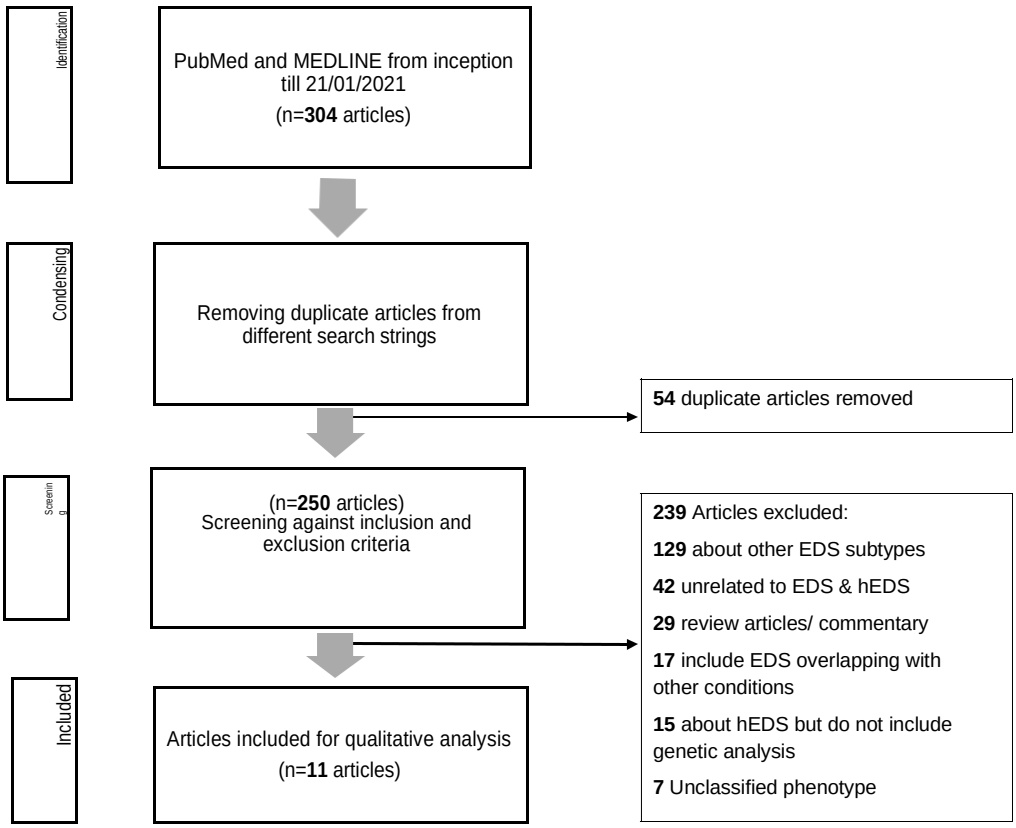


FIGURE 1 Quantitative overview of the selection process

OrphaNet (ORPHA:285) were inconclusive as these pages are either still under development or only state that the genetic cause of hEDS is unknown.Recent reviews on hEDS also mention studies already identified and included for analysis..[7,39,40,42,43] The reviews by Gensemer et al.[38] and Tinkle[9] also included a novel study[44] in which a phenotype closely related to hEDS was associated with an altered copy number of TPSAB1, resulting in an increase in serum tryptase levels. Despite the similarity of the described phenotype to hEDS, patients were not assessed using any diagnostic criteria for EDS. Furthermore, only 28% of the cohort had generalized joint hypermobility, a major phenotypic feature of hEDS. While the authors of this review acknowledge that further studies are indeed required to test this intriguing association, this study was excluded from our analysis for these reasons.

3.2 | Characteristics of selected studies

Data extracted from the 11 selected publications is given in Table 2. Articles retrieved and included in our review were published between 1992 and 2021. The majority of studies conducted between 2003 and 2016 diagnosed their cohort as EDS-HT using the Villefranche Criteria, with some inclusions of BJHS and JHS. The two studies conducted prior to this date used the Berlin Nosology (EDS III) and provisional criteria for JHS respectively. Studies conducted in 2018 and 2019 adopted the terminology used in the then-newly defined 2017 International Criteria.

Most studies included in this review used Beighton's hyper-mobility score in order to assess their cohorts, which in turn is com-monly featured in past and present diagnostic criteria for EDS. Clinical examination and clinical history of the subjects was performed by cli-nicians or specialists in all of these articles, in contrast to self-reporting of symptoms by the patients themselves. Of particular note is the thorough strategy used by Syx et al.[7] which included clinical examinations at baseline, as well as after 24 months by two separate clinical geneticists. A total of 8 studies out of 11 also indicated the source of recruitment, which were chiefly conducted in a hospital set-ting and through patient support networks. During extraction of data for this review, it was observed that some patients were recruited to more than one study. It was also noted that in the study by Weerakkody et al.,[45] patients with vascular phenotypes were favored over those with non-vascular manifestations only.

While the primary objective of most of these studies was to uncover genes and gene variants which potentially may be associated with hEDS,[2,7,39–41,46,47] one study did not focus solely on hEDS, but aimed to test the efficiency of two research-based HTS panels on a mixed EDS cohort.[45] Despite the fact that a novel TNXB variant was indeed discovered in another study, the principal aim was not the discovery of a genetic variant but to test the utility of measuring serum Tenascin-X levels.[48] Furthermore, studies by Chiarelli et al.[38] and Zoppi et al.[39] did not directly test for the presence of variants, but rather examined perturbations in gene expression and/or cellular pathways.

TABLE 2 Overview of data extracted from selected studies

Study	No. and gender of subjects	Diagnostic Synonyms	Diagnostic Criteria	Sample type	Method of genetic analysis	Outcome of genetic testing
1992-Henney et al.[44] Tested the association between JHS and variants in genes coding for Type III, V and VI collagen	Two families, Family A n = 10 (4 M,6 F), Family B n = 21 (11 M, 10F)	JHS	Provisional criteria for Hypermobility Syndrome[45]	DNA from peripheral blood	Restriction Fragment Length Polymorphism (RFLP)	No evidence that JHS is caused by a variant in the COL3A1, COL5A2 or COL6A3 genes, variants in COL1A1 and COL1A2 genes unlikely
1994-Narcisi et al.[29] Genetic examination of proband and his family members	1 family n = 5 (4 M, 1F)	EDS-III	Berlin Nosology[46]	cDNA from dermal fibroblasts DNA from peripheral blood	Sanger sequencing Allele specific oligonucleotide hybridization	Discovered variant in COL3A1 p.G637S
2003-Zweers et al.[33] Tested the association between TNX haplo-insufficiency and HT-EDS/ BJHS	n = 24 from 4 families: 4 unaffected, 20 affected (6 M, 14 F). Six additional subjects from an EDS-HT cohort	EDS-HT/BJHS	Villefranche Criteria[47] Revised Criteria for BJHS[48]	gDNA from peripheral blood and cultured fibroblasts	Sanger sequencing Southern blotting	17 subjects positive for variants in TNXB: a 2 bp deletion in exon 8 [c._AA56063], a 2 bp insertion in exon 3 [c.GT44906] and a 30 kb deletion. All heterozygous variants accompanied by decreased serum TNX concentration.
2005-Zweers et al.[3] Determined the association of missense TNXB variants with HT-EDS and the effect of these variants on dermal tissue.	n = 33: 24 HT-EDS subjects and 9 controls	EDS-HT	Villefranche Criteria[47]	gDNA from peripheral blood	Sanger sequencing	3 TNXB variants (2 of which are novel) discovered in 3 EDS-HT individuals with normal serum TNX levels: p. Arg29Trp, p.Val1195Met and p.Leu4033Ile
2005-Zweers et al.[34] Investigated whether patients with BJHS also have decreased serum TNX concentration	n = 54 (1 M, 53F)	BJHS	Revised Criteria for BJHS[48]	Undisclosed	Undisclosed	No previously-described TNXB variants found in this cohort
2008-Hoffman et al.[28] Tested whether COL5A3 variants are linked with EDS-HT	n = 13 (3 M, 10F)	EDS-HT	Villefranche Criteria[47]	cDNA and gDNA from dermal fibroblasts	RT-PCR Sanger sequencing	11 SNPs found in COL5A3 not deemed as pathogenic, excluded COL5A3 as a candidate gene
2015-Syx et al[8] Identified linked locus in a three-generation affected family	1 family n = 35, genetic analysis on 14 members (5 M, 9F)	EDS-HT	Villefranche Criteria[47] Revised Criteria for BJHS[48]	cDNA from dermal fibroblasts gDNA from fibroblasts and peripheral blood	Linkage analysis Genome-wide linkage analysis Sanger sequencing Whole exome sequencing Array CGH	p.His211Gln substitution found in LZTS1. 3 additional variants in LZTS1 found in other probands: p. His17Asp, p.Arg529Trp, p. Ser250*

(Continues)

TABLE 2 (Continued)

Study	No. and gender of subjects	Diagnostic Synonyms	Diagnostic Criteria	Sample type	Method of genetic analysis	Outcome of genetic testing
2016-Chiarelli et al.[30] Aimed to uncover mechanisms for the pathogenesis of JHS/EDS-HT	n = 11. 5 affected females, 6 female controls	EDS-HT/JHS	Villefranche Criteria[47] Brighton Criteria for JHS[48]	Total RNA from dermal fibroblasts	Transcriptome-wide expression profiling by microarray analysis Quantitative real-time PCR	46 up-regulated genes and 162 down-regulated genes in JHS/EDS-HT when compared to controls. Disruption in signaling transduction pathways PI3K-AKT, TGFβ, canonical Wnt and JAK–STAT signaling
2016-Weerakkody et al[35] Screened a mixed EDS cohort using 2 NGS panels featuring collagen and aortopathy genes	n = 177 EDS individuals, 43% (n = 76) EDS-HT/BJHS (18 M, 56F, 2 undisclosed)	EDS-HT/BJHS	Villefranche Criteria[47]	gDNA from peripheral blood gDNA from saliva	Whole exome sequencing Sanger sequencing	Pathogenic variant in COL5A1 (p.Gly1522Cys) in one EDS-HT subject Likely pathogenic variant in COL1A1 (p.Arg994Cys) in other EDS-HT subject 16 VUS in COL1A1, COL1A2, COL3A1, COL5A1, COL5A2, TGFBR2 found in 14 EDS-HT subjects.
2018-Zoppi et al.[32] Uncovered perturbations in cellular mechanisms possibly involved in hEDS and HSD	n = 17, including 4 hEDS (4 F) and 3 HSD (1 M, 2F). 2 controls	hEDS, HSD	2017 International classification of EDS[1]	Total RNA from dermal fibroblasts	Quantitative real-time PCR	Increased mRNA expression of matrix metallopeptidase 9, an ECM-degrading protease, and transcription factor SNAIL1 in hEDS and HSD.
2019-Yamada et al.[31] Determined the feasibility of diagnosing JHS/hEDS by measuring serum TNX	n = 17 affected females, 50 unaffected females as controls	JHS/hEDS	2017 International classification of EDS[1]	gDNA from peripheral blood serum/plasma from peripheral blood	Whole exome sequencing nano-LC/MS/MS and Western blot analysis for measuring serum TNX	Nonsense variant in TNXB (p. Arg1653*) in one subject with normal serum TNX levels. No association observed between serum TNX and TNXB variants

Abbreviations: BJHS, benign joint hypermobility syndrome; cDNA, complementary deoxyribonucleic acid; CGH, comparative genomic hybridization; DNA, deoxyribonucleic acid; ECM, extracellular matrix; EDS, Ehlers–Danlos syndrome; EDS-HT, Ehlers–Danlos syndrome hypermobility type; F, females; gDNA, genomic deoxyribonucleic acid; hEDS, hypermobile Ehlers–Danlos syndrome; HSD, hypermobility spectrum disorders; JHS, joint hypermobility syndrome; LC/MS/MS, liquid chromatography tandem mass spectrometry; M, males; mRNA, messenger ribonucleic acid; NGS, next generation sequencing; PCR, polymerase chain reaction; RNA, ribonucleic acid; RT-PCR, reverse transcription polymerase chain reaction; SNP, single nucleotide polymorphism; TNX, tenascin-X; VUS, variant of uncertain significance.

In the majority of studies, genetic data were obtained in the form of DNA and RNA extracted from peripheral blood or from dermal fibroblasts. Only one study did not disclose the nature of the speci-men and the genetic template used for analysis.[41] The study by Weerakkody et al.[45] also extracted DNA from saliva. The methods used for genetic analysis vary between studies, mainly due to differ-ences in the aims of the research project, their study designs and, most significantly, according to the available technology at the time of analysis. Notably, the use of Sanger sequencing is consistent in both past and modern studies, while techniques such as whole exome sequencing (WES) are only featured in literature from 2015 onwards.

3.3 | Findings and conclusions

Two studies from a total of 11 excluded the involvement of genes involved in the production of collagen,[46,47] whereas one study reported the absence of previously described TNXB variants in their cohort.[41] In the remaining eight studies, authors either reported the discovery of novel variants in multiple genes[2,7,39–41,46,47] or the altered expression of numerous gene transcripts and signaling pathways.[42,43] Nevertheless, in light of the significant differences in the methodologies used and their respective limitations, it is crucial to evaluate further the applicability of these findings to diagnostics and future research. These minutiae, as well as the deficits in the literature available to date and potential areas of interest in prospective studies, will be discussed below.

4 | DISCUSSION

The studies featured in this review will be discussed in tandem with other selected studies which feature a common gene, protein or sig-naling pathway in order to better identify and compare the limitations and outcomes of these studies. Apart from genetic testing, most stud-ies were also augmented by histochemical and biochemical testing, as well as proteomic analysis. However, in order to adhere to the scope of this review, only the molecular aspects will be discussed. Through-out this discussion the term hEDS will be used to encompass HT-EDS, BJHS, and JHS as these have been considered to refer to the same condition.[8] Nevertheless, it is important to point out that most of these studies were conducted prior to the current classification. Hence, as previously stated, some JHS/BJHS/HT-EDS patients rec-ruited in these cohorts may now be classified as having other EDS subtypes or hypermobility spectrum disorders as they might not fit the new criteria for hEDS if re-evaluated. This limitation was highlighted in previous publications.[49] Therefore, studies conducted prior to 2017 must be interpreted with a degree of caution.Early liter-ature on the molecular basis of hEDS was directly focused on the genes encoding collagen I, III, V, and VI, as defects in the respective coding genes were already known to cause other EDS subtypes. The earliest study included in this review performed segregation analysis of collagen gene markers in two families with hEDS.[46] Consequently, the involvement of the collagen genes COL3A1, COL5A2, or COL6A3 in these two families was excluded, thus ruling out collagen type III as the defective protein, but not collagens V and VI, as the genes coding for other chains of collagen V and VI were not assessed. The authors remark that the involvement of COL1A1 and COL1A2 is also unlikely due to the lack of strong linkage in either case. Interestingly, it was suggested in this study that hEDS may be the result of other factors including the presence of other variants, which, on their own, might not affect the phenotype to such an extent as to be clinically notice-able. This hypothesis is later reflected in the threshold model pro-posed by Castori et al.[13] In this model, Castori et al. suggested that a single inherited variant may serve as a predisposing factor or" suscep-tibility locus", which in turn results in the expression of the disease only when supplemented with other inherited or acquired factors.[13]

In contrast, Narcisi et al.[39] describe a glycine to serine substitu-tion in COL3A1 which was discovered in one family with features sim-ilar to hEDS. However, as also remarked by the author, variants in type III collagen are usually found in the vascular EDS subtype (vEDS) which in turn may show similar clinical features found in hEDS.[50] In addition, the diagnostic criteria used for clinical assessment of this cohort are highly non-specific by modern standards. Thus, the findings of this study are instead thought to reflect a family with vEDS with subclinical or yet undeveloped vascular complications which may include aneurysms, dissections or arterial ruptures rather than hEDS.[51] In both of these studies, the male: female ratio (approxi-mately 1:1) is unusually high compared to later studies in which the ratio skews significantly towards the female sex. This might be explained by the fact that females tend to present with more advanced and severe phenotypes[6] which may satisfy the newer, strin-gent hEDS criteria more than their male counterparts. Thus, older studies applying broader diagnostic criteria might have led to the inclusion of more "affected" males that would otherwise be excluded when following the current stricter criteria if re-examined.

In a more recent study[47] the gene coding for type V collagen (COL5A3) was deemed as unlikely to be associated with hEDS due to the absence of obvious pathogenic variants in 13 individuals. Eleven variants were discovered; however, only the two novel variants were disclosed namely, NP_056534.2: p.Asn226Tyr (rs62104337) and NP_056534.2: p.Ala280Pro (rs61742765). The same two variants were also found to be present in 9 out of 26 healthy individuals and were thus deemed as unlikely to be associated with hEDS. The degree of association between variants affecting collagen and hEDS was fur-ther investigated by applying research-based HTS panels, in which a pathogenic variant in COL5A1 (c.4564G > T, p.Gly1522Cys) and vari-ants of uncertain significance (VUS) in COL1A1, COL1A2, COL3A1, and COL5A2 were identified in hEDS individuals.[45] The discovery of variants in COL3A1 is seemingly common with the findings by Narcisi et al.[39] while the involvement of COL5A1 has never been determined in hEDS. COL3A1 variants result in the synthesis of unstable collagen precursors, which disrupt the successful assembly of type III collagen, in turn reducing the ability of organs to withstand mechanical stress.[52] However, it should be pointed out that this study had a selection bias for vascular phenotypes to the extent that a significant number of

patients deemed as hEDS also met the Villefranche criteria for classical and vEDS. Furthermore, the percentage of hEDS subjects presenting with arterial aneurysms or dissection (20%) is unusually high as these clinical features do not typically form part of the hEDS phenotype spectrum.[53] Hence, as a result of possible misclassification of the cohort, the findings reported in this study may be attributed to vEDS and/or other EDS subtypes.

Apart from the collagens, the ECM protein Tenascin-X (TNX) has also been associated with the pathogenesis of some EDS subtypes such as hEDS and the more defined classical-like EDS (clEDS). In the latter, homozygous or compound heterozygous autosomal recessive variants in the gene coding for Tenascin-X (TNXB) result in either bi-allelic gene deletion or defective mRNA transcripts which undergo nonsense mediated decay.[54] Consequently, this leads to a complete deficiency of the TNX protein.[54] Sequencing and molecular analysis of the TNXB gene is thus crucial in the diagnosis of clEDS,[1] which can be supplemented by quantification of the TNX protein in serum.[55] Indi-viduals with clEDS present with a hypermobile phenotype along with skin hyperextensibility, cardiac-valvular abnormalities and chronic pain, amongst others.[1,55] The chromosomal loci for TNXB and CYP21A2 overlap. Thus, deletions at the locus give rise to continuous CYP21A2 and TNXB chimeric genes.[54,56] CYP21A2 codes for 21-hydroxylase, an essential catalytic enzyme in the glucocorticoid and mineralocorticoid pathways.[57] Perturbations in the synthesis of this protein leads to 21-hydroxylase deficiency, resulting in insuffi-cient production of cortisol and/or aldosterone, also known as congenital adrenal hyperplasia (CAH).[57] The presence of this chimeric gene affect the level and function of both TNX and 21-hydroxylase, resulting in a contiguous gene deletion syndrome termed CAH-X, which present with symptoms seen in both CAH and hEDS.[56] In addi-tion to clinical evidence supporting the link between TNX and connec-tive tissue disorders, the general functional implications of Tenascin-X on the musculoskeletal system have also been confirmed in a mouse model.[58] Tnx knockout mice showed muscle weakness, reduced colla-gen fibril diameter and elevated ECM turnover in skeletal muscle.[58]

The effect of these autosomal recessive variants in TNXB on EDS phenotypes was first highlighted[54] and later substantiated by Zweers et al.[40] who demonstrated the presence of a hEDS-like phenotype in 20 family members with heterozygous variants in TNXB. In contrast to the previous study, these novel variants in TNXB were inherited in an autosomal dominant manner, resulting in haploinsufficiency of the gene product. Additionally, while members within the same families had the same TNXB variants, no two families in this study shared a common variant, demonstrating a high level of inter-familial locus heterogeneity.

In a later study, Zweers et al.[2] demonstrated the presence of three additional variants in TNXB which did not directly lead to haploinsufficiency but instead altered the morphology of elastic fibers. These variants were not replicated in another unrelated cohort rec-ruited by the authors.[41] However, the method used for genetic analy-sis was not well-defined in the latter cohort, with serum samples being collected instead of peripheral blood as required for DNA test-ing. The lack of correlation between variants in TNXB and serum TNX levels was also shown by Yamada et al.,[48] in which only one research subject had a heterozygous nonsense variant [NM_019105.7: c.4957C > T (p.Arg1653*; rs201184519) but had normal serum TNX levels. In contrast to previous studies on TNX,[2] this study did not assess the effect of this seemingly non-functional variant on elastic fiber morphology. A significant reduction in serum TNX levels was also found in subjects without variants in TNXB, the cause of which was attributed to epigenetic changes. Since exome sequencing was performed rather than whole genome analysis, this observation can also be explained by the presence of undetected variants in deep intronic regions or intergenic sequence which could alter splicing and/or transcription regulation. This study, despite having comparable results, is distinguished from those conducted by Zweers et al.[2,40,41] as subjects were assessed using the 2017 diagnostic criteria and analysed using HTS.

In a comprehensive study, a large multigenerational Belgian family with hEDS was investigated.[7] First, any association between the phe-notype and genes encoding type I, III, and V collagen as well as TNXB was excluded in 14 out of 35 family members. A subsequent genome-wide linkage scan identified a linked locus which was narrowed down to a region spanning 8p22-8p21.1. WES was performed in two family members and only variants residing in the identified locus were con-sidered. Further filtering and in silico prediction tools narrowed the search to one missense variant: a histidine to glutamine substitution NP_066300.1:p.His211Gln (rs755849215) in the leucine zipper, puta-tive tumor suppressor-1 (LZTS1) gene. Using Sanger sequencing, this variant was also observed to segregate with the disease phenotype. The presence of variants in the LZTS1 gene was then tested in addi-tional probands, which identified three other likely pathogenic vari-ants. Some of the limitations of this study pertaining to the use of WES were duly acknowledged by the authors, namely that pathogenic variants in deep intronic sequences and regulatory regions could have been present but were not detected. Despite the fact that this study preceded the 2017 diagnostic criteria and thus utilized the broader diagnostic criteria available at the time, it is by far unique in terms of the type of cohort recruited, the level of clinical assessment of patients, as well as the detail of the methodology used. Due to these reasons, this article was particularly noted to be a fitting model for future family studies aiming to investigate the genetics of hEDS or other hereditary connective tissue disorders.

A recent transcriptome-wide expression analysis conducted on patient fibroblasts identified 46 up-regulated genes and 162 down-regulated genes in hEDS subjects when compared to healthy con-trols.[42] Genes with increased expression were classified by gene ontology (GO) categories and included genes involved in the immune response and redox process. Conversely, reduced expression was observed in transcripts involved in transcription regulation, kinase activity and inflammation. Transcripts involved in organization of cell– cell adhesion, and differentiation of osteoclasts were also under-expressed. Additionally, this study also noted the decreased expres-sion of proteins involved in calcium, TGF-β and JAK–STAT signaling pathways as well as the PI3K-AKT pathway. Analysis was performed on fibroblasts obtained from five patients only, and thus these

findings have to be corroborated using larger cohorts. In contrast to studies discussed above, the authors did not uncover any causative variant, however by identifying and understanding these transcriptional changes, one can potentially trace back the genetic changes resulting in these alterations. In their annotations, the authors demonstrate the numerous functional implications of these transcriptional changes and in turn, how these may result in the clinical manifestations observed in hEDS.[42]

The utility of transcriptional analysis in hEDS was further demon-strated in a study which reported a higher expression of matrix metallopeptidase 9 (MMP9, a bone and muscle ECM-degrading enzyme) and SNAI1 (a transcriptional repressor) in hEDS fibroblasts.[43] The decreased concentration of fibronectin described previously[42] was hypothesized to be a result of enhanced MMP9 expression which in turn, generates fibronectin fragments and activates an inflammatory response. After deducing that hEDS fibroblasts undergo a conforma-tional change into myofibroblasts which are strongly associated with the presence and proliferation of inflammation,[59] the author con-cludes that hEDS and HSDs are parts of the same entity which results in a chronic inflammatory state.[43] This hypothesis is supported by the increased levels of complement proteins observed in hEDS subjects[60] as well as the enhanced expression of immune mediators and reduced inflammatory regulators.[61] Apart from transcriptomics, this study is also supplemented by cell migration assays, immunofluores-cence and western blotting which, in synthesis, provide a holistic analysis of the effects of multiple pathways on the cellular and extracellular environment in hEDS.

In light of the considerations discussed above, it can be hypothe-sized that it is unlikely that hEDS is solely caused by variants in genes coding for the collagen proteins. Previous studies which reported the presence of such variants are hindered by possible misclassification of cohorts due to broad criteria and failure to reproduce similar results in subsequent cohorts. On the other hand, studies investigating the role of TNXB do indicate that a minority of hEDS patients may have het-erozygous variants in TNXB which can potentially cause pathogenic structural alterations in connective tissue without affecting serum TNX levels. However, these variants were only identified in a small subset of research patients and thus the genetic aetiology of the remaining majority is still unaccounted for. Meanwhile, the role of LZTS1 in hEDS is not yet confirmed as it is normally associated with prevention of oncogenesis.[62] However, it may be implicated in neuronal growth and migration[63] which is more functionally applicable to the pathophysiology of the condition. The genetic heterogeneity in hEDS was also highlighted by the lack of linkage of the identified locus in four other hEDS families, thus demonstrating the need for further research to test the role of LZTS1 variants and their prevalence in hEDS subjects. As previously discussed, the recent transcriptome studies[42,43] did not directly uncover causative or linked variants but instead demonstrated alterations in the transcriptional expression of several genes as well as the involvement of the immune response. Considering that the complement pathway has already been impli-cated in the periodontal EDS subtype,[1] the effect of a defective immune response on connective tissue is an interesting area of study for future research. Needless to say, further evidence is required for this correlation to have any merit. However, this hypothesis may func-tionally validate any discovered variants linked to the immune system.

i. | LIMITATIONS AND OTHER OBSERVATIONS

The use of a single database is a significant limitation of this review, as any articles which are not featured on PubMed might have been missed. Other well-resourced databases were excluded as some were either accessible by subscription basis only (EMBASE, CINAHL) or focus on the efficacy of interventions or clinical trials (COCHRANE). In order to counteract this deficit, the literature search was sup-plemented by other sources such as comprehensive contemporary reviews from experts in the field and online research tools in order to converge all information available on the subject. Furthermore, most of the articles selected for further analysis featured a small number of affected and unaffected individuals, which hinders the statistical power and applicability of these findings.

At the same time, this review is opportune as it presents the genetic literature on hEDS available to date, prior to the advent of the much anticipated Hypermobile Ehlers–Danlos Genetic Evaluation (HEDGE) study which aims to sequence 1000 hEDS subjects from around the globe.[64] The obvious difficulty of finding and recruiting well-sized representative cohorts is well-known in the study of rare disorders, a limitation also observed in all of the studies reviewed in this paper. The HEDGE study will be the largest and most diverse hEDS genetic cohort to date and one anticipates a better understand-ing on the genetic aetiology of hEDS once this study is concluded.

It is interesting to note that in contrast to the published literature discussed above, which classified patients according to the contempo-rary clinical criteria at the time, the often-resourced online genetic tools used in our review are still referring to the disorder using out-dated ontology terminology. A search on OMIM using the contempo-rary term “hEDS” initially yielded results for the other EDS subtypes, as the page for the condition (OMIM 130020) was archived under the previous terminology “Ehlers–Danlos Syndrome, hypermobility type”. Furthermore, the Genetic Testing Registry website refers to hEDS as EDS type III. While the need for classifying patients using contempo-rary criteria is more evident in the more-formal published literature, articles in the last decade are often supplemented by data from more-accessible online sources. These sources also need to be updated accordingly.

12. | RECOMMENDATIONS FOR FUTURE WORK

In light of the scant literature published to date, a search was also conducted for ongoing projects related to the genetics of hEDS on clinicaltrials.gov and the Research Portfolio Online Reporting Tool (RePORT). Regardless of the transparent and often-remarked need for

further studies on the subject, it was noted that there is no indication of further molecular studies other than the aforementioned HEDGE study and a proposed study at the University of Bristol (Proposal B3207). As is evident in this work and mostly due to the reasons out-lined above, studying hEDS using conventional approaches has not yielded sufficient comprehension about its aetiology. One must there-fore consider the introduction of novel technologies in conjunction with established methods featured in some studies discussed previ-ously. The inclusion of large, multigenerational families with clear modes of inheritance as in the case of Syx et al.[7] as opposed to stud-ies on individual unrelated subjects would shed more light on the pen-etrance of any discovered variant within the same family. As this is not always feasible, studies need to counteract this limitation through collaborative efforts to establish consortia which share larger patient cohorts. On this topic, studies may recruit individual subjects from multiple medical centers and record their phenotype using standard-ized terminology such as the Human Phenotype Ontology system. This information can then be integrated with genomic and trans-criptomic data and with the use of bioinformatics tools, detect genotype–phenotype patterns and patient sub-clusters with similar phenotypes using developed statistical algorithms. This approach has already been applied for neuromuscular and neurodevelopmental dis-orders[65,66] and may thus show some promise in the study of EDS and musculoskeletal diseases. Assessing the effect, if any, of epigenetic modifications on the development of the disorder may also be an intriguing approach for future work.

7 | CONCLUSIONS

To our knowledge, this is the first review to analyze in synthesis and condense the results of research studies available on the genetics of hEDS, which studies included cohorts diagnosed under all the termi-nologies used for hEDS throughout its history. It is clear that uncovering the underlying genetic aetiology of hEDS is a complex pro-cess which is further aggravated by the disorder's clinical heterogene-ity, as well as the variation in its classification. Nevertheless, recent studies have highlighted a number of valid candidate mechanisms which will serve as a point of reference in future studies, when assessing the functional and physiological effect of any discovered variant. As the utility of HTS is becoming more established in the fields of research and diagnostics, this technology in conjunction with the newly defined clinical criteria present ample opportunities for the hEDS researcher to expand the molecular understanding of this condition.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/cge.14026.

DATA AVAILABILITY STATEMENT

Data presented in this review is published and readily available.

ORCID

Kirsty Scicluna https://orcid.org/0000-0001-7064-0773
Melissa M. Formosa https://orcid.org/0000-0002-7852-2701
Rosienne Farrugia https://orcid.org/0000-0001-7038-2874
Isabella Borg https://orcid.org/0000-0003-0639-3055

REFERENCES

- Malfait F, Francomano C, Byers P, et al. The 2017 international classi-fication of the Ehlers-Danlos syndromes. Am J Med Genet. 2017;175 (1):8-26. https://doi.org/10.1002/ajmg.c.31552
- Zweers MC, Dean WB, van Kuppevelt TH, Bristow J, Schalkwijk J. Elastic fiber abnormalities in hypermobility type Ehlers-Danlos syn-drome patients with tenascin-X mutations. Clin Genet. 2005;67(4): 330-334. https://doi.org/10.1111/j.1399-0004.2005.00401.x
- Bascom R, Schubart JR, Mills S, et al. Heritable disorders of connec-tive tissue: description of a data repository and initial cohort charac-terization. Am J Med Genet A. 2019;179(4):552-560. https://doi.org/ 10.1002/ajmg.a.61054
- Sulli A, Talarico R, Scirè CA, et al. Ehlers-Danlos syndromes: state of the art on clinical practice guidelines. RMD Open. 2018;4(1):e000790. https://doi.org/10.1136/rmdopen-2018-000790
- Colombi M, Dordoni C, Chiarelli N, Ritelli M. Differential diagnosis and diagnostic flow chart of joint hypermobility syndrome/ehlers-danlos syndrome hypermobility type compared to other heritable connective tissue disorders. Am J Med Genet C Semin Med Genet. 2015;169(1):6-22. https://doi.org/10.1002/ajmg.c.31429
- Copetti M, Morlino S, Colombi M, Grammatico P, Fontana A, Castori M. Severity classes in adults with hypermobile Ehlers-Danlos syndrome/hypermobility spectrum disorders: a pilot study of 105 Ital-ian patients. Rheumatology (Oxford). 2019;58(10):1722-1730. https:// doi.org/10.1093/rheumatology/kez029
- Syx D, Symoens S, Steyaert W, De Paepe A, Coucke PJ, Malfait F. Ehlers-Danlos syndrome, hypermobility type, is linked to chromosome 8p22-8p21.1 in an extended Belgian family. Dis Markers. 2015;2015: 828970. https://doi.org/10.1155/2015/828970
- Levy HP. Hypermobile Ehlers-Danlos Syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews®. Seattle, WA: Uni-versity of Washington; 2018.
- Tinkle B, Castori M, Berglund B, et al. Hypermobile Ehlers-Danlos syndrome (a.k.a. Ehlers-Danlos syndrome type III and Ehlers-Danlos-syndrome hypermobility type): clinical description and natural history. Am J Med Genet. 2017;175(1):48-69. https://doi.org/10.1002/ajmg.c. 31538
- Bendik EM, Tinkle BT, Al-shuik E, et al. Joint hypermobility syndrome: a common clinical disorder associated with migraine in women. Cephalalgia. 2011;31(5):603-613. https://doi.org/10.1177/ 0333102410392606
- Castori M, Camerota F, Celletti C, Grammatico P, Padua L. Ehlers-Danlos syndrome hypermobility type and the excess of affected females: possible mechanisms and perspectives. Am J Med Genet A. 2010;152A(9):2406-2408. https://doi.org/10.1002/ajmg.a.33585
- Murray B, Yashar BM, Uhlmann WR, Clauw DJ, Petty EM. Ehlers-Danlos syndrome, hypermobility type: a characterization of the

- patients' lived experience. Am J Med Genet A. 2013;161(12):2981-2988. https://doi.org/10.1002/ajmg.a.36293
- Castori M, Sperduti I, Celletti C, Camerota F, Grammatico P. Symptom and joint mobility progression in the joint hypermobility syndrome (Ehlers-Danlos syndrome, hypermobility type). Clin Exp Rheumatol. 2011;29:9.
- Challal S, Minichiello E, Funalot B, Boissier M-C. Ehlers–Danlos syn-drome in rheumatology: diagnostic and therapeutic challenges. Joint Bone Spine. 2015;82(5):305-307. https://doi.org/10.1016/j.jbspin. 2015.04.002
- Castori M, Camerota F, Celletti C, et al. Natural history and manifes-tations of the hypermobility type Ehlers–Danlos syndrome: a pilot study on 21 patients. Am J Med Genet A. 2010;152(3):556-564. https://doi.org/10.1002/ajmg.a.33231
- Miglis MG, Schultz B, Muppidi S. Postural tachycardia in hypermobile Ehlers-Danlos syndrome: a distinct subtype? Auton Neurosci. 2017; 208:146-149. https://doi.org/10.1016/j.autneu.2017.10.001
- Scheper MC, de Vries JE, Verbunt J, Engelbert RH. Chronic pain in hypermobility syndrome and Ehlers–Danlos syndrome (hypermobility type): it is a challenge. J Pain Res. 2015;8:591-601. https://doi.org/ 10.2147/JPR.S64251
- Castori M, Dordoni C, Morlino S, et al. Spectrum of mucocutaneous manifestations in 277 patients with joint hypermobility syn-drome/Ehlers-Danlos syndrome, hypermobility type. Am J Med Genet C Semin Med Genet. 2015;169(1):43-53. https://doi.org/10.1002/ ajmg.c.31425
- Berglund B, Pettersson C, Pigg M, Kristiansson P. Self-reported qual-ity of life, anxiety and depression in individuals with Ehlers-Danlos syndrome (EDS): a questionnaire study. BMC Musculoskelet Disord. 2015;16:89. https://doi.org/10.1186/s12891-015-0549-7
- Castori M. Ehlers-Danlos syndrome, hypermobility type: an under-diagnosed hereditary connective tissue disorder with mucocutaneous, articular, and systemic manifestations. ISRN Dermatology; New York, NY. 2012.
- Castori M, Dordoni C, Valiante M, et al. Nosology and inheritance pattern(s) of joint hypermobility syndrome and Ehlers-Danlos syn-drome, hypermobility type: a study of intrafamilial and interfamilial variability in 23 Italian pedigrees. Am J Med Genet A. 2014;164(12): 3010-3020. https://doi.org/10.1002/ajmg.a.36805
- Ross J, Grahame R. Joint hypermobility syndrome. BMJ. 2011;342: 7167. https://doi.org/10.1136/bmj.c7167
- Forghani I. Updates in clinical and genetics aspects of hypermobile Ehlers Danlos syndrome. Balkan Med J. 2019;36(1):12-16. https://doi. org/10.4274/balkanmedj.2018.1113
- Beighton P, De Paepe A, Steinmann B, Tsipouras P, Wenstrup RJ. Ehlers-Danlos syndromes: revised nosology, Villefranche, 1997. Ehlers-Danlos National Foundation (USA) and Ehlers-Danlos support group (UK). Am J Med Genet. 1998;77(1):31-37. https://doi.org/10. 1002/(sici)1096-8628(19980428)77:1<31::aid-ajmg8>3.0.co;2-o
- Beighton P, de Paepe A, Danks D, et al. International nosology of her-itable disorders of connective tissue, Berlin, 1986. Am J Med Genet. 1988;29(3):581-594. https://doi.org/10.1002/ajmg.1320290316
- Castori M, Hakim A. Contemporary approach to joint hypermobility and related disorders. Curr Opin Pediatr. 2017;29(6):640-649. https:// doi.org/10.1097/MOP.0000000000000541
- McGillis L, Mittal N, Santa Mina D, et al. Utilization of the 2017 diag-nostic criteria for hEDS by the Toronto GoodHope Ehlers–Danlos syndrome clinic: a retrospective review. Am J Med Genet. 2020;182 (3):484-492. https://doi.org/10.1002/ajmg.a.61459
- Tinkle BT, Bird HA, Grahame R, Lavallee M, Levy HP, Sillence D. The lack of clinical distinction between the hypermobility type of Ehlers-Danlos syndrome and the joint hypermobility syndrome (a.k.a. hypermobility syndrome). Am J Med Genet A. 2009;149A(11): 2368-2370. https://doi.org/10.1002/ajmg.a.33070
- Castori M, Tinkle B, Levy H, Grahame R, Malfait F, Hakim A. A frame-work for the classification of joint hypermobility and related condi-tions. Am J Med Genet. 2017;175(1):148-157. https://doi.org/10. 1002/ajmg.c.31539
- Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol. 2018;18(1):143. https://doi.org/10.1186/s12874-018-0611-x
- Pham MT, Rajic A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods. 2014;5 (4):371-385. https://doi.org/10.1002/jrsm.1123
- Johns Hopkins University, Baltimore, MD. Online Mendelian Inheri-tance in Man, OMIM® Accessed January 19, 2021. World Wide Web URL: https://omim.org/
- National Organization for rare disorders. Ehlers Danlos Syndromes. NORD (National Organization for Rare Disorders). Accessed January 19, 2021. https://rarediseases.org/rare-diseases/ehlers-danlos-syndrome/.
- Orphanet. Hypermobile Ehlers Danlos syndrome. Accessed January 19, 2021. https://www.orpha.net/consor/cgi-bin/Disease_Search.php?lng=EN&data_id=4041&Disease_Disease_Search_diseaseGroup=ehlers-danlos&Disease_Disease_Search_diseaseType=Pat&Disease(s)/group%20of%20diseases=Hypermobile-Ehlers-Danlos-syndrome&title=Hypermobile%20Ehlers-Danlos%20syndrome&search=Disease_Search_Simple.
- Genetic Testing Registry. Ehlers-Danlos syndrome, type 3. Accessed January 19, 2021. https://www.ncbi.nlm.nih.gov/gtr/conditions/C0268337/#related-conditions
- Genetic and Rare Diseases Information Center (GARD). Ehlers-Danlos syndromes. Accessed January 19, 2021. https://rarediseases.info.nih.gov/diseases/6322/ehlers-danlos-syndrome.
- SNPedia. Ehlers-Danlos syndrome. Accessed January 19, 2021. https://www.snpedia.com/index.php/Ehlers-Danlos_syndrome
- Gensemer C, Burks R, Kautz S, Judge DP, Lavallee M, Norris RA. Hyp-ermobile Ehlers-Danlos syndromes: complex phenotypes, challenging diagnoses, and poorly understood causes. Dev Dyn. 2020;1:27-344. https://doi.org/10.1002/dvdy.220
- Narcisi P, Richards J, Ferguson SD, Pope FM. A family with Ehlers — Danlos syndrome type III/articular hypermobility syndrome has a gly-cine 637 to serine substitution in type III collagen. Hum Mol Genet. 1994;3(9):1617-1620. https://doi.org/10.1093/hmg/3.9.1617
- Zweers MC, Bristow J, Steijlen PM, et al. Haploinsufficiency of TNXB is associated with hypermobility type of Ehlers-Danlos syndrome. Am J Hum Genet. 2003;73(1):214-217.
- Zweers MC, Kucharekova M, Schalkwijk J. Tenascin-X: a candidate gene for benign joint hypermobility syndrome and hypermobility type Ehlers-Danlos syndrome? Ann Rheum Dis. 2005;64(3):504-505. https://doi.org/10.1136/ard.2004.026559
- Chiarelli N, Carini G, Zoppi N, et al. Transcriptome-wide expression profiling in skin fibroblasts of patients with joint hypermobility syn-drome/Ehlers-Danlos syndrome hypermobility type. PLoS One. 2016;

- 11(8):e0161347. https://doi.org/10.1371/journal.pone.0161347
- Zoppi N, Chiarelli N, Binetti S, Ritelli M, Colombi M. Dermal fibroblast-to-myofibroblast transition sustained by αvß3 integrin-ILK-Snail1/slug signaling is a common feature for hypermobile Ehlers-Danlos syndrome and hypermobility spectrum disorders. Biochim Biophys Acta Mol Basis Dis. 2018;1864(4):1010-1023. https://doi.org/ 10.1016/j.bbadis.2018.01.005
- Lyons JJ, Yu X, Hughes JD, et al. Elevated basal serum tryptase iden-tifies a multisystem disorder associated with increased TPSAB1 copy number. Nat Genet. 2016;48(12):1564-1569. https://doi.org/10. 1038/ng.3696
- Weerakkody RA, Vandrovcova J, Kanonidou C, et al. Targeted next-generation sequencing makes new molecular diagnoses and expands genotype-phenotype relationship in Ehlers-Danlos syndrome. Genet Med. 2016;18(11):1119-1127. https://doi.org/10.1038/gim.2016.14
- Henney AM, Brotherton DH, Child AH, Humphries SE, Grahame R. Segregation analysis of collagen genes in two families with joint hypermobility syndrome. Rheumatology. 1992;31(3):169-174. https:// doi.org/10.1093/rheumatology/31.3.169
- Hoffman GG, Dodson GE, Cole WG, Greenspan DS. Absence of apparent disease causing mutations in COL5A3 in 13 patients with hypermobility Ehlers–Danlos syndrome. Am J Med Genet A. 2008; 146A(24):3240-3241. https://doi.org/10.1002/ajmg.a.32586
- Yamada K, Watanabe A, Takeshita H, et al. Measurement of serum Tenascin-X in joint hypermobility syndrome patients. Biol Pharm Bull. 2019;42(9):1596-1599. https://doi.org/10.1248/bpb.b19-00168
- Demmler JC, Atkinson MD, Reinhold EJ, Choy E, Lyons RA, Brophy ST. Diagnosed prevalence of Ehlers-Danlos syndrome and hypermobility spectrum disorder in Wales, UK: a national electronic cohort study and case-control comparison. BMJ Open. 2019;9(11): e031365. https://doi.org/10.1136/bmjopen-2019-031365
- Byers PH. Vascular Ehlers-Danlos syndrome. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews®. Seattle, WA: University of Washington; 2019.
- Malfait F, Hakim AJ, De Paepe A, Grahame R. The genetic basis of the joint hypermobility syndromes. Rheumatology (Oxford). 2006;45 (5):502-507.
- Frank M, Albuisson J, Ranque B, et al. The type of variants at the COL3A1 gene associates with the phenotype and severity of vascular Ehlers-Danlos syndrome. Eur J Hum Genet. 2015;23(12):1657-1664. https://doi.org/10.1038/ejhg.2015.32
- Gazit Y, Jacob G, Grahame R. Ehlers–Danlos syndrome—hypermobility type: a much neglected multisystemic disorder. Rambam Maimonides Med J. 2016;7(4):34. https://doi.org/10.5041/RMMJ.10261
- Schalkwijk J, Zweers MC, Steijlen PM, et al. A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X deficiency. N Engl J Med. 2001;345(16):1167-1175. https://doi.org/10.1056/NEJMoa002939
- Demirdas S, Dulfer E, Robert L, et al. Recognizing the tenascin-X deficient type of Ehlers–Danlos syndrome: a cross-sectional study in 17 patients. Clin Genet. 2017;91(3):411-425. https://doi.org/10.1111/cge.12853
- Merke DP, Chen W, Morissette R, et al. Tenascin-X haploinsufficiency associated with Ehlers-Danlos syndrome in patients with congenital adrenal hyperplasia. J Clin Endocrinol Metab. 2013;98(2):379-387. https://doi.org/10.1210/jc.2012-3148
- El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. The Lancet. 2017;390(10108):2194-2210. https://doi.org/10.1016/ S0140-6736(17)31431-9
- Voermans NC, Verrijp K, Eshuis L, et al. Mild muscular features in tenascin-X knockout mice, a model of Ehlers-danlos syndrome. Connect Tissue Res. 2011;52(5):422-432. https://doi.org/10.3109/ 03008207.2010.551616
- Bagalad BS, Mohan Kumar KP, Puneeth HK. Myofibroblasts: master of disguise. J Oral Maxillofac Pathol. 2017;21(3):462-463. https://doi. org/10.4103/jomfp.JOMFP_146_15
- Watanabe A, Satoh K, Maniwa T, Matsumoto K-I. Proteomic analysis for the identification of serum diagnostic markers for joint hyper-mobility syndrome. Int J Mol Med. 2016;37(2):461-467. https://doi. org/10.3892/ijmm.2015.2437
- Chiarelli N, Carini G, Zoppi N, Ritelli M, Colombi M. Molecular insights in the pathogenesis of classical Ehlers-Danlos syndrome from transcriptome-wide expression profiling of patients' skin fibroblasts. PLoS One. 2019;14 (2):e0211647. https://doi.org/10.1371/journal.pone.0211647
- He Y, Liu X. The tumor-suppressor gene LZTS1 suppresses hepatocel-lular carcinoma proliferation by impairing PI3K/Akt pathway. Biomed Pharmacother. 2015;76:141-146. https://doi.org/10.1016/j.biopha. 2015.10.006
- Kropp M, Wilson SI. The expression profile of the tumor suppressor gene Lzts1 suggests a role in neuronal development. Dev Dyn. 2012; 241(5):984-994. https://doi.org/10.1002/dvdy.23777
- The Ehlers Danlos Society. HEDGE study- hypermobile Ehlers-Danlos genetic evaluation study. The Ehlers Danlos Society Accessed January 21, 2021. https://www.ehlers-danlos.com/hedge/
- Díaz-Santiago E, Claros MG, Yahyaoui R, et al. Decoding neuromus-cular disorders using phenotypic clusters obtained from co-occurrence networks. Front Mol Biosci. 2021;8:635074. https://doi. org/10.3389/fmolb.2021.635074
- Bertoli-Avella AM, Kandaswamy KK, Khan S, et al. Combining exome/genome sequencing with data repository analysis reveals novel gene-disease associations for a wide range of genetic disorders. Genet Med. 2021. https://doi.org/10.1038/s41436-021-01159-0

How to cite this article: Scicluna K, Formosa MM, Farrugia R, Borg I. Hypermobile Ehlers–Danlos syndrome: A review and a critical appraisal of published genetic research to date. Clinical Genetics. 2021;1-12. https://doi.org/10.111