Identifying and characterising the genes and proteins underlying Coats disease and Familial Exudative Vitreoretinopathy (FEVR)

Denisa Džulová

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds
School of Medicine

October 2018

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First of all, I would like to express my gratitude to my supervisors Carmel, James and Ian for giving me the opportunity to challenge myself and carry out this PhD. I would like to extend my thank you to James, who had especially been there through all the ups and downs, and for that I am forever grateful.

I would also like to thank everyone on Level 8 who has made the last 3 years full of cocktails, friendships, support and never-ending laughs. Marilena, Jo, Claire, Carla, Jose, Layal, Cam, Nikki, Amy, Eleanor, Evi, Alice, Becky, Mark, Tom, Alexia and Ewa: my time here would never be so unforgettable if it wasn't for the people I shared it with.

One of my biggest thank you must go to you, Emma. I'm finding it hard to describe just how much of a support system you had been throughout this rollercoaster and I want you to know that I wouldn't be finishing this journey if it wasn't for you. I'm beyond thankful for the amazing memories and friendship PhD has given us.

Michelle, you have been there from the day I screamed in your face on the very first day of our scientific careers. We went through it all, from heartbreaks to dancing the nights away to having the best possible friendship I could have ever asked for, and for that I would like to thank you.

Gratitude I can't really express in words goes to you, Dylan. You've been my rock through all this since day one and I'm sorry for everything I've put you through. There are words that won't make it between these tiny lines but just know that I feel extremely lucky to share life together.

Last but not least, to my parents, Jarmila and Peter. Thank you for giving me the life I've always wanted and for giving me the chance to be the best version of myself even if you had to sacrifice so much. This thesis belongs to both of you and your years of hard work and shaping me to the person that I am today. Z celého srdca Vám ďakujem.

Abstract

Coats disease is an idiopathic, non-hereditary condition characterised by retinal telangiectasia, exudation and, in severe cases, total retinal detachment. It usually manifests as a unilateral disease and mainly affects young males. The Coats phenotype shares many features with familial exudative vitreoretinopathy (FEVR), a Mendelian disorder frequently associated with mutations affecting the Norrin-β-catenin signalling pathway. Somatic mutations in *NDP*, the gene encoding Norrin, have been previously identified in Coats disease, but these only account for a small proportion of cases.

In this study, whole-exome sequencing (WES) was used to analyse DNA extracted from seven formalin-fixed, paraffin-embedded Coats eyes to uncover new genes associated with this disorder. A novel heterozygous missense mutation in *LRP5* was found in one Coats eye, c.2951A>G; p.(Tyr984Cys). *LRP5* encodes a co-receptor for the Norrin-β-catenin pathway. A cell based-luciferase reporter assay for β-catenin transcription (TOPflash) showed that the mutant p.(Tyr984Cys) LRP5 resulted in significantly reduced levels of reporter activation compared to wild-type LRP5 ($p = \leq 0.05$). This is the first reported case of an *LRP5* mutation in Coats disease.

Given that *LRP5* and *NDP* are both mutated in FEVR, WES was undertaken in a cohort of 20, predominantly pre-screened, FEVR patients to identify novel FEVR mutations and genes, and to facilitate the identification of additional pathogenic mutations from the Coats dataset. This analysis led to the identification of mutations in known FEVR genes in five cases, a putative

mutation in a recently reported transcript of *LRP5* and a potential mutation in a novel FEVR gene.

These findings strengthen the evidence that Coats disease is caused by defects in the Norrin-β-catenin pathway and is a somatic form of FEVR. They also impact FEVR families by facilitating accurate molecular diagnosis and genetic counselling. Furthermore, the data generated will assist the development of therapies for these and other retinal vascular disorders.

Abbreviations

A	Adenine
adFEVR	Autosomal dominant FEVR
AMD	Age-macular degeneration
AP	Alkaline phosphatase
APC	Adenomatous polyposis coli
arFEVR	Autosomal recessive FEVR
ATOH7	Atonal homolog 7
bHLH	Basic helix-loop-helix
BAM	Binary alignment map
BBB	Blood brain barrier
BCA	Bicinchoninic acid assay
BEST1	Bestrophin 1
BLAST	Basic local alignment tool
BMD	Bone mineral density
bp	Base pair
BWA	Burrows-wheeler aligner
C	Cytosine
CADD	Combined annotation dependent depletion
CDMMR	Chorioretinal dysplasia, microcephaly and mental retardation
cDNA	Complementary DNA
CK1α	Casein kinase 1 alpha
CNS	Central nervous system
CNV	Copy number variant
CRD	Cysteine rich domain
CTNNB1	Beta-1 catenin
dbSNP	Single nucleotide polymorphism database
DECIPHER	Database of chromosomal imbalance and phenotype in Humans using Ensembl resources
DEXA	Dual energy X-ray absorptiometry
dH_2O	De-ionised water
Dkk1	Dickkopf inhibitor 1
DKO	Double knock-out
DLG1	Disc large MAGUK scaffold protein 1
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotides
DR	Diabetic retinopathy
DVL	Dishevelled
EC	Endothelial cells

EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ER	Endoplasmic reticulum
EVS	Exome variant server
ExAC	Exome aggregation consortium
ExoSAP-IT	Exonuclease I shrimp alkaline phosphatase
EST	Expressed sequence tags
FA	Fluorescein angiography
FCS	Fetal calf serum
FEVR	Familial exudative vitreoretinopathy
FF	Fresh frozen
FFPE	Formalin-fixed, paraffin-embedded
FZD	Frizzled receptor
FZD4	Frizzled 4
G	Guanine
GATK	Genome analysis toolkit
GCL	Ganglion cell layer
gDNA	Germline DNA
GFP	Green fluorescent protein
gnomAD	Genome aggregation database
GSK3β	Glycogen synthase kinase 3 beta
H&E	Hematoxylin and Eosin
HeLa	Henrietta Lacks
HEK293	Human embryonic kidney 293
HSMM	Hotshot mastermix
HUVEC	Human umbilical vein endothelial cells
IF	Immunofluorescence
IGV	Integrative genomics viewer
ILM	Inner limiting membrane
INL	Inner nuclear layer
Indel	Insertion/deletion
IPL	Inner plexiform layer
iPSC	Induced pluripotent stem cells
ISPD	Isoprenoid synthase domain-containing protein
IVFA	Intravenous fluorescein angiography
kb	Kilobase
KCNJ13	Potassium voltage-gated channel subfamily J member13
kDa	Kilodalton
KIF11	Kinesin family member 11
KO	Knock-out
LB	Luria-Bertani
LCA	Leber congenital amaurosis
LCI	Live cell imaging
LDL-R	Low density lipoprotein receptor

LEF	Lymphoid enhancing factor
LEL	Large extracellular loop
LoF	Loss of function
LOVD	Leiden open variation database
LRP5	Low-density lipoprotein receptor-related protein 5
LRR	Leucine-rich repeats
M	Molar
MAF	Minor allele frequency
Mb	Megabase
MCLMR	Microcephaly with or without chorioretinopathy, lymphedema, or mental retardation syndrome
MIM	Mendelian inheritance in man
ml	Millilitre
MLCRD	Microcephaly, lymphedema & chorioretinal dysplasia
mM	Millimolar
MRI	Magnetic resonance imaging
mRNA	Messenger RNA
Mut	Mutant
NCBI	National center for biotechnology information
ND	Norrie disease
NDP	Norrie disease pseudoglioma
NER	Nucleotide excision repair
NFL	Nerve fibre layer
ng	Nanogram
NGS	Next generation sequencing
NMD	Nonsense-mediated decay
OCT	Optical coherence tomography
OFA	Oral fluorescein angiography
ONL	Outer nuclear layer
OPL	Outer plexiform layer
OPPG	Osteoporosis-pseudoglioma syndrome
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFV	Persistent foetal vasculature
PHPV	Persistent hyperplastic primary vitreous
PL	Photoreceptor layer
pmol	Picomoles
PROVEAN	Protein variant effect analyser
PTT	Protein truncation test
QC	Quality control
qPCR	Quantitative real-time PCR
RCBTB1	RCC1 and BTB domain containing protein 1
RGC	Retinal ganglion cells

RNA	Ribonucleic acid
ROP	Retinopathy of prematurity
RP	Retinitis pigmentosa
RPE	Retinal pigment epithelium
RT-PCR	Reverse transcription PCR
SAM	Sequence alignment map
SDM	Site-directed mutagenesis
SEL	Small extracellular loop
SIFT	Sorting intolerant from tolerant
SMART	Simple modular architecture research tool
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SOC	Super optimal broth with catabolite repression
T	Thymine
TAE	Tris-acetate-EDTA
TBE	Tris HCl-borate EDTA
Taq	*Thermus aquaticus* derived DNA polymerase
TCF	T cell factor
TD-PCR	Touch down polymerase chain reaction
TE	Tris EDTA
TM	Transmembrane domain
Tris	Tris (hydroxymethyl) aminomethane
TSPAN12	Tetraspanin-12
U2OS	Human osteosarcoma cell line
UCSC	University of California Santa Cruz
UDG	Uracil-DNA glycosylase
USER	Uracil-specific excision reagent
UTR	Untranslated region
V	Volt
VCAN	Versican
VCF	Variant call format
VEGF	Vascular endothelial growth factor
VEP	Variant effect predictor
gVCF	Genomic variant call format
WB	Western blot/blotting
WES	Whole exome sequencing
WGA	Whole genome amplification
WG	Weeks gestation
WGS	Whole genome sequencing
WT	Wild-type
μl	Microliter
μM	Micromolar
ZNF408	Zinc finger protein 408

Table of Contents

List of Figures

List of Tables

1 Chapter: Introduction

1.1 General Overview

In recent years, the use of the next generation sequencing (NGS) technology has allowed for efficient sequencing of an individual's exome or genome, which ultimately leads to faster diagnosis and, if applicable, treatment. DNA sequencing of the entire human exome or genome has revolutionised research with the increased number of disease-causing genes that have been identified (Bahassi and Stambrook, 2014). There has been a significant rise in identification of retinal disease genes since the arrival of NGS technologies (Eisenberger et al., 2013; Carrigan et al., 2016; Ellingford et al., 2016; Chaitankar et al., 2016). Genetics is not only a powerful tool for determining the underlying cause of eye disease but also for understanding the mechanisms leading to pathogenesis. The identification of novel defective genes also increases the understanding of normal development and function of the eye. This in turn, leads to faster diagnoses, the establishment of treatments and facilitates accurate genetic counselling for the affected patients and their families.

However, there are still a number of eye diseases that remain to be elucidated at their genetic basis. One of those diseases is Coats disease, a disorder affecting the retinal vasculature and the focus of this thesis. Vasculature plays a significant role in pathogenesis of many diseases and, therefore, creates a very important avenue for research (Rajendran et al., 2013). In particular,

abnormal retinal vasculature underlies a significant proportion of all eye diseases and, therefore, identifying genes that cause retinal vascular defects when mutated, and studying their mechanisms, is essential. Fortunately, the retina serves as a powerful tool for studying the vasculature as it is easily accessible and visualised. Upon the identification of defective genes, there are still many points to consider. The understanding of the normal or abnormal function of proteins that are encoded by the defective genes is very crucial for the development of therapies for these disorders.

1.2 Structure of the human eye

The fully developed human eye is composed of three main layers (Figure 1.1). The external layer is formed by the sclera and cornea. The sclera and cornea provide support and maintain the shape of the eye (McLananahan, 2008). The middle layer is composed of two parts, the anterior and posterior chamber. The anterior chamber consists of the iris and ciliary body whereas the posterior consists of the choroid. The iris is responsible for the amount of light that enters the eye by contracting or relaxing the size of the pupil. The ciliary body consists of a muscle that controls the shape of the lens and it also produces the aqueous humour that is present within the anterior and posterior eye chamber. The lens is located behind these chambers and its role lies in the refraction and focus of the light onto the retina. Immediately behind the lens is the vitreous, which is responsible for maintaining the structure and shape of the eye. The choroid supplies nourishment to the eye and is also responsible for the removal of waste products (Kolb et al., 2005). The internal layer is the retina (Kolb et al., 2005) which detects the light and converts it into a signal

sent to the visual cortex through the process of visual phototransduction (Section 1.2.1).

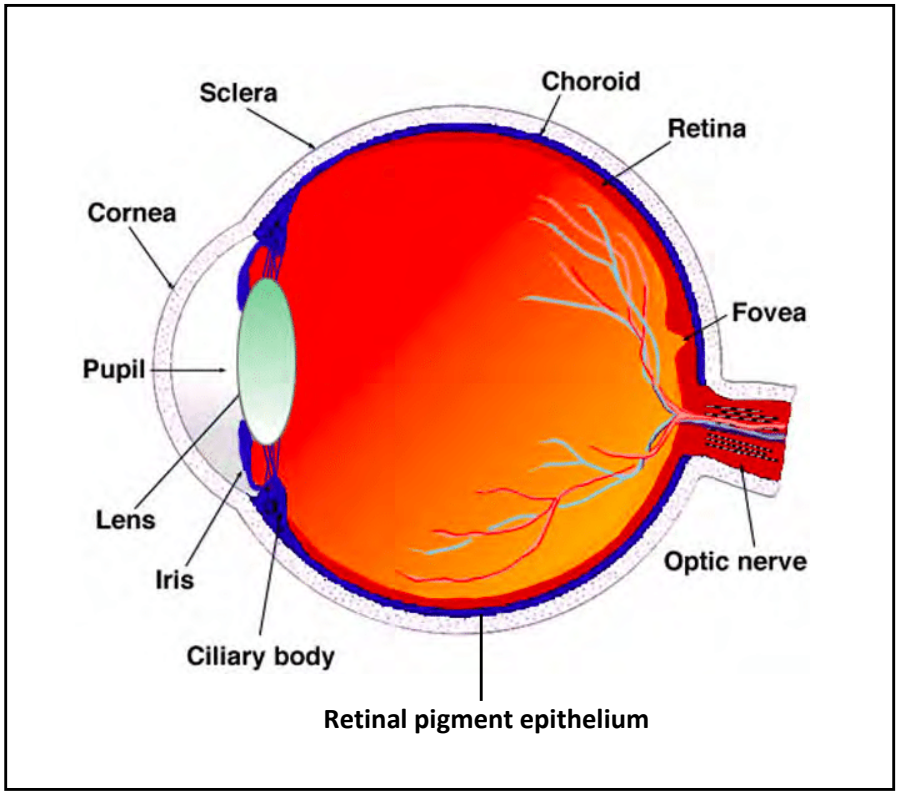


Figure 1.1 Structure of the human eye.
Sagittal section of the human eye with the representative structures. Figure adapted from Webvision (www.webvision.med.utah.edu). Image used with non-exclusive rights under Attribution, Non-commercial, No Derivative Works Creative Commons license.

1.2.1 Structure of the retina

The development of the retina begins as an outgrowth of the brain. This makes the retina a very good model system for vascular development of the central nervous system (CNS) (Hughes et al., 2000). It is a metabolically active organ and it contains the highest consumption of oxygen per gram in the body (Warburg, 1928; Anderson, 1968).

The vertebrate neural retina is composed of seven cell types. Five cell types are neurons and three are types of glia. Photoreceptors, horizontal, bipolar,

amacrine and ganglion cells are all neurons. The three types of glia consist of Müller cells, microglia and astrocytes. These cell types are organised into layers that are separated by two plexiform layers to form a structured mature retina (Stenkamp, 2015). The layers are, starting from the back of the retina - retinal pigment epithelium (RPE), photoreceptor layer (PL), outer nuclear layer (ONL), outer plexiform layer (IPL), inner nuclear layer (INL), inner plexiform layer (OPL), ganglion cell layer (GCL) and nerve fibre layer (NFL) (Figure 1.2).

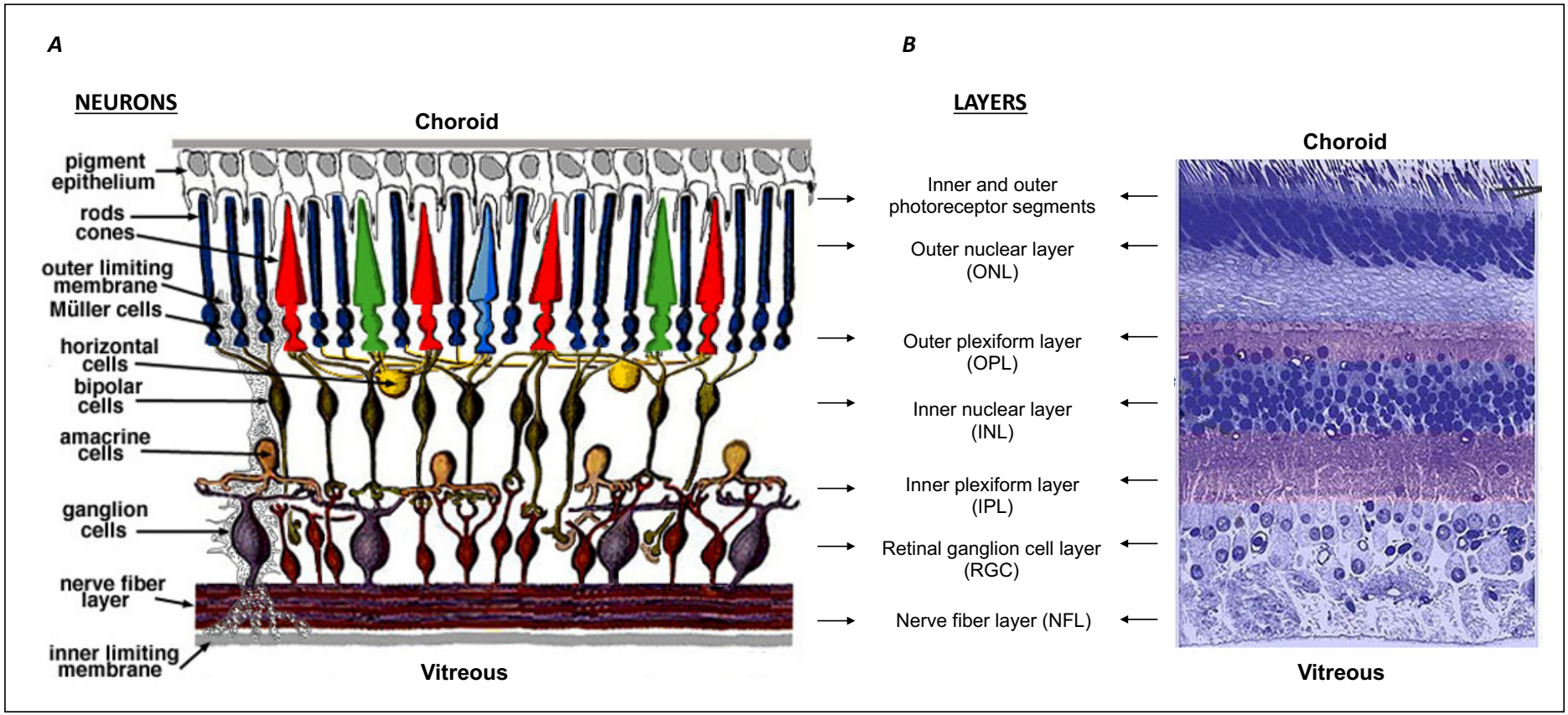


Figure 1.2 Representation of the retinal layers and neurons.
(A) Schematic representation of all the cell types in the retina. All cell types and corresponding layers are identified. **(B)** Light micrograph of the central human retina at a vertical cross section. Both images were adapted from Webvision (www.webvision.med.utah.edu). Image used with non-exclusive rights under Attribution, Non-commercial, No Derivative Works Creative Commonslicense.

Beginning from the choroid at the back of the retina, the first layer is the RPE which is responsible for a number of functions including the recycling of the visual pigment, phagocytosis of the photoreceptor outer segments and also the absorption of excess light. In the next layer moving towards the vitreous, is the photoreceptor layer composed of the outer segment of the photoreceptors. Their cell bodies and nuclei reside in the ONL. Photoreceptors are composed of rods and cones which detect the light and subsequently convert it into visual signals by phototransduction. Retinal phototransduction is a process through which the light entering the eye is converted into a form of visual information that can be perceived by the brain. It is a process of converting light energy into electrochemical energy (Mannu, 2014). The retina is crucial for phototransduction as it contains light sensitive cells called photoreceptors, rods and cones. Pigment molecules in photoreceptors contain aldehyde vitamin A (retinaldehyde) and opsin. Cones are further specified into three subtypes based on which opsin they acquire (Nickle and Robinson, 2007). Cones absorb light at different frequencies based on the opsins and then convert it differentially to the brain (Alpern, Bastian & Moeller, 1982). Cone cells are concentrated in the fovea, have higher visual sensitivity in light conditions and are used for central vision. Rod cells function in dark conditions, are found throughout the retina and are mainly responsible for peripheral vision. The pigment that is present in rod cells is called rhodopsin (Tam and Moritz, 2009). Together, they convey the visual signal to the bipolar cells in the OPL. Bipolar cells are responsible for transferring the signal from the photoreceptors to RGCs. They also interact with amacrine and horizontal cells to ensure that visual signals are correctly relayed to the brain (Stenkamp,

2015). The INL contains the cell bodies of amacrine cells, horizontal cells and bipolar cells. These cells transmit the interactions and information passed on from the photoreceptors to the RGCs (Stenkamp, 2015). RGCs reside in the innermost layer of the retina closest to the vitreous. Their axons spread along the vitreal surface in the NFL. RGCs receive visual signals from bipolar and amacrine cells and convey them to the brain via their axons which make up the optic nerve (Ye et al., 2010).

The three types of glia include the microglia which migrate from the optic disc to the vessels. Initially, a set of microglia cells grow within the retina before the vasculature system develops, followed by a second set of microglia cells that migrate into the retina at the same time as the vasculature is establishing (Provis, 2001). Astrocytes are the second type of glia. They enter the retina through the optic nerve, are only present in vascularised retina and their cell bodies are located in the NFL (Ling and Stone, 1988; Provis, 2001). Astrocytes were found to migrate into the neuroretina during retinal angiogenesis, process of vessels sprouting from pre-existing vessels, along with endothelial cells to guide the vasculature process and also to establish the blood-brain barrier (Tao and Zhang, 2014). The last glial cell type that is present within the retinal structure are the Müller glia cells which span the entire length of the retina. It is the only glia cell type in the retina that is derived from retinal progenitor cells and their function lies in the establishment of the metabolic and homeostatic protection of the neurons (Hoon et al., 2014; Livesey et al., 2001; Nathans, 1999; Stenkamp, 2015; Ye et al., 2010). The stalk connecting the retina and the brain is the optic nerve, which is made up of axons from the RGCs (Ye et al., 2010).

The retinal vasculature first develops as choroidal vasculature that is located beyond the RPE. Once the choroidal vasculature regresses, the retinal vasculature begins to emerge from the optic nerve to create plexuses at three different depths within the retina (Section 1.4.2).

1.3 Development of the human eye

The human eye starts to develop at week three of the embryonic development. The retina develops from the retinal neuroepithelium which is derived from the neural tube (Stenkamp, 2015). At this stage the single eye field forms from the neural tube and then develops bilaterally, creating the optic vesicles and the optic cup (Graw, 2010). The optic vesicles come into contact with the surface ectoderm which induces the invagination of the optic vesicles to generate the bi-layered optic cups (Fuhrmann, 2010). The optic cups begin to grow to establish the main structure of the eye. At 5 weeks gestation (WG), the surface ectoderm in contact with the optic vesicles forms the lens placode. The lens placode then detaches to form the lens and the surface ectoderm seals over the lens to generate corneal epithelium. At 7 WG, the outer layer of the optic cup develops into the RPE. The neural retina develops from the inner layer of the optic cup (Fuhrmann, 2010) (Figure 1.3).

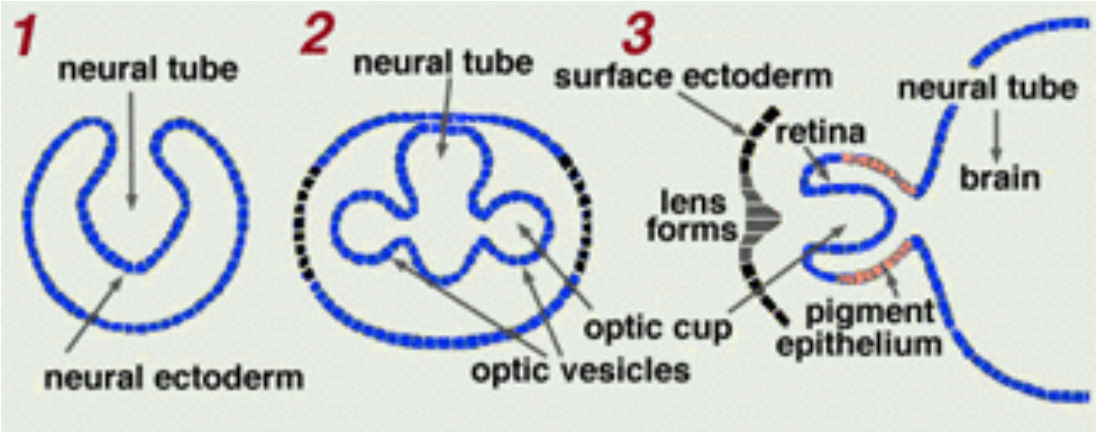


Figure 1.3 Schematic of the human eye development process.
(1) The eye develops from the neural tube. **(2)** Both eyes are formed from the optic vesicles at each side of the neural tube. **(3)** The optic cups are then formed from the optic vesicles. The outside part of the optic cup forms into retinal pigment epithelium whilst the inside of the cup forms the retina. The lens forms from the surface ectoderm. Image from webvision (www.webvision.med.utah.edu). Image used with non-exclusive rights under an Attribution, Non-commercial, No Derivative Works Creative Commons license.

The differentiation of the cells forming the neural retina begins at day 47 after gestation until the ninth month of gestation (Graw, 2003). The cells are initially differentiated into ganglion cells, horizontal cells, amacrine cells and then cone photoreceptors. Differentiation of rod photoreceptors begins before birth and continues shortly after birth. Bipolar cells and Müller glia cells are differentiated within the first week of birth (Marquardt and Gruss, 2002). The optic nerve forms from the growth of the RGCs. The establishment of the optic nerve is generated by axons forming from RGCs and their growth along the inner wall of the optic stalks to the brain. Before the optic nerves reach the visual cortex, the optic chiasm is formed by the nerve fibres from each eye partially crossing at the midline forming an X-shaped like structure (Jeffery, 2001).

1.4 Vascularisation of the human eye

The retina of the human eye spans the back inner surface of the eyeball. It is a highly complex and organised tissue with neuronal and vascular layers

(Figure 1.2). For its proper function, all cell types must be organised into the correct layers (Section 1.2.1) to ensure that the visual stimuli can be passed from the photoreceptors to the brain. The vasculature within the retina comprises of three vascular plexuses, which are crucial for the proper function of the tissue. The primary vascular plexus develops first from the optic nerve and spreads through the inner surface of the retina. As it continues its growth, it further sprouts into two connected intra-retinal plexuses that are flanking the INL layer on both sides (Section 1.4.2). Abnormalities of this vascular system can have pathogenic effects. If the plexuses do not form properly, or if they are present in areas of the retina that are usually avascular, abnormal vascularization and vision defects can occur (Wechsler-Reya and Barres, 1997).

1.4.1 The hyaloid vasculature

The ocular vasculature is established in human development during the second and third trimester. Prior to this, the retina is supplied by the hyaloid and choroidal vessels (Dorrell et al., 2007) (Figure 1.4). The fetal vasculature, which includes the hyaloid vasculature, starts to develop at 4-6 weeks gestation (WG) (Lutty and McLeod, 2018). Hyaloid vasculature does not serve as a permanent blood supply to the retina and it is fully developed after 10 WG, begins to expand around 12 WG by angiogenesis and serves as a blood supply until a retinal vasculature begins to form (Zhu et al., 2000). At this point, it is the main source of nutrients for the developing eye in utero until the retinal vasculature has fully developed. The blood of the hyaloid vasculature network enters the eye via the hyaloid artery through the optic nerve. The hyaloid network runs into the vitreous and the blood is carried through the vasculature

to the choroidal vein where it is collected and carried out of the eye. The hyaloid vasculature regresses before birth and the ocular vasculature becomes established (Provis, 2001) (Figure 1.4).

The primary plexus of the retinal vasculature is formed first through vasculogenesis, the process of de novo formation of vessels through differentiation of endothelial cells from the optic nerve head (Hughes et al., 2000). The secondary deeper plexuses (or intraretinal capillary plexuses) are formed by angiogenesis. The intraretinal plexuses are developed from the primary plexus and span the length of the retina (Fruttiger, 2007).

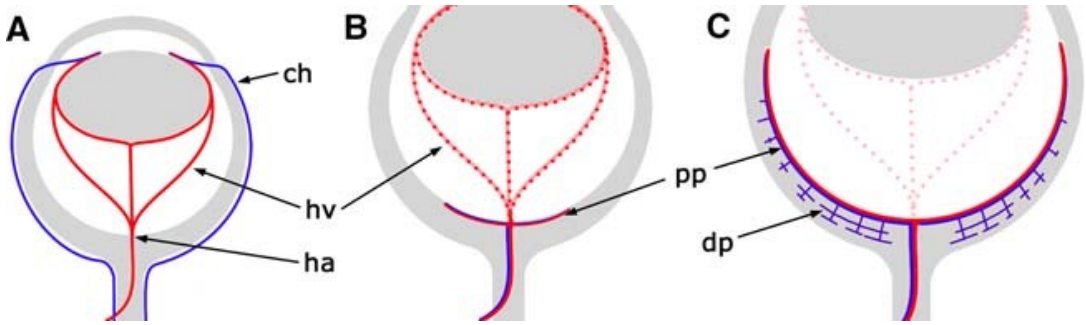


Figure 1.4 Regression of the hyaloid vasculature and the development of the retinal vasculature in the human eye.
(A) The hyaloid vasculature (hv) supplies the blood to the developing retina through the hyaloid artery (ha). The blood exits the eye through the choroidal net (ch). **(B)** The hyaloid vasculature is starting to regress and the primary plexus (pp) starts to develop from the optic nerve. **(C)** A secondary deeper plexus (dp) starts to develop from the pp as the growth of the retina progresses. Image used with permission from Fruttiger (2007).

The process of spontaneous regression of hyaloid vasculature was recently found to be triggered by neurons in mouse models through the regulation of vascular endothelial growth factor receptor 2 (VEGFR2) (Yoshikawa et al., 2016). Previously, a study by Lang & Bishop (1993) showed the involvement of macrophages and their role in remodelling the hyaloid vasculature. The retreat of the hyaloid vessels was documented to be controlled by factors such as Wnt7b, macrophage secreting factor and a signal required for WNT

signalling to initiate apoptosis and promote cell death within the endothelium (Lobov et al., 2005).

Persistent hyperplastic primary vitreous (PHPV) (MIM 221900) is a disease characterised by the improper regression of the hyaloid vasculature (Haddad et al., 1978). This phenotype is also seen in many FEVR patients and was shown to be present in mouse studies of the known FEVR genes, frizzled-4 (*Fzd4*), norrin (*Ndp*), low-density lipoprotein receptor-related protein 5 (*Lrp5*) and tetraspanin-12 (*Tspan12*) (Section 1.8.2.1- 1.8.2.4) (Kato et al., 2002; Xu et al., 2004; Junge et al., 2009).

1.4.2 Vascularisation of the retina

As the hyaloid vasculature regresses, the human retinal vasculature is beginning to develop by vasculogenesis. The formation of the first vessels within the retina is controlled by vasculogenesis, whereas angiogenesis is responsible for the expansion of the vascular network (Hughes et al., 2000). The formation of vessels by vasculogenesis occurs by the assembly of CXCR4+/CD39+ (progenitor markers for hematopoietic and endothelial cells that are known to be involved in vasculogenesis) angioblasts (Lutty & McLeod, 2018; McLeod et al., 2006). Recently, it was also shown that vascular progenitors could be using Müller cell Notch1 or axonal neuropilinin-1 (NRP1) for guidance of semaphorin 3A-expressing angioblasts (Lutty and McLeod, 2018).

The distribution of the retinal blood vessels is influenced by the formation and distribution of astrocytes (Watanabe and Raff, 1988). Astrocytes in the retina

develop from precursors that express the transcription factor Pax2. The fate of the retinal astrocytes is influenced by the expression of the platelet derived growth factor receptor alpha (PDGFRα) marker. This expression distinguishes between the retinal astrocytes, as they express the PDGFRα marker, and optic nerve astrocytes (Mudhar et al., 1993). The proliferation of the retinal astrocytes is mediated by the platelet derived growth factor A (PDGFA) secreted by RGCs (Fruttiger et al., 2000). The expansion of the astrocytes continues until it forms a template for the vascular network. As the blood vessels start to develop, they span the full size of the retina, apart from the fovea, which remains avascular (Schnitzer, 1987). With the formation of the vascular network, the primary vascular plexus develops from the optic nerve and then sprouts into two connected intra-retinal plexuses that are flanking the INL (Figure 1.5) (Provis, 2001). This process, once complete, establishes the complete vasculature of the retina.

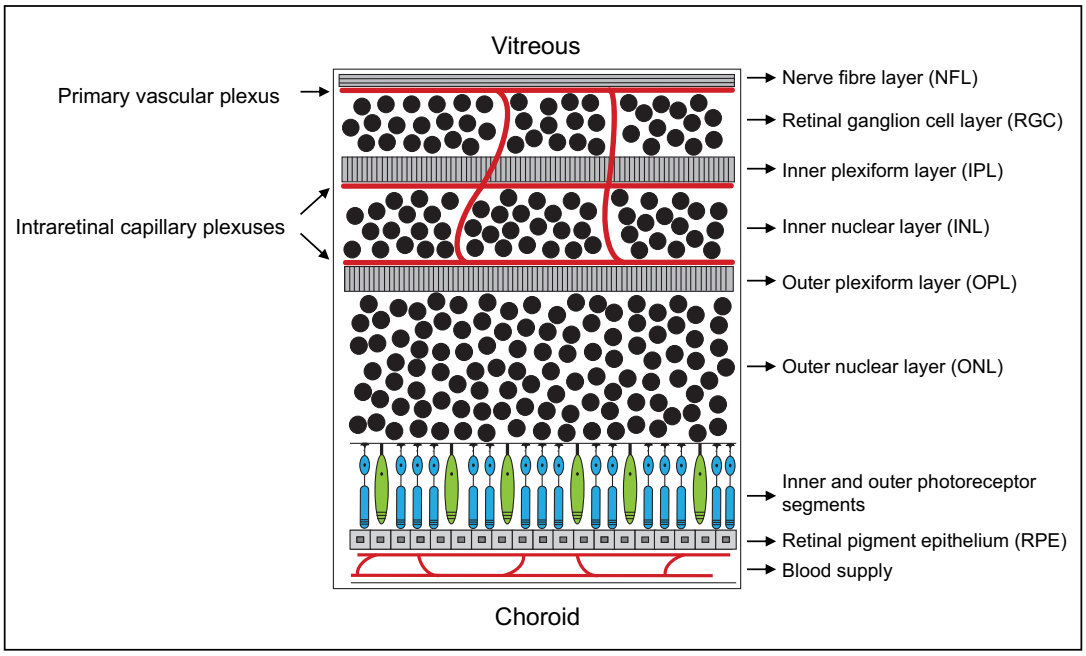


Figure 1.5 Schematic diagram of the retinal vasculature.
The primary vascular plexus is located on the vitreous surface and connected through the ganglion cell layer (GCL) and inner plexiform layer (IPL) to two intra-retinal plexuses. These two intra-retinal capillary beds span the inner nuclear layer (INL). All layers of the retina are identified.

1.5 Coats Disease

In 1908, George Coats described a number of cases with unilateral retinal vascular abnormalities, exudation and retinal detachment (Coats, 1908). Four years later, in 1912, Leber described a very similar condition that presented with multiple retinal aneurysms and retinal detachment, named Leber's miliary aneurysms. However, this condition had no abnormal vessel growth (Leber, 1912). Further in 1956, Reese noticed the overlap between the two conditions and concluded that Leber's miliary aneurysms described the progression of Coats disease (Reese, 1956). Ever since this revelation, Coats telangiectasia and Leber's retinal aneurysms have been termed Coats disease (MIM 300216) (Shields et al., 2001a).

Coats disease has been widely characterised as a rare, idiopathic condition with abnormal growth of the retinal vasculature (also known as telangiectasia) with intraretinal and subretinal exudation, which leads to lipid accumulation. In severe cases, it can progress to partial or full retinal detachment (Shields et al., 2001b) (Figure 1.6).

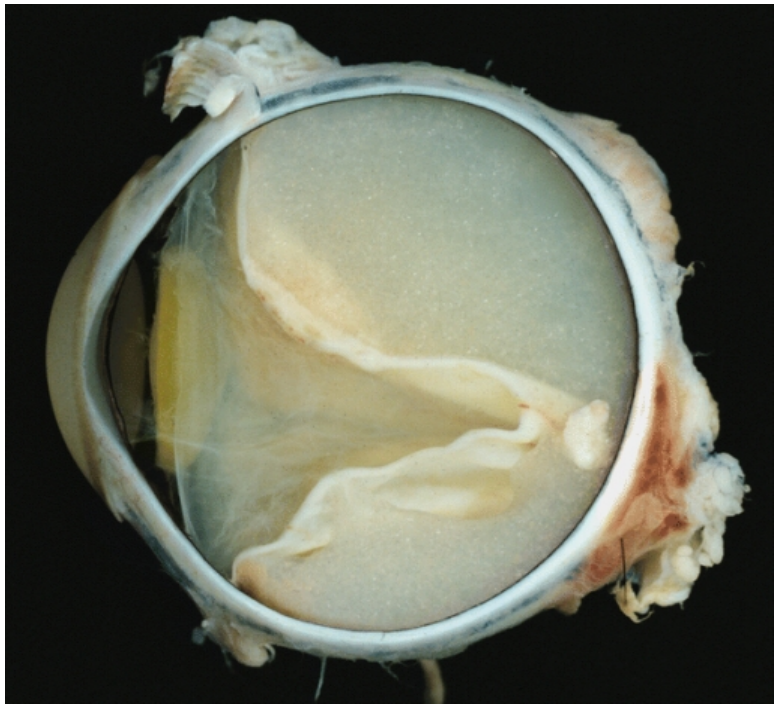


Figure 1.6 Cross-section image of an enucleated eye from a Coats disease patient.
A retinal detachment can be seen with the retina attached at the optic nerve head forming a characteristic funnel shape. Image sourced from Wikipedia, with permission: The Armed Forces Institute of Pathology (AFIP) [CC BY 2.0 (http://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons.

Most cases present unilaterally but there have been cases documented with bilateral disease (McGettrick and Loeffler, 1987; Alexandridou and Stavrou, 2002), although some of these cases presented with other associated diseases (Bass et al., 2011; Gursoy et al., 2015). The majority of cases of Coats disease reported have occurred in young males (76%, 114/150)

(Shields et al., 2001a). The disease generally presents itself in the first decade of life, however, there are cases reported of the disease occurring after 35 years of age, but these are usually milder (Smithen et al. 2005). The estimated frequency of Coats disease is 0.09 in every 100,000 people in the UK (Morris et al., 2010).

The pathogenesis of Coats disease has also been elucidated over recent years. The abnormal vessels are caused by a build-up of plasma in the vessel wall due to lipid exudates which makes vessels necrotic and abnormal. This is thought to be due to the endothelial cells in the blood retinal barrier not fully developing (Egbert et al., 1976). This phenomenon was also documented in a study by Fernandes et al. (2006) where the number of endothelial cells in the telangiectatic retinal vessels was reduced in Coats disease eyes. In the advanced stage of the disease, retinal detachment is caused by contracting vitreous condensations (Reichstein and Recchia, 2011). Shields et al. (2001b) have proposed that Coats disease is an advancing disease that might not have any symptoms in early life and can progress to severe stages.

1.5.1 Diagnosis

Coats disease was shown to be a progressive disease that in early years can be asymptomatic. The presenting features of Coats disease include reduced vision, strabismus and leukocoria (Shields et al., 2001a). Retinal telangiectasia is the key feature that is present at the early stages of the disease. However, as the disease develops, vascular abnormalities, exudation with lipid deposits and retinal detachment can be seen on fundus examination (Ghorbanian et al. 2012 & Shields et al., 2001b). Coats disease is usually diagnosed at a routine ophthalmologic examination. However, in most cases, its diagnosis requires

more advanced forms of detection. Fluorescein angiography, optical coherence tomography (OCT) and magnetic resonance imaging (MRI) are frequently used (Edward et al., 1998).

Fluorescein angiography is one of the most common diagnostic procedures for Coats disease. If the ophthalmic examination does not reveal any definitive phenotypes, fluorescein angiography is used (Gass, 1968b). The fluorescent dye (fluorescein) is intravenously injected into the patient (IVFA) or can be taken orally (OFA). An angiogram then detects the fluorescence that is emitted. It helps to visualise the structural changes within the vasculature and, in particular, locate the permeability changes in the affected vessels. This aids in detecting the severity of the disease, such as the presence of exudates within the intraocular space or beneath the retina (Gass, 1968a).

OCT or MRI are often used if the disease is very severe and other differential diagnoses, such as retinoblastoma, need to be dismissed (Section1.5.4.1). In a study by Edward et al. (1998), OCT of a severe case of Coats disease showed denser substances posterior to the vitreous between the vitreous and exudation but the precise phenotype can be quite variable which makes it difficult to establish an accurate diagnosis. Edward et al. (1998) concluded that the MRI is more precise and effective in detecting any masses within the ocular space.

1.5.2 Classification

Coats disease has been categorised into 5 different phenotypic stages (Shields et al., 2001a) (Figure 1.7). The stages are divided into:

Stage 1 - telangiectasia of the retinal vessels only.

Stage 2 - telangiectasia and exudation. Stage 2 was further divided into stage 2A - extrafoveal exudation and 2B - foveal exudation.

Stage 3 - exudative retinal detachment. This stage was further classified into 3A - subtotal retinal detachment and 3B - total retinal detachment.

Stage 4 - total detachment and secondary glaucoma.

Stage 5 - blind eye with complete retinal detachment (Shields et al., 2001a).

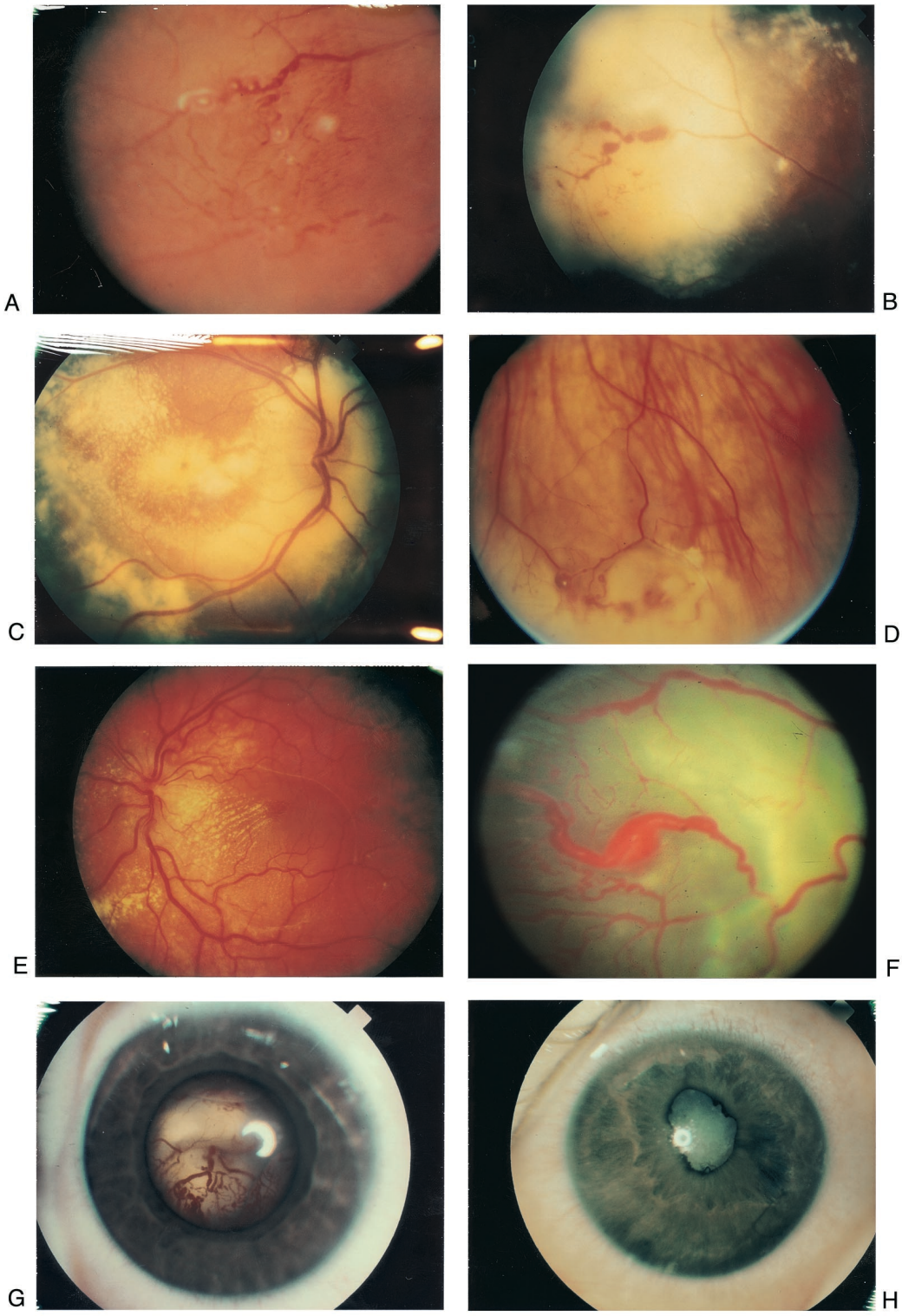


Figure 1.7 Images of all the stages as classified by Shields et al. (2001a).
(A) Retinal telangiectasia only – Stage 1. **(B)** Telangiectasia and extrafoveal exudation – Stage 2A. **(C)** Foveal exudation– Stage 2B. **(D)** Subtotal retinal detachment not involving the fovea – Stage 3A1. **(E)** Subtotal retinal detachment involving the fovea – Stage 3A2. **(F)** Total exudative retinal detachment – Stage 3B. **(G)** Total retinal detachment with secondary glaucoma –Stage 4. **(H)** Advanced end stage disease with chronic inflammation and cataract, secondary to longstanding retinal detachment – Stage 5. Images used with permission from Dr Carol Shields (Shields et al., 2001a).

This system is routinely used today to medically describe the progression of Coats disease in affected patients.

1.5.3 Treatment

Currently, there is no cure for Coats disease and all of the treatments available are in place to prevent any further vision loss. There are a number of treatment options available, depending on the severity presented by the patient. The treatment ranges from observation, photocoagulation, cryotherapy, anti-VEGF therapy and subretinal fluid drainage in combination with photocoagulation or cryotherapy. Enucleation may become the only option in severe cases that display complete retinal detachment and secondary glaucoma with a painful eye (Shields et al., 2001b).

1.5.3.1 Observation

Observation is part of the treatment plan for all patients. Patients are observed when in very early stages of the disease to allow for the most effective treatment route to be taken. Patients on the more severe spectrum of the disease are also subjected to observation only when either no other treatment options were effective, or the progression of the disease remains stable. For the extreme cases, observation is crucial in case of major decline in health of the affected eye that may result in possible enucleation.

1.5.3.2 Laser photocoagulation

This method was first described by Meyer-Schwickerath (Meyer-Schwickerath, 1959; Spitznas et al., 1976) and it still remains one of the most effective treatments for early stages of Coats disease. Photocoagulation is based on

treating the eye with a laser and the conversion of the light energy into heat or thermal energy. Once the laser is applied to live tissue, the temperature of the local area rises which in turn denatures proteins and coagulates tissue (McGrand, 1970). This method is used to cauterise the abnormal vasculature within the retina. Patients are usually checked within a couple of months of treatment to assess if more sessions of laser photocoagulation are needed to fully seize the growth of abnormal vessels (Shields et al., 2001a). The laser photocoagulation kills sections of the retina, and this prevents the hypoxic signals that would stimulate the growth of abnormal vessels. This treatment is also used in combination with other treatments, such as subretinal fluid drainage (Section 1.5.3.4).

1.5.3.3 Cryotherapy

Cryotherapy is another treatment option for Coats disease patients and it is usually used when laser photocoagulation is not successful or not possible in certain cases (Tarkkanen and Laatikainen, 1983). This method involves the freezing of the abnormal vessels to prevent more leakage and lipid deposits. The cryotherapy also kills sections of the retina, similarly to laser photocoagulation. The affected areas are treated three times in one session in a form of a freeze/thaw therapy and it standardly requires more than one session (Sigler et al., 2014). Follow up is needed to evaluate the effectiveness of the treatment and whether repeating is necessary (Shields et al., 2001a).

Photocoagulation and cryotherapy are not effective treatment options in cases of complete detachment of the retina (Haik, 1991). In these cases, the combination of various therapies is used, which are described in the sections below.

1.5.3.4 Subretinal fluid drainage and vitrectomy

In late stages of Coats disease, retinal detachment may occur due to the aggregation of exudates within the intraocular space. In these cases, subretinal fluid drainage or vitrectomy are used to prevent further retinal detachment and encourage retinal reattachment (Yoshizumi et al., 1995 & Imaizumi et al., 2016). This method of treatment can sometimes be also used along with photocoagulation and cryotherapy for optimal results.

1.5.3.5 Anti-VEGF therapy

Vascular endothelial growth factor (VEGF) is an angiogenic cytokine that was documented to be involved in leakage of retinal vasculature (Aiello et al., 1994). The use of intravitreal anti-VEGF compounds has become one of the common treatment options, either as sole treatment or in conjunction with others described earlier in this section (He et al., 2010; Lin et al., 2010; Cackett et al., 2010; Villegas et al., 2014; Park et al.,2016; Goel & Kumar, 2016).

1.5.4 Differential diagnosis

Coats disease is known to be widely misdiagnosed as it contains phenotypic features of many other eye disorders. During a 24-year study, 150 patients were reviewed and diagnosed with Coats disease by specialists Drs Shields and Shields (Shields et al., 2001b). However, the referral diagnoses for the patients from physicians ranged from Coats disease (41%), retinoblastoma (27%), retinal detachment (8%), retinal haemorrhage (4%) and various others (Shields et al., 2001b). The most common differential diagnoses of Coats disease are detailed in the following sections 1.5.4.1 - 1.5.4.5.

1.5.4.1 Retinoblastoma (MIM 180200)

Both disease spectrums can have signs of retinal detachment, defective vasculature and subretinal mass (Jaffe et al., 1977; Lam et al., 2008). The differentiation between the two entities used to be very difficult (Steidl et al., 1996) but nowadays it has become easier with the advances in molecular diagnoses. The misdiagnosis was due to the poor view of posterior segment during ophthalmoscopic examination. For correct diagnosis, axial contrast enhanced computed tomography was used. This allowed for detection of masses within the intraocular space to distinguish between retinal detachment. Incorrect diagnosis of retinoblastoma in patients with Coats disease used to result in unnecessary enucleation of a potentially treatable eye (Jaffe et al., 1977). On the other hand, misdiagnosis of retinoblastoma for Coats disease could potentially have had severe effects on patients' health due to the possibility of tumour metastasis. OCT is more frequently used in retinoblastoma detection than MRI due to its ability to detect calcification (Mafee et al., 1989).

1.5.4.2 Retinitis pigmentosa (RP) (MIM 268000)

An RP phenotype can be simulated in Coats-like retinopathy (Khan et al., 1988). However, RP presents bilaterally, and the underlying genetic basis are not related to Coats disease. In a study by den Hollander et al. (2004), mutations in *CRB1* have never been identified in patients diagnosed with RP with Coats-like retinopathy.

1.5.4.3 Norrie disease (MIM 310600)

Norrie disease is an X-linked recessive disorder which causes congenital blindness, deafness and, sometimes, mental deficiency (Norrie, 1927; Warburg, 1961; Warburg, 1966). It is a severe form of FEVR-related retinopathy with congenital retinal detachment/dysplasia. Norrin, the protein product of *NDP* (norrie disease pseudoglioma), is defective in Norrie disease patients. It can be distinguished from Coats disease by the bilateral presentation and family history of the disease.

1.5.4.4 Familial exudative vitreoretinopathy (FEVR) (MIM 133780)

FEVR is a hereditary disorder that shares many phenotypic similarities to Coats disease, such as abnormal vasculature, exudation and retinal detachment (Section 1.8) (Salvo et al., 2015; Toomes et al., 2005; Zhang et al., 2011). FEVR can be inherited in an autosomal dominant (ad), autosomal recessive (ar) or X-linked inheritance pattern. It presents bilaterally and occurs in both males and females. Patients with radical FEVR can be blind from birth, while patients with milder forms of the disease may not have any noticeable symptoms (Ober et al., 1980).

1.5.4.5 Macular telangiectasia (MIM 187300)

There are three types of macular telangiectasia. Only type 1 is often confused for Coats disease as it is characterised by a phenotype of abnormal vasculature, so called telangiectasia and macular edema, that is only localised around the fovea. It was first described in 1982 as an idiopathic condition

(Gass & Blodi, 1993). It also presents unilaterally and in young males (Machkour et al., 2017).

1.6 Norrin-β-catenin signalling pathway

The Norrin-β-catenin signalling pathway has been defined over a number of years by identifying and studying the genes and proteins involved in FEVR-related retinopathies (Junge et al., 2009; Xu et al., 2004; Ye et al., 2010). The target genes of this pathway are involved in correct vascularisation of the retina (Section 1.4.2). The pathway itself is similar to the Wnt-β-catenin signalling pathway but there are a number of key differences (Clevers & Nusse 2012). One of the crucial changes is the change of the Wnt ligand for Norrin (Xu et al., 2004). The second change, in order to transduce the signal, the ligand acts only through a specific frizzled receptor, FZD4, whereas the Wnt-β-catenin pathway uses any of the ten frizzled receptors (Smallwood et al., 2007). Finally, the pathway is enhanced by the auxiliary protein TSPAN12 which is only involved in activation of the Norrin-β-catenin pathway (Ye at al. 2009).

1.6.1 Activation of the Norrin-β-catenin signalling pathway

The pathway is activated when the ligand, Norrin, binds to the FZD4-LRP5-TSPAN12 receptor complex at the cell membrane (Figure 1.9). Norrin binds to the extracellular CRD of FZD4 (Chang et al., 2015). The co-receptor of LRP5 is also essential for the signalling to occur. There is no evidence of binding of Norrin to LRP5 during the signal induction but studies have elucidated potential binding sites between these two proteins (Ke et al., 2013; Chang et al., 2015). This binding initiates a signal that is transduced inside the cell and prevents the degradation of the cytoplasmic β-catenin. The signal recruits the

destruction complex which binds the C-terminal of LRP5 through Axin together with glycogen synthase kinase 3 beta (GSK3β). The destruction complex is composed of Axin and adenomatous polyposis coli (APC) proteins together with GSK3β and casein kinase 1 (CK1) which facilitate the phosphorylation of β-catenin (Stamos and Weis, 2013). DVL is also activated and binds to FZD4 inside the cytoplasm (Mao et al., 2001; Punchihewa et al., 2009). Axin is subsequently degraded and the inhibition of GSK3β reduces the phosphorylation of β-catenin. The relocation of the degradation complex further stabilises β-catenin which allows it to accumulate within the cytoplasm. β-catenin then translocates into the nucleus where it binds to the T-cell factor (TCF)/Lymphoid enhancing factor (LEF) transcription factors and promotes the activation of Norrin target genes (Ye et al., 2010).

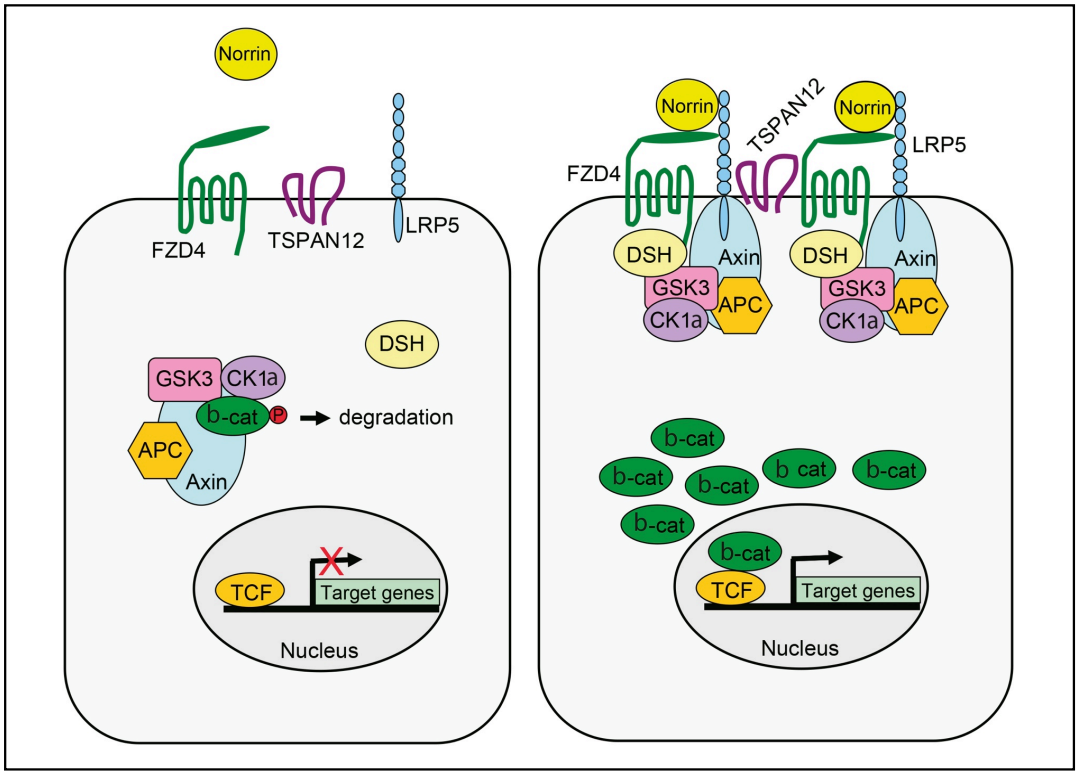


Figure 1.8 Schematic of the Norrin-β-catenin signalling pathway.
Image on the left-hand side shows the absence of Norrin and β-catenin undergoing degradation and expression of target genes is inhibited. The image on the right-hand side

shows Norrin binding to the receptor complex which prevents the degradation of β-catenin. This allows for β-catenin's accumulation and translocation into the nucleus where it activates the transcription of target genes. Schematic provided by Dr Carmel Toomes (University of Leeds).

1.6.2 Degradation of β-catenin in the absence of the ligand Norrin

In the case of Norrin being absent and no ligand binding occurs, the destruction complex does not relocate to the cytoplasmic C-terminal of LRP5 and β-catenin is phosphorylated and subsequently removed by ubiquitin-mediated protein degradation (Ye et al., 2010). Within the nucleus, the target genes remain unexpressed by interaction with the TCF/LEF factors and co-repressors. Therefore, in the off state of β-catenin signalling, β-catenin levels and pathway activation remains low (Figure 1.9,left-hand side image).

1.6.3 The role of Norrin-β-catenin pathway in vascularisation

The Norrin-β-catenin signalling pathway has been documented to control the development of the retinal vasculature in a number of studies (Junge et al., 2009; Kato et al., 2002; Rehm et al., 2002; Xu et al., 2004; Qin et al., 2007; Ye et al., 2009). Norrin-β-catenin signalling is responsible for proliferation, migration and invasion of endothelial cells in retinal vascular development, making it essential for proper angiogenesis that is also required for blood-brain barrier (BBB) and CNS development (Wang et al., 2012). The NDP/FZD4 signalling expression has been predominantly documented in vascular endothelial cells where it is of high importance (Ye et al., 2009; Zhang et al.,

2017). In the same study, it was also documented that KO of *Fzd4* in all endothelial cells disrupted the blood brain barrier (Ye et al., 2009).

To further establish the role of the pathway in retinal vascularisation, published mutations were modelled to establish their effects on the activation of the Norrin-β-catenin signalling pathway. In a study by Qin et al. (2007) it was found that nonsense mutation in the receptor *FZD4* completely abolished the signalling activity and missense mutations in *LRP5* and *FZD4* showed less of an effect but still decreased the signalling by 36% on average. Double missense mutations had a more severe effect on the reduction of signalling, 71%. Tspan12 siRNA was shown to abolish Norrin signalling in retinal endothelial cells (Junge et al., 2009). Mutations in *CTNNB1* were only recently published to have an effect on Norrin signalling (Panagiotou et al., 2017). In this study, a nonsense mutation showed a significant decrease in signalling, which is in agreement with mutations in the other components of the Norrin-β-catenin signalling pathway. However, a missense change showed significantly increased levels of Norrin signalling, which is not usually linked to defects in vasculature. This phenomenon was also documented in another study, where an increase in Norrin signalling was shown to disrupt angiogenesis (Ye et al., 2009).

1.7 Genetics of Coats disease

Coats disease is an idiopathic condition which has not been explored extensively at the genetic level. The first study that proposed a genetic cause of this disease was a study by Graeme Black and colleagues (1999) who

identified somatic mutations in the *NDP* (MIM 300658) gene. *NDP* is very well characterised and known to be mutated in Norrie disease and FEVR-related retinopathy (Chen et al., 1993a). It encodes the ligand, Norrin, that is utilised in the Norrin-β-catenin signalling pathway. In the breakthrough study by Black et al. (1999), a mother with unilateral Coats disease gave birth to a son affected with Norrie disease (Section 1.5.4.3). Mother and son were both found to carry the same missense mutation in *NDP* which was located in exon 3, c.288C>G, p.(Cys96Trp). Based on this finding, the authors hypothesized that Coats disease may be a somatic form of *NDP*-related disease. They went on to screen *NDP* in the DNA of nine enucleated Coats eyes from unrelated males presenting with unilateral, classical Coats disease. The screening of this somatic tissue identified the same missense mutation identified in the original mother and child. They showed that the mutation was somatic by confirming its absence in non-retinal tissue. This mutation is not present in an online variant database gnomAD (Section 2.19.7) and it occurs within a highly conserved region.

Recently, a mutation in a new gene called RCC1 and BTB domain containing protein 1 (*RCBTB1*) was identified in a single Coats disease patient. In this study, a heterozygous single nucleotide deletion c.707delA, p.(Asn236ThrfsTer11) was identified in germline DNA (gDNA) in a male patient diagnosed with Coats disease in his right eye by WES. The left eye was healthy. The father also carried the same nucleotide deletion but was unaffected (Wu et al., 2016). In the same study, *RCBTB1* was speculated to

have a negative effect on the Norrin-β-catenin signalling pathway by reducing the nuclear accumulation of β-catenin.

No further Coats disease cases have been identified with *NDP*, *RCBTB1* or any other gene mutations in the literature to date. Studies that have screened gDNA or somatic tissue in Coats disease patients included the study by Black et al. (1999) where gDNA of two patients (one affected with Coats disease) was used to identify the *NDP* mutation. This study also screened the retina of nine enucleated eyes and identified the same mutation. In the *RCBTB1* study by Wu et al. (2016) only gDNA was screened in two patients (one affected with Coats disease) who presented with the described mutation. In another study, *FZD4* was screened in gDNA of a cohort of 16 Coats disease patients but no pathogenic mutations were identified (Robitaille et al., 2011).

1.8 Familial exudative vitreoretinopathy (FEVR)

FEVR (MIM 133780) is a hereditary disorder that is characterized by the incomplete vascularization of the peripheral retina. The disease was first described by Criswick & Schepens (1969). It was defined as an inherited retinal disease that shares many clinical features with retinopathy of prematurity (ROP) (Toomes and Downey, 2005). FEVR is known to be caused by mutations in the components of the Norrin-β-catenin signalling pathway (Section 1.6) (Chen et al., 1993a; Robitaille et al., 2002; Jiao et al., 2004; Toomes et al., 2004b; Junge et al., 2009; Poulter et al., 2010; Dixon et al., 2016;Panagiotou et al., 2017).

1.8.1 Phenotypic features of FEVR

The primary feature of this disorder is the premature arrest of retinal vasculature development that leads to incomplete vascularisation of the peripheral retina (van Nouhuys, 1991). Lack of vascularization in the retina can lead to retinal ischemia, which can cause adverse changes that result in vision loss such as retinal exudation, neovascularization, retinal folds and detachment (Figure 1.9) (Toomes & Downey, 2005, last revision 2011).

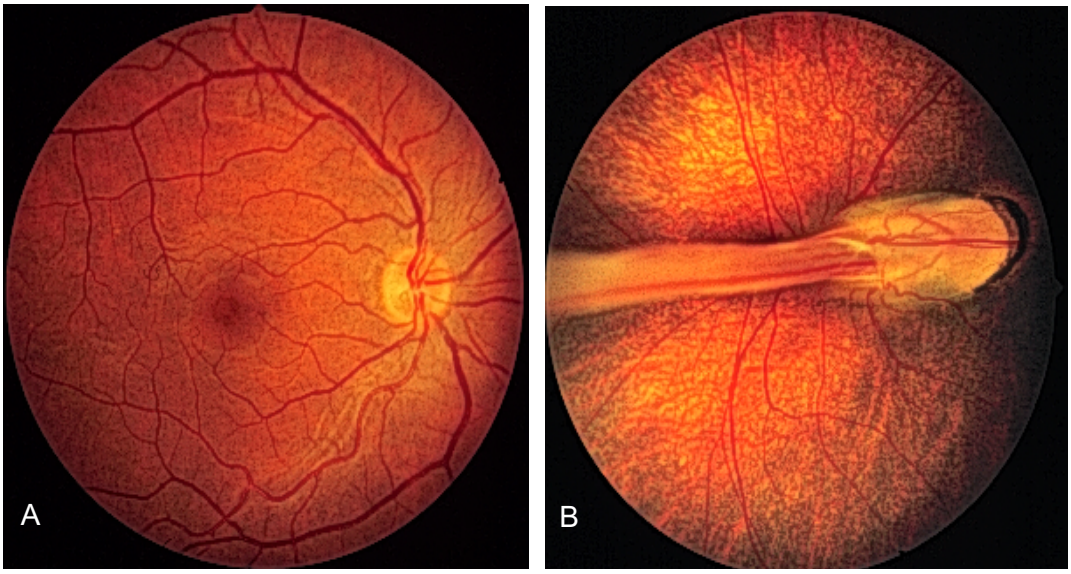


Figure 1.9 Clinical feature of FEVR.
(A) The fundus image shows a healthy human eye with healthy vasculature. **(B)** The fundus image is from a child with FEVR and shows a retinal fold crossing the macular area and dragging of the optic disc. Both images were provided by Dr Carmel Toomes.

The symptoms of the disease can be extremely variable, even within the same family (Toomes et al., 2004a). Patients with severe FEVR can be blind from birth, while patients with milder forms of the disease may not have any noticeable symptoms throughout life (Ober et al., 1980). Some relatives can be asymptomatic while others can display more severe complications such as retinal detachments (Robitaille et al., 2009). Although FEVR is typically

reported as a bilateral condition, previous studies have also documented patients presenting with a high degree of asymmetry between their eyes or even unilateral disease (van Nouhuys, 1991; Ranchod et al., 2011).

To detect and diagnose FEVR, a basic fundus image can be carried out. Milder forms can also be detected by the use of intravenous fluorescein angiography (IVFA) as it is a very sensitive method for detection of vascular defects even in patients with a very mild phenotype (Canny and Oliver, 1976). The more severe forms are detected by sonography, OCT and MRI.

1.8.2 Genetics of FEVR

FEVR can be inherited in an autosomal dominant (adFEVR, MIM 133780), autosomal recessive (arFEVR, MIM 601813) or X-linked (MIM 305390) mode of inheritance. adFEVR is the most common form (Kondo et al., 2001). A number of genes identified to cause this disorder form the components of the established Norrin-β-catenin signalling pathway (Section 1.6).

To date, there have been nine genes reported to cause FEVR when mutated. These encode Norrin (*NDP*, MIM 300658) (Chen et al., 1993a); frizzled receptor 4 (*FZD4*, MIM 604579) (Robitaille et al., 2002); low-density lipoprotein receptor-related protein 5 (*LRP5,* MIM 603506) (Jiao et al., 2004; Toomes et al., 2004a); tetraspanin-12 (*TSPAN12*, MIM 613138) (Nikopoulos et al., 2010; Poulter et al., 2010); atonal homolog 7 (*ATOH7*, MIM 609875) (Khan et al., 2012); zinc finger protein 408 (*ZNF408*, MIM 616454) (Collin et al., 2013); kinesin family member 11 (*KIF11*, MIM 148760) (Robitaille et al., 2014); RCC1 and BTB domain containing protein 1 (*RCBTB1,* MIM 607867) (Wu et al., 2016) and beta-1 catenin (*CTNNB1*, MIM 116806) (Dixon et al., 2016; Panagiotou et al., 2017).

The following genes: *FZD4*, *LRP5*, *TSPAN12, ZNF408, KIF11, RCBTB1 and CTNNB1* when mutated cause adFEVR. *FZD4*, *LRP5*, *TSPAN12* and *ATOH7* when mutated can cause arFEVR. Mutations in *NDP* gene cause X-linked FEVR.

All the genes involved in the pathogenesis of FEVR are documented in detail in the following sections and summarised in Table 1.1.

Gene ID	Inheritance pattern
Frizzled-4 (*FZD4*)	Dominant & Recessive
Norrin (*NDP*)	X-linked
Low density lipoprotein receptor related protein 5 (*LRP5*)	Dominant & Recessive
Tetraspanin 12 (*TSPAN12*)	Dominant & Recessive
Beta-1 catenin (*CTNNB1*)	Dominant
Atonal basic helix–loop–helix (bHLH) transcription factor 7 (*ATOH7*)	Recessive
Zinc finger protein 408 (*ZNF408*)	Dominant
Kinesin family member 11 (*KIF11*)	Dominant
RCC1 and BTB domain containing protein 1 (*RCBTB1*)	Dominant

Table 1.1 Summary of all genes involved in pathogenesis of FEVR. Names, gene symbols and inheritance patterns of all genes mutated in FEVR to date.

1.8.2.1 Frizzled-4 (*FZD4*)

FZD4 (MIM 604579) was first identified to play a role in retinal angiogenesis and to harbour mutations causing adFEVR in 2002 (Robitaille et al., 2002). FZD4 acts as a receptor in both the Wnt-β-catenin and Norrin-β-catenin signalling pathways (Xu et al., 2004). The N-terminus of FZD4 is extracellular and contains the cysteine-rich domain (CRD) which has been shown to bind to Norrin and Wnt to initiate the signalling pathway (Smallwood et al., 2007). The C-terminal of FZD4 is cytoplasmic and contains the Dishevelled (DVL) association domain and also the PDZ binding domain. In the cytoplasm, DVL binds to FZD4 when Norrin signal activates the β-catenin signalling. This causes the destruction complex to relocate and bind to cytoplasmic LRP5 and FZD4 (Punchihewa et al., 2009). PDZ domain was also shown to be crucial for β-catenin signalling (Bertalovitz et al., 2016).

Heterozygous *Fzd4*$^{+/-}$ mice don't have any vascular defects as seen in humans with heterozygous *FZD4* mutations. A Fzd4 knockout (KO) mouse model showed that the development of and spread of the primary vascular plexus across the vitreous was diminished by P5-P10 and the intraretinal deeper plexuses failed to develop. Hyaloid vasculature regression was also delayed (Xu et al., 2004).

1.8.2.2 Norrin (*NDP*)

Mutations in *NDP* (MIM 300658) are known to cause Norrie disease (MIM 310600, Section 1.5.4.3) and X-linked FEVR (Humbert et al., 1993). Norrin, the protein encoded by *NDP*, shares homology with the CRD of mucins and it is structurally similar to the transforming growth factor β (TGFβ) (Meindl et al.,

1992; Meitinger et al., 1993). The mouse and human protein share 94% homology.

To initially show that Norrin plays a role in the development of the retinal vasculature, a knock-out C57Bl/6 Ndp mouse model was generated and it was shown that the vasculature was not established into its defined retinal layers (Richter et al., 1998). The intraretinal capillary plexuses spanning the INL layer were shown to be absent. The GCL was shown to have an increased number of vessels and there was also penetration of vessels into the vitreous through the inner limiting membrane (ILM). The hyaloid vasculature also failed to fully regress (Richter et al., 1998; Berger et al., 1996). The ocular findings in the hemizygous Ndp mice replicated the phenotypes observed in Norrie disease human patients (Berger et al., 1996). In another study, Norrin was found to be essential for proper vasculature development in the eye, brain and ear (Rehm et al., 2002). In 2004, it was identified as a ligand for the FZD4 receptor in the Norrin-β-catenin signalling pathway (Xu et al., 2004). Recently, Norrin was found to cause FZD4 ubiquitination leading to receptor complex internalisation (Zhang et al., 2017). Norrin was confirmed to bind to the CRD domain of FZD4 specifically, and not to any of the other frizzled family receptors, to initiate signalling. This study determined the binding between FZD4 and Norrin, but no binding was shown between Norrin and LRP5/6 (Xu et al., 2004).

1.8.2.3 Low density lipoprotein receptor related protein 5 (*LRP5*)

LRP5 (MIM 603506) was first identified by DNA sequencing on chromosome 11q13 (Hey et al., 1998). *LRP5* was identified by Toomes et al. (2004b) as one

of the FEVR genes in which mutations in *LRP5* that cause adFEVR were documented. In a separate study, homozygous mutations in *LRP5* were found in three consanguineous families (Jiao et al. 2004), identifying *LRP5* as the first gene causing arFEVR.

Homozygous mutations in *LRP5* have been found in patients with low bone mass as a cause of autosomal recessive osteoporosis pseudoglioma (OPPG, MIM 259770) (Gong et al., 2001). In a study by Downey et al. (2006), a consanguineous recessive family with FEVR was found to have homozygous LRP5 p.Gly505Arg mutation and subsequently assessed to show that affected patients also had low bone mass. Heterozygous carriers were shown to also have reduced bone mass (Gong et al., 2001). Additional studies have since identified further *LRP5* mutations for both recessive and dominant forms of FEVR (Qin et al., 2005; Yang et al., 2012; Rao et al., 2017).

Studies have elucidated the role of LRP5/6 in Wnt signalling by determining that LRP5/6 comes into contact with frizzled receptors, which interact and bind to Wnt proteins, and in turn form the ligand receptor complex that activates signalling (Pinson et al., 2000; Tamai et al., 2000; Wehrli et al., 2000). In 2001, a study showed that LRP5 was able to mediate Wnt signalling *in vitro* and interact with Axin, which is part of the inhibitory complex for degradation of β-catenin (Section 1.6.2) (Mao et al., 2001). In a further study a number of years later, LRP5 was found to serve as a co-receptor in the Norrin-β-catenin signalling pathway involved in proper vasculature within the retina (Xu et al., 2004). Eventually, it was determined that the addition of Lrp5/6 further enhanced the signalling activity when WT Norrin was bound to Fzd4 (Ke et al.,

2013).

Over the years, there have been a number of Lrp5 KO mouse models described. In 2002, a mouse model deficient of Lrp5 showed the delay in regression of the hyaloid vasculature (Kato et al., 2002). A recessive frameshift mutation that causes a premature stop codon was introduced to create a mouse model by Xia et al. (2008). This frameshift mutation impaired the recruiting of the destruction complex and, therefore, prevented β-catenin from degradation and activation of the signalling pathway (He et al., 2004). The phenotype of the mutant mouse was similar to that of FEVR in humans. In a heterozygous *Lrp5*$^{+/-}$ mice, endothelial cells were shown to form clusters in the IPL instead of continuing on to form the intraretinal plexuses (Xia et al., 2010). Another study by Xia et al. (2010) developed an *Lrp5*$^{-/-}$ knockin LacZ mouse model to establish the role of LRP5 in vascular development. It was concluded that Lrp5 expression is in Müller cells and vascular endothelial cells and in *Lrp5*$^{-/-}$ mice they form cell clusters in the IPL and GCL. This indicated that the endothelial cells do not sprout and migrate to form plexuses. However, in the *Lrp5*$^{+/-}$ mouse, the vascular phenotype does not change and the plexuses remain organised (Xia et al., 2010).

One mouse model generated focused on a double knock-out (DKO) of *Lrp5* and *Vldlr*. A single homozygous knock out of Lrp5 only showed the mice lacking proper vasculature and Lrp5 was not shown to be expressed in deeper layers of the retina. This showed that the endothelial cells forming the vessels have failed to form and migrate through the layers. This confirmed that LRP5 is required for proper angiogenesis within the retina (Xia et al., 2013).

1.8.2.4 Tetraspanin 12 (*TSPAN12*)

Dominant mutations in *TSPAN12* (MIM 613138) were identified in FEVR patients in a number of studies (Nikopoulos et al., 2010; Poulter et al., 2010; Kondo et al., 2011). In another study by the same lab, it was suggested that the FEVR phenotype can depend on the dosage of TSPAN12. Patients with homozygous or compound heterozygous mutations (arFEVR) in *TSPAN12* showed a more severe phenotype in comparison to patients with only one heterozygous mutation (Poulter, et al., 2012).

KO mouse studies of Tspan12 also showed failure of hyaloid vessel regression and loss of the intraretinal plexuses, an identical phenotype to KO mouse studies of Fzd4, Ndp and Lrp5 (Junge et al., 2009). In the same study, it was also found that TSPAN12 was interacting with the Norrin-β-catenin receptor complex by increasing its activation when introduced into endothelial cells. It was concluded that TSPAN12 promotes multimerization of FZD4 together with Norrin (Junge et al., 2009). In a study by Lai et al. (2017), it was shown, through cell-based assays and *in vivo* experiments, that TSPAN12 stabilised the receptor-ligand complex of FZD4 and Norrin.

1.8.2.5 Beta-1 catenin (*CTNNB1*)

CTNNB1 is involved in binding of cadherins and plays an important role in Wnt signalling as the nuclear effector (Valenta et al., 2012). Mutations in *CTNNB1* (MIM 116806) are documented to be associated with many cancers (Anastas and Moon, 2013). Haploinsufficiency of *CTNNB1* also causes syndromic intellectual disability that is accompanied by microcephaly, brain abnormalities, abnormal muscle tone, speech impairment and distinctive facial features (de Ligt et al., 2012; Dubruc et al., 2014; Kuechler et al., 2014).

CTNNB1 was found to be expressed in most retinal layers in mouse retinal development and in mature retina (Fu et al., 2006). In the same study, it was shown that *CTNNB1* null retinas affected the arrangement of retinal progenitor cells and also the laminar structure. Other KO mice were documented to lack proper vasculature within the CNS and also in the retina (Daneman et al., 2009; Zhou et al., 2014). *CTNNB1* was only recently described to harbour mutations in FEVR patients (Dixon et al., 2016; Panagiotou et al., 2017). Panagiotou et al. (2017) observed a clustering of heterozygous mutations at the C-terminus of *CTNNB1* in adFEVR families and a sporadic case and it is the C-terminal mutations which appear to cause isolated FEVR.

The genes described have all been found to be involved in the Norrin-β-catenin signalling pathway. However, additional genes have been found to be mutated in FEVR cases but have not yet been shown to be implicated in Norrin-β-catenin signalling. These genes are described in sections 1.8.2.6 - 1.8.2.9 below.

1.8.2.6 Atonal basic helix–loop–helix (bHLH) transcription factor 7 (*ATOH7*)

ATOH7 (MIM 609875) is responsible for RGC fate specification (Prasov et al., 2012) and *Atoh7*$^{-/-}$ mice show a decrease in the number of RGCs and the lack of the optic nerve (Brown et al., 2001). There is a close link between the RGCs and the retinal vasculature. Defects in RGC development often lead to retinal vasculature defects and a failure of hyaloid vasculature regression, meaning it remains within the retina (Brzezinski et al., 2003). Homozygous mutations were identified in *ATOH7* in patients suffering from ocular symptoms that show

similar defects to those seen in FEVR, such as abnormal vasculature and persistence of hyaloid vasculature (Khan et al., 2012). Mutations in *ATOH7* have only been found in severe cases of FEVR and their phenotype was shown to be more similar to retinal dysplasia as seen in Norrie disease and arFEVR caused by mutations in *LRP5* and *TSPAN12*.

1.8.2.7 Zinc finger protein 408 (*ZNF408*)

ZNF408 (MIM 616454) encodes the zinc finger transcription factor and dominant mutations in this gene have been associated with adFEVR. This is a rare form of FEVR with only small number of mutations reported to date. In a study of Dutch FEVR families, missense mutations were identified in *ZNF408*. One of the missense mutations p.His455Tyr was shown to act as dominant negative mutation. Additionally, the knockdown of *znf408* in zebrafish demonstrated abnormal retinal vasculature and trunk vasculature, which was shown to be rescued by the WT ZNF408 RNA (Collin et al., 2013). *ZNF408* was also shown to cause autosomal recessive retinal dystrophies (such as RP) (Coppieters et al., 2016).

1.8.2.8 Kinesin family member 11 (*KIF11*)

Dominant mutations in *KIF11* (MIM 148760) were shown to cause FEVR (Robitaille et al., 2014), microcephaly, lymphedema and chorioretinal dysplasia (MLCRD, MIM 152950), chorioretinal dysplasia, microcephaly and mental retardation (CDMMR) and microcephaly with or without chorioretinopathy, lymphedema, or mental retardation syndrome (MCLMR) (Hazan et al., 2012; Mears et al., 2015; Mirzaa et al., 2014; Jones et al., 2014; Ostergaard et al., 2012; Robitaille et al., 2014; Schlögel et al., 2015). Robitaille

(2014) showed that patients with a diagnosis of FEVR had *KIF11* mutations but retrospective analysis showed that some of these patients had also mild microcephaly that had been missed on initial consultation. No other features of MLCRD and CDMMR were noted (Robitaille et al., 2014).

1.8.2.9 RCC1 and BTB domain containing protein 1 (*RCBTB1*)

In 2016, a gene called *RCBTB1* (MIM 607867) was identified to carry heterozygous mutations in Taiwanese FEVR patients. To assess for the involvement of the gene in angiogenesis and retinal vasculature, a knockdown of *RCBTB1* by shRNA registered reduced β-catenin levels when using both Norrin and Wnt3a as ligands. The study also generated a knockdown of *rcbtb1* in zebrafish and documented abnormal intraocular vessels (Wu et al., 2016). A canonical splice site variant in *RCBTB1* was identified in two cousins diagnosed with FEVR and the variant was also present in their asymptomatic mothers. *RCBTB1* was also shown to cause autosomal recessive retinal dystrophies (such as RP), similar to *ZNF408* (Coppieters et al., 2016).

Only around 50% of FEVR cases are documented to carry mutations in any of the known FEVR genes mentioned above (Salvo et al., 2015; Seo et al., 2015; Kashani et al., 2014). Thus, there is a need for identification of novel genes that may be implicated in this disorder.

1.9 Formalin-fixed, paraffin-embedded (FFPE) samples

Formalin-fixation and paraffin-embedding of tissues has been widely used for many years in order to preserve precious tissues to create historical archives and to facilitate analysis of many cancers or other diseases. Initially, fixation of

samples and tissues was undertaken to preserve irreplaceable populations and species. It's only been in the last 40 years that the preservation of samples for DNA related purposes has become routine (Schander et al., 2003).

1.9.1 Formaldehyde as the first fixative

Formaldehyde was created in 1859 by Butlerov and was the first known fixative (Schander & Halanych, 2003). Nine years later, von Hofmann also created and identified formaldehyde. This substance was then commercially produced from 1901 (Walker, 1964). The use of formaldehyde as a fixative for tissues was assessed by French & Edsall (1945). The formaldehyde quickly became the main component of preserving biological samples (Jones, 1976; Fox et al., 1985;Johnson et al., 1995). Formaldehyde is gas but for fixing tissues it is mixed with water at the concentration of 37%.

The use of formaldehyde to preserve biological and medical specimens has allowed for the ability to continue on with research for many diseases which require biological samples. Histological fixation can aid in long term storage of specimens, thus creating a vast archive that can be used for many forthcoming years.

1.9.2 Formalin fixation

Formalin has fast become the most frequently used reagent for fixing tissues. Formalin is a saturated formaldehyde aqueous solution with the addition of a stabiliser, methyl alcohol (methanol). The addition of the stabiliser prevents the polymerisation of formaldehyde. Without it, the solution is unstable, and the polymerisation can produce molecules that are insoluble, forming precipitation.

10% buffered formalin is standardly used as a fixative and this solution contains 4% formaldehyde (Ibeh, 1998). Alcohols such as acetone were documented to have replaced formaldehyde in the past (Fukatsu, 1999). However, acetone is not in frequent use today as a fixative.

Besides using formalin as a fixative, fresh-frozen (FF) samples are another method of preserving tissues. Hedegaard et al. (2014) compared gene expression levels between tissue preserved by either FF or FFPE. DNA and RNA were isolated from FFPE and FF samples and exome and transcriptome analysis performed. The exome analysis showed that 70-80% of variants were present in both FF and FFPE samples that were preserved for more than 3 years. The transcriptome analysis found that 1,494 genes varied between paired FF and FFPE samples. This study also showed the longer the samples were stored, the lower the yield of DNA and RNA. Overall, the study concluded that FFPE samples are a viable source of DNA that can be analysed by NGS technologies, however, caution must be taken when interpreting the findings due to the damage caused by the fixation process. This topic is discussed in further detail in Chapter 3.

1.10 Aims

The primary aim of this thesis was to identify and characterise genes and proteins involved in Coats disease. The use of WES was undertaken to screen a cohort of FFPE Coats disease eye globes to identify causative mutations and genes that could play a role in this disease.

The second aim of this thesis was to identify mutations in known FEVR genes by WES in a cohort of 30 FEVR patients, and to identify new disease genes

for FEVR that may also be involved in the Norrin-β-catenin signalling pathway. If a new FEVR gene was identified, this would also be a good candidate gene for Coats disease.

2 Chapter: Materials & Methods

2.1 General buffers

1x Tris-Ethylenediaminetetraacetic acid EDTA (TE) buffer (pH 8.0)

10mM	Tris (pH 8.0)
1mM	EDTA

10x Tris-Borate-EDTA (TBE) buffer, pH 8.0

0.89M	Tris
0.89M	Boric Acid (H_3BO_3)
25mM	EDTA (pH 8.0)

10x Gel loading dye

3x	TAE
20%	Ficoll 400
0.1%	Bromophenol Blue
0.2%	Xylene Cyanol

Luria-Bertani (LB) broth

1%	Tryptone
0.5%	Yeast Extract
1%	Sodium Chloride (NaCl)

For plates, 1.5% agar was added.

Kanamycin final concentration of 25μg/ml and ampicillin final concentration of 50μg/ml.

NZY+ Broth

10g	NZ amine (casein hydrolysate)
5g	Yeast extract
5g	NaCl

Add dH_2O to a final volume of 1 litre and adjust pH to 7.5. Autoclave and add the following prior to use:

12.5ml	1M $MgCl_2$
12.5ml	1M $MgSO_4$
20ml	20% w/v glucose

1x TAE electrophoresis buffer

40mM	Tris
20mM	Acetic acid (CH3COOH)
1mM	EDTA

Elution Buffer (EB)

10mM	Tris-HCl (pH 8.0)

1x PBS Tween (PBS-T)

1x	PBS (Phosphate Buffered Saline)
0.05%	Tween 20

2.2 Patients and eye samples

2.2.1 Coats disease eye globes

The enucleated human Coats eye samples were recruited and consented by collaborators Dr J. A. Shields, Dr C. L. Shields and Dr A.V. Levine (Wills Eye Hospital, Philadelphia, USA). This work was done under the approval of the Wills Eye Hospital Institutional Review Board (IRB #10-024 The Wnt signalling pathway and genetics of Coats disease). Eight half eye globes, which were FFPE preserved, were received from seven unrelated patients. Two half eye globes belonged to the same patient. Sample identifiers were as follows: 20147, 48753, 31276, 67580, 63170-1, 67662ON, 71208M1 and 71208M2.

2.2.2 FEVR patients

The FEVR patients (and their family members) were diagnosed and recruited by local, national and international ophthalmologists, clinical geneticists and genetic nurses. Consent was obtained from all subjects tested, and the research adhered to the tenets of the Declaration of Helsinki. Ethics approval was obtained from the Leeds East Teaching Hospitals NHS Trust Research Ethics Committee ("Investigating the molecular basis of human inherited retinal degeneration" Project numbers 03/362 and 17/YH/003).

2.3 Sectioning and deparaffinising of eye globes

Blocks of embedded eye globe were chilled overnight at 4°C. Prior to sectioning, they were placed on ice with dH_2O to soften the tissue. From each block, 50 5μm sections were cut using a microtome (Leica RM2255). Sections were placed in a 45°C water bath that had been pre-cleaned with DNA*Zap*™ (ThermoFisher Scientific) to prevent any cross contamination. Sections were picked up with slides and dried for 15 minutes on a 70°C hot plate. Sections were subsequently deparaffinised by placing them in xylene for 4 x 5 minutes followed by 100% ethanol for 3 x 5 minutes. Sections were then placed into a series of graded ethanol dilutions (100%, 75%, 50%, 25%) for 5 minutes each before placing in dH_2O prior to macrodissection.

2.4 Hematoxylin and Eosin (H&E) staining

After the sectioning and deparaffinisation of the sections in Section 2.3, slides were placed into Hematoxylin stain (acidic) for 2 minutes and washed with H_2O to remove any unbound stain. Slides were then moved to Scotts H_2O (alkaline) for further 2 minutes. After this wash, slides were placed into Eosin stain for 2

minutes and washed with 100% ethanol 4x for 2 minutes each. Finally, slides were placed in xylene 3x for 2 minutes each to fix the stain. Once finished, coverslips are mounted on top of the stained tissue section and dried overnight at room temperature (RT).

2.5 Isolation of DNA from FFPE eye globes using the QIAamp DNA FFPE Tissue Kit (Qiagen)

Eye globes were sectioned and mounted onto slides as described in Section 2.3. DNA was extracted using the QIAamp DNA FFPE Tissue Kit (Qiagen), using a modified version of the manufacturer's instructions. During dissection, slides were kept hydrated to prevent tissue flaking. Tissue was scrapped from the slide, using a 2mm scalpel, into a 1.5ml eppendorf containing 180μl of ATE buffer and 20μl of proteinase-K. A total of 12 sections were processed per eppendorf, with the exception of samples 67580 and 67662ON, where only 4 sections were processed per eppendorf (to prevent clogging of the spin columns). Samples were then incubated at 56°C in a water bath for 72 hours, with an additional 20μl of proteinase-K added to each sample after the first 24 hours. The remaining protocol steps were followed exactly from manufacturer's instructions (June 2012 edition). Briefly, following the incubation, the samples were pulse centrifuged and 200μl of Buffer AL (lysis buffer) was added. After mixing by vortexing, 200μl of 100% ethanol was added, and the samples were vortexed again and pulse centrifuged. The entire lysate from a single eppendorf was then transferred to a QIAamp MinElute column (in a 2ml collection tube) and centrifuged at 6000 x *g* for 1 minute. Each column was then placed in a clean 2ml collection tube and 500μl of Buffer AW1

(wash buffer 1) was added. After centrifuging at 6000 x *g* for 1 minute, the columns were placed in a clean 2ml collection tube and 500μl of Buffer AW2 (wash buffer 2) was added to each column before spinning again at 6000 x *g* for 1 minute. The columns were then placed in new 2ml collection tubes and centrifuged at 20,000 x *g* for 3 minutes. The QIAamp MinElute columns were placed in 1.5ml microcentrifuge tubes and 50μl of Buffer ATE (elution buffer) was added to the centre of each column membrane. Samples were subsequently incubated at RT for 5 minutes, which generally increased DNA yield, and centrifuged at full speed 20,000 x *g* for 1 minute. Eluted DNA was stored at -20°C.

2.6 Qubit Fluorometer

The concentration of DNA was measured using the Qubit™ dsDNA BR and HS Assay Kit (Invitrogen) on a Qubit™ fluorometer (Invitrogen) according to the manufacturer's instructions. All reagents were allowed to reach RT prior to the assay. A master mix containing 199μl iQuant™ Buffer BR and 1μl of iQuant™ Reagent per sample and standard was made. From the master mix, 190μl was aliquoted into two Qubit™ Assay tubes. 10μl of Qubit™ dsDNA BR Standard #1 (0ng/μl) and Qubit™ dsDNA BR Standard #2 (100ng/μl) (Invitrogen) were added to each appropriate assay tube. For each sample, 180-199μl of iQuant™ master mix solution was aliquoted into a Qubit™ Assay tubes and 1-20μl of DNA sample was added. The Qubit™ fluorometer was calibrated using each standard before measuring the samples. Resulting concentrations were multiplied by the dilution factor used for the DNA sample.

2.7 Vacuum Concentration of DNA Samples

Samples were run on the Speed Vac (SC110, Savant) powered by refrigerated Vapor Trap (RCT4104, Savant) for 90 minutes. Samples were then re-suspended in 15-30μl of ATE elution buffer provided from QIAamp DNA FFPE Tissue Kit (Qiagen) or H_2O.

2.8 Quantification and qualification of FFPE-extracted DNA using NGS FFPE QC kit (Agilent)

qPCR based quantification and qualification of FFPE-derived DNA samples was carried out using the NGS FFPE QC kit (Agilent) kit according to the manufacturer's instructions (handbook version). DNA concentrations were initially measured using the Qubit fluorometer (Section 2.6). Each DNA sample was subsequently diluted with nuclease-free H_2O to 2ng/μl in a 1.5ml eppendorf tube. Each sample was vortexed and spun to collect any liquid from the sides of the tube and placed on ice. In a new 1.5ml eppendorf tube, 5μl of each sample was added into 75μl of nuclease-free H_2O to reach a final concentration of 125pg/μl. All samples were vortexed and spun thoroughly and stored on ice. The dilution factor used for each sample was recorded for future use.

The qPCR was carried out with two sets of primers which amplified fragments of different sizes, Primer Set A of 42bp and Primer Set B of 123bp. Reference dye was used to compensate for non-PCR related fluorescence and provided a baseline to which samples were normalised. For the use of Agilent's AriaMx Mx3005P, the reference dye was diluted to 1:500 in nuclease-free H_2O to a final concentration of 2μM. To prepare the Primer Set A (quantification) master

mix, components were added into 1.5ml tube in the same order according to manufacturer's instructions. Briefly, master mix was prepared with Primer Set A containing the following reagents per one reaction: 10μl of 2x Brilliant III SYBR® Green QPCR Master Mix, 1μl of Primer Set A, 0.3μl of diluted reference dye and 4.7μl of nuclease-free PCR-grade H_2O. One reaction was required per one sample. Samples were vortexed, spun and kept on ice.

Primer Set B (qualification) was then prepared containing the following reagents per one reaction: 10μl of 2x Brilliant III SYBR® Green QPCR Master Mix, 1μl of Primer Set B, 0.3μl of diluted reference dye and 4.7μl of nuclease-free PCR-grade H_2O. One reaction was required per one sample. Samples were then vortexed, spun and kept on ice. 16μl of Primer Set A and Primer Set B reactions were added into individual wells of a 96well cluster flat bottom plate (Costar®) and kept on ice. 4μl of each DNA sample (or nuclease-free H_2O) was added to each well. The plate was then sealed with strip caps and vortexed. The plate was spun to collect any residual liquid from the sides of the tubes and run on an Mx3005P qPCR system (Agilent) using MxPro QPCR Software (Agilent) at the following conditions: denaturation at 95°C for 3 minutes for one cycle and 40 cycles at 95°C for 10 seconds followed by 63°C for 30 seconds. To identify the amount of amplifiable DNA, a series of calculations was used using the formula described in the handbook. This step is explained in Chapter 3 (Section 3.2.7.1).

2.9 Polymerase Chain Reaction

2.9.1 Standard PCR

Reactions were performed in a volume of 25µl using 1µl of 10µM of each primer (forward and reverse) (Sigma), 1µl of 200µM of each dNTP, 2.5µl of 10x PCR Buffer (Invitrogen), 0.75µl of $MgCl_2$ to achieve final concentration of 1.5mM, 1U Taq polymerase (Invitrogen) and 2µl of 25ng of gDNA. dH_2O was added to the final volume of 25µl. Samples were generally run at 96°C for 5 minutes to denature, followed by 35 cycles of 92°C for 20 seconds, 55-65°C for 30 seconds and 72°C for 30 seconds. The final extension step was performed at 72°C for 10 minutes.

2.9.2 HotShot Diamond PCR

Reactions were prepared with HotShot Diamond mastermix (Clent Life Science) according to manufacturer's instructions. Touchdown (TD) PCR or standard PCR cycles were used (Section 2.9.1). TD-PCR cycle begins with a denaturation step at 95°C for 10 minutes, followed by 14 cycles of 95°C for 30 seconds, 67.5°C-72.5°C for 30 seconds (decreasing by 0.5°C/cycle) and 72°C for 30 seconds. For the following 13 cycles, the annealing temperature was maintained between 60-65°C. A final extension step was performed at 72°C for 7 minutes. Samples were then kept at 4°C until required.

2.9.3 Agarose Gel Electrophoresis

DNA fragment separation was carried out on 0.5-2% w/v gel prepared using molecular biology grade agarose powder (Bioline) dissolved in 1x TAE buffer

(Section 2.1), followed by addition of ethidium bromide at a concentration of 0.5μl/ml or 4μl per 100ml of Midori Green Advance (Nippon Genetics, Germany). Samples were diluted with 10x gel loading dye (Section 2.1) and run alongside an EasyLadder I (Bioline) or GeneRuler 1kb (ThermoFisher Scientific) size standards. Electrophoresis was performed at 120V for 45-60min depending on size of expected fragments. A ChemiDoc™ MP Imaging System (Bio-Rad) was used to visualise all agarose gels. Images were generated by Image Lab 6.0 analysis software (Bio-Rad).

2.10 Purification of PCR products

2.10.1 Purification of PCR product using the QIAquick® PCR purification kit (Qiagen)

5x volume of Buffer PB (binding buffer) was added to 1x volume of the PCR reaction. The solution was transferred into a QIAquick® column and centrifuged for 30-60 seconds at 17,900 x *g*. Flow-through was discarded and 750μl of Buffer PE (wash buffer) was added to the sample which was then centrifuged for 30-60 seconds at 17,900 x *g* and the flow-through was discarded. To remove any residual wash buffer, centrifugation was repeated. At this point, the QIAquick® column was placed in a sterile 1.5ml microcentrifuge tube and 30μl of Buffer EB (elution buffer) was added to the centre of the membrane. Sample was incubated for 1 minute and centrifuged in microcentrifuge at 17,900 x *g*. Eluted purified PCR product was quantified using NanoDrop (Section 2.13).

2.10.2 Purification of PCR product from agarose gel using the QIAquick Gel Extraction kit (Qiagen)

The QIAquick gel extraction kit (Qiagen) was used according to the manufacturer's instructions. The DNA band of interest was excised from the agarose gel using a UV transilluminator. The slice was weighed and 3x the volume of buffer QG (solubilisation buffer) to 1x volume of the gel (100mg = 100μl) was added and the solution was then incubated at 50°C for 10 minutes until dissolved. 1x gel volume of isopropanol was then added to the solution and mixed and sample was centrifuged for 1 minute at 17,900 x *g* to bind the DNA in MinElute™ spin column. To remove any residual traces of agarose 500μl of buffer QG was added to the column followed by the addition of 750μl of buffer PE to wash the column and centrifuging for 1 minute at 17,900 x *g*. Flow-through was discarded and sample was centrifuged for additional 1 minute at 17,900 x *g*. MinElute™ spin column was then placed into a sterile 1.5ml microcentrifuge tube and DNA was eluted with the addition of 30μl of dH_2O to the centre of the column. The column was left to stand for 1 minute to increase DNA yield and was then centrifuged for 1 minute at 17,900 x *g.* Eluted DNA was stored at -20°C.

2.10.3 Purification of PCR product and plasmids from agarose gel using the GeneClean® II Kit (MP Biomedicals)

GeneClean® II Kit (MP Biomedicals) was used for purifying DNA from agarose gels. The DNA band was excised from agarose gel using a sterile scalpel. The gel slice was weighed and placed in a clean 1.5ml microcentrifuge tube where

3x volume of NaI solution was added to 1x volume of the gel slice and incubated at 45-55°C for 1 minute or until the agarose band had dissolved. The volume of GLASSMILK® needed was calculated based on the amount of DNA and the volume of NaI solution. 1μl of GLASSMILK® binds 1-2μg of DNA. For example, for <500μl of DNA and NaI final volume, the maximum bound DNA was <5μg and, therefore, 5μl of GLASSMILK® was used. GLASSMILK® was re-dissolved by vortexing for 1 minute and the required volume of GLASSMILK® was added to the NaI/DNA solution which was then incubated at RT for 5 minutes. This allowed for the binding of DNA to the silica matrix. Suspension was mixed every 1-2 minutes to prevent GLASSMILK® sedimenting to the bottom of the tube. Suspension was pelleted by centrifugation at 14,000 x *g* for 5 seconds and supernatant was discarded. 500μl of NEW wash buffer was added and centrifuged for further 5 seconds. The wash step was repeated one more time and afterwards, the pellet was dried to remove residual ethanol. The exact same volume of dH_2O was added as of GLASSMILK® calculated earlier to resuspend the pellet and elute the DNA. Sample was centrifuged at 14,000 *x g* for 30 seconds and eluted DNA was very carefully transferred into a new tube.

2.11 Sanger Sequencing

2.11.1 Clean-up of DNA PCR product with ExoSAP-IT (Affymetrix)

PCR products were treated with ExoSAP-IT (Exonuclease I and Shrimp Alkaline Phosphatase) (Affymetrix) according to the manufacturer's

instructions. ExoSAP-IT was added to PCR products in a 5:2 ratio. Samples were heated for 15 minutes at 37°C followed by 15 minutes at 85°C.

2.11.2 Sanger Sequencing reaction and precipitation

Each sequencing reaction included 1μl BigDye Terminator 3.1 (Applied Biosystems), 1.5μl of 5x BigDye Terminator Sequencing Buffer (Applied Biosystems), 1.6pmol of sequencing primer, 5.5μl of de-ionised H_2O and 1μl of ExoSAP-IT treated PCR product or 100ng of purified plasmid DNA. After an initial denaturation step was carried out at 96°C for 1 minute, 25 cycles of 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes were performed. All stages were ramped at 1°C/second. Following cycling, samples were precipitated by adding 5μl of 125mM EDTA and 60μl of 100% ethanol. Samples were centrifuged at 3,061 x *g* at 22°C for 30 minutes, supernatant was removed and freshly prepared 70% ethanol was added. Samples were centrifuged at 805 x *g* at 4°C for 15 minutes. Pellets were re-suspended in HiDi Formamide (Applied Biosystems) and run on an ABI3130xl Genetic Analyser (Applied Biosystems) using POP7 polymer (Applied Biosystems) and the default RapidSeq36POP7 module. The results were analysed using Seqscape (V2.5, Applied Biosystems).

2.12 Whole genome amplification (WGA)

The Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare) was used according to manufacturer's instructions. Under the General Protocol section, 9μl of Sample Buffer was added to 1μl of 10ng/μl DNA template. The reactions were heated to 95°C for 3 minutes and then cooled to 4°C on ice. Master mix

was prepared on ice by combining 9μl of Reaction Buffer with 1μl of Enzyme Mix. 10μl of prepared master mix was added to each cooled sample. Samples were kept on ice before the reactions were incubated for 1.5 hours at 30°C followed by incubation for 10 minutes at 65°C and then cooled to 4°C. After amplification, samples were either used neat or diluted 1:20 ratio in sterile dH_2O. Using control primers, a PCR was performed to make sure the concentration of DNA was sufficient for amplification.

2.13 NanoDrop 1000 Spectrophotometer

NanoDrop software (ThermoFisher Scientific) was initialized by loading 1μl of nuclease-free H_2O. It was then blanked by loading 1μl of diluent used in samples to be measured, i.e. EB, TE buffer or H_2O. Samples were then loaded in the same volume and quantified in ng/μl. The NanoDrop used the spectrum of 260nm to measure the absorbance of each sample.

2.14 High Sensitivity DNA Kit for Bioanalyzer (Agilent)

Extracted DNA was measured in size and concentration on a 2100 Bioanalyzer (Agilent) using the Agilent High Sensitivity DNA Kit (Agilent) according to manufacturer's instructions. The High Sensitivity DNA kit measures fragments between 50-7,000bp long.

2.15 TapeStation 2200 (Agilent)

Extracted DNA from FFPE tissues was measured in size and concentration on 2200 TapeStation (Agilent) using the High sensitivity D1000 DNA screen tape

(Agilent) as per manufacturer's instructions with no deviations. D1000 screen tape measures fragments between 35-1,000bp long.

2.16 WES library preparation using SureSelect^{XT} and NEBNext® Ultra II DNA Library Prep Kit

The libraries were prepared for all patients using the NEBNext® Ultra II DNA Library Prep Kit (Illumina) and SureSelect^{XT} (Agilent Technologies) kit that were combined into one protocol. This protocol was optimised for FFPE samples with low DNA input by a colleague Laura Crinnion (Leeds Next Generation Sequencing Facility). The reagents used in the NEBNext® Ultra™ II DNA Library Prep Kit (NEB) included the NEBNext Ultra II End Prep Enzyme Mix, NEBNext Ultra II End Prep Reaction Buffer, NEBNext Ultra II Ligation Master Mix, the NEBNext Ligation Enhancer and the NEBNext Ultra II Q5 Master Mix. Reagents used from the SureSelect^{XT} included the SureSelect Adaptor Oligo Mix, SureSelect Primer and SureSelect ILM Indexing Pre-Capture PCR Reverse Primer.

First, the DNA quality, quantity and fragment size of the libraries were assessed using the 2100 Bioanalyzer (Agilent Technologies) using the Agilent High Sensitivity DNA Kit (Agilent Technologies) (Section 2.14). Each sample was sheared for 2 minutes 20 seconds to 150-250bp using the Covaris S2 system (Covaris®). Post-shearing, samples were assessed by the 2100 Bioanalyzer (Agilent Technologies) using the Agilent High Sensitivity DNA Kit (Agilent Technologies).

Ends were repaired by adding 3μl of NEBNext Ultra II End Prep Enzyme Mix, 6.5μl of NEBNext Ultra II End Prep Reaction Buffer, 7.5μl of EB buffer and 48μl

of sheared DNA. Samples were heated at 20°C for 30 minutes and at 65°C for 30 minutes. To ligate paired-end adaptors to each fragment, 15μl of NEBNext Ultra II Ligation Master Mix, 2.5μl of SureSelect Adaptor Oligo Mix and 1μl of the NEBNext Ligation Enhancer were added to 65μl of end repaired DNA sample and incubated at 20°C for 15 minutes. Samples were then cleaned up using the Agencourt AMPure XP beads (Beckman Coulter). 83.5μl of beads were placed in magnetic stand to separate the PCR product, which binds to the beads, from unwanted contaminants, such as primers and nucleotides. Samples were washed with 200μl 70% ethanol and re-dissolved in 28μl of 0.1X TE buffer.

PCR amplification of all samples was carried out using 25μl of NEBNext Ultra II Q5 Master Mix along with 1μl of SureSelect Primer, 1μl of SureSelect ILM Indexing Pre-Capture PCR Reverse Primer and 23μl of adaptor ligated DNA. Initial denaturation was at 98°C for 30 seconds and annealing temperature at 98°C for 10 seconds, 65°C for 30 seconds and 72°C for 30 seconds for a total of 12 cycles. The extension step was 72°C for 5 minutes and samples were held at 4°C. Samples were cleaned up using the Agencourt AMPure XP beads (Beckman Coulter) as described earlier but now eluting in nuclease-free H_2O instead of TE buffer. The quality was assessed on the 2100 Bioanalyzer (Agilent Technologies) using the Agilent High Sensitivity DNA Kit (Agilent Technologies). The amplified DNA library was hybridised to the capture library of biotinylated RNA oligo probes by preparing 2.3-3.4μl (maximum amount available for FFPE samples) of 221ng/μl prepped library (final concentration of 250-1000ng for FFPE library) and preparing the Hybridisation Buffer mixture

by adding 6.63µl of SureSelect Hyb 1, 0.27µl of SureSelect Hub 2, 2.65µl of SureSelect Hyb 3 and 3.45µl of SureSelect Hyb 4. The SureSelect Block Mix was prepared by adding 2.5µl SureSelect Indexing Block 1, 2.5µl of SureSelect Block 2 and 0.6µl of SureSelect ILM Indexing Block 3. 5.6µl of prepared SureSelect Block Mix was added to each DNA sample. Samples were incubated for 5 minutes at 95°C. Preparation of Capture Library Hybridisation Mix was carried out as follows: 13µl of the Hybridisation Buffer mixture, 2µl of RNase Block Dilution (1:3 RNase block: H_2O) and 5µl of SureSelect^{XT} Human All Exon v5 (Agilent) library. 20µl of the Capture Library Hybridisation Mix was added to the SureSelect Block Mix containing the DNA samples. The hybridisation mixture was incubated for 16 to 24 hours at 65°C with a heated lid at 105°C. Streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1) (Invitrogen) were prepared for selection of the captured library by washing with binding buffer. The captured library was then selected using streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1) which were separated from the buffer using a magnetic separator. The supernatant was removed, and the beads were re-dissolved in 200µl of SureSelect Wash 1. The beads were then washed using 200µl of SureSelect Wash 2 for a total of 3 washes. The library was recovered with the magnetic separator and re-dissolved in 30µl of H_2O.

Index tags were added to each sample by PCR by adding 30µl of each library, 2.5µl of nuclease-free H_2O, 10µl of 5x Herculase II Reaction Buffer, 1µl of Herculase II Fusion DNA Polymerase, 0.5µl of 100mM dNTP Mix and 1µl of SureSelect ILM Indexing Post Capture Forward PCR Primer. 5µl of the appropriate indexing primer of 6bp (PCR Primer Index 1-16 depending on the

number of libraries) was added to each sample. The samples were amplified using the following settings: 98°C for 2 minutes, 10 cycles at 98°C for 30 seconds, 57°C for 30 seconds and 72°C for 1 minute followed by 72°C for 10 minutes and hold temperature of 4°C. The captured libraries were purified by Agencourt AMPure XP beads (Beckman Coulter) as previously described with the final elution volume of 30µl. Libraries were assessed using the 2100 Bioanalyzer (Agilent Technologies) using the Agilent High Sensitivity DNA Kit (Agilent Technologies) with the expected DNA fragments of 200-300bp.

All seven samples were now pooled into one lane at a final concentration of 10nM in a final volume of 50µl by using the following formula: volume of library = (Vf x Cf)/(# x Ci), where Vf is the final desired volume of the pool (50µl), Cf is the final desired DNA concentration (10nM), # is the number of samples (7) and Ci is the initial concentration of each sample. Next generation sequencing was carried out as 100bp paired-end read on an Illumina HiSeq™2500 by the University of Leeds Next Generation Sequencing Facility (http://dna.leeds.ac.uk/genomics/).

2.17 WES library preparation using SureSelect^{XT} 200ng kit (Agilent)

Exome capture was carried out using the SureSelect^{XT} 200ng kit (Agilent) for both, FFPE DNA (Coats samples) and non FFPE DNA (FEVR samples). WES library preparation for all FEVR samples was carried out by the NGS Sequencing Facility (University of Leeds) and the protocol was followed according to the manufacturer's instructions with no modifications. For FFPE DNA for all Coats disease samples, the protocol was followed according to

instructions with the recommended modifications for FFPE DNA, described in detail below.

The reagents used in the SureSelect^{XT} kit included the ILM, SureSelect Target Enrichment Kit ILM Indexing Hyb Module Box #2, Herculase II Fusion DNA Polymerase and SureSelect Target Enrichment Kit Box #1 (Agilent).

First, the DNA was assessed by the NGS FFPE QC Kit (Agilent) (Section 2.8) and the input of 200ng of amplifiable DNA was used. After this quantification, the quality, quantity and fragment size of the libraries were assessed with the High Sensitivity D1000 Screen Tape and run on the Tapestation 2200 (Agilent) (Section 2.15). Each sample was sheared for 4 minutes for FFPE DNA instead of 6 minutes to 150-250bp using the Covaris S2 system (Covaris®). Post-shearing, samples were assessed by the High Sensitivity D1000 Screen Tape (Agilent) using the Tapestation 2000 (Agilent). The electropherograms showed a distribution with a peak size of 150 to 200bp. Ends were repaired by adding 10μl of 10x End Repair Buffer, 1.6μl dNTP Mix, 1μl of T4 DNA polymerase, 2μl of Klenow DNA polymerase, 2.2μl of T4 Polynucleotide Kinase and 35.2μl of nuclease-free H_2O to the purified sheared DNA. Samples were incubated at 20°C for 30 minutes and then purified by the addition of 180μl of Agencourt AMPure XP beads (Beckman Coulter). Beads with sample were placed in magnetic stand to separate the PCR product, which binds to the beads, from unwanted contaminants, such as primers and nucleotides. Samples were washed with 500μl 70% ethanol and re-dissolved in 32μl of nuclease-free H_2O. The 3' ends of DNA fragments were adenylated by mixing 30μl of each sample, 11μl of nuclease-free H_2O, 5μl of 10x Klenow Polymerase Buffer, 1μl of dATP and 3μl of Exo(-) Klenow. The mix was incubated at 37°C for 30 minutes.

Samples were now purified as described earlier using the Agencourt AMPure XP beads (Beckman Coulter) with the final elution volume of 13µl. To ligate paired-end adaptors, 13µl of each sample was used with the addition of 15.5µl of nuclease-free H_2O, 10µl of 5x T4 DNA Ligase Buffer, 10µl of SureSelect Adaptor Oligo Mix and 1.5µl of T4 DNA Ligase. The reaction was incubated at 20°C for 15 minutes. Samples were again purified as described by Agencourt AMPure XP beads (Beckman Coulter) with the final elution volume of 30µl. After the purification, the adaptor-ligated libraries were PCR amplified by adding 30µl of each sample together with 6µl of nuclease-free H_2O, 1.25µl of SureSelect Primer, 1.25µl of SureSelect ILM Indexing Pre-Capture PCR Reverse Primer, 10µl of 5x Herculase II Rxn Buffer, 0.5µl of 100mM dNTP Mix and 1µl of Herculase II Fusion DNA Polymerase. Samples were amplified under the following PCR conditions: initial denaturation step at 98°C for 2 minutes and annealing temperatures at 98°C for 30 seconds, 65°C for 30 seconds, 72°C for 30 seconds for 10-13 cycles (depended on the DNA integrity score obtained for each sample by the NGS FFPE QC kit, Agilent) (Section 2.8). The extension step was at 72°C for 10 minutes and samples were held at 4°C. Samples were then cleaned up using Agencourt AMPure XP beads (Beckman Coulter) as described previously with the final elution volume of 30µl. The samples were assessed using the High Sensitivity D1000 Screen Tape and Tapestation 2200 (Agilent) and clear peaks at 225-275bp were obtained. The amplified DNA library was hybridised to the capture library of biotinylated RNA oligo probes by preparing 2.3-3.4µl (maximum amount available for FFPE samples) of 221ng/µl prepped library (final concentration of 500-750ng for FFPE library) and preparing the Hybridisation Buffer mixture by

adding 6.63µl of SureSelect Hyb 1, 0.27µl of SureSelect Hub 2, 2.65µl of SureSelect Hyb 3 and 3.45µl of SureSelect Hyb 4. The SureSelect Block Mix was prepared by adding 2.5µl SureSelect Indexing Block 1, 2.5µl of SureSelect Block 2 and 0.6µl of SureSelect ILM Indexing Block 3. 5.6µl of prepared SureSelect Block Mix was added to each DNA sample. Samples were incubated for 5 minutes at 95°C. Preparation of Capture Library Hybridisation Mix was carried out as follows: 13µl of the Hybridisation Buffer mixture, 2µl of RNase Block Dilution (1:3 RNase block: H_2O) and 5µl of SureSelect^{XT} Human All Exon v6 (Agilent) library. 20µl of the Capture Library Hybridisation Mix was added to the SureSelect Block Mix containing the DNA samples. The hybridisation mixture was incubated for 16 to 24 hours at 65°C with a heated lid at 105°C. Streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1) (Invitrogen) were prepared for selection of the captured library by washing with binding buffer. The captured library was then selected using streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1) which were separated from the buffer using a magnetic separator. The supernatant was removed, and the beads were re-dissolved in 200µl of SureSelect Wash 1. The beads were then washed using 200µl of SureSelect Wash 2 for a total of 3 washes. The library was recovered with the magnetic separator and re-dissolved in 30µl of H_2O.

Index tags were added to each sample by PCR by adding 30µl of each library (for FFPE samples to account for greater volume of captured DNA), 2.5µl of nuclease-free H_2O (reduced volume of H_2O for FFPE samples), 10µl of 5x Herculase II Reaction Buffer, 1µl of Herculase II Fusion DNA Polymerase, 0.5µl of 100mM dNTP Mix and 1µl of SureSelect ILM Indexing Post Capture

Forward PCR Primer. 5µl of the appropriate indexing primer of 6bp (PCR Primer Index 1-16 depending on the number of libraries) was added to each sample. The samples were amplified using the following settings: 98°C for 2 minutes, 10 cycles at 98°C for 30 seconds, 57°C for 30 seconds and 72°C for 1 minute followed by 72°C for 10 minutes and hold temperature of 4°C. The captured libraries were purified by Agencourt AMPure XP beads (Beckman Coulter) as previously described with the final elution volume of 30µl. Libraries were assessed using the High Sensitivity D1000 Screen Tape and run on the Tapestation 2200 (Agilent) with the expected fragment size of 300bp. All seven samples were now pooled into one lane at a final concentration of 10nM in a final volume of 50µl by using the following formula: volume of library = (Vf x Cf)/(# x Ci), where Vf is the final desired volume of the pool (50µl), Cf is the final desired DNA concentration (10nM), # is the number of samples (7) and Ci is the initial concentration of each sample. Next generation sequencing was carried out as 150bp paired-end reads on an Illumina HiSeq™3000 sequencing platform by the University of Leeds Next Generation Sequencing Facility (http://dna.leeds.ac.uk/genomics/).

2.18 Bioinformatics Analysis

Bioinformatics analysis was performed with the same pipeline for both, Coats FFPE and FEVR samples. Any differences between the two pipelines are described in great detail in Chapter 3, Section 3.2.7.3.

The raw fastq files generated from WES sequencing were trimmed using trim galore and aligned to human reference genome GRCh37 (build hg19) using Burrow's wheeler aligner (BWA) (Li and Durbin, 2009) to generate Sequence

Alignment/Map format (SAM) files. SAM file composes of a header and alignment of all bases. SAM file was then converted to a BAM file using SAM tools (Li et al., 2009). BAM file is a binary version of SAM file. All duplicates were then removed with Picard (http://broadinstitute.github.io/picard) and the indel realignment targets, base recalibration and variant calling were created by GATK. At this point, determination of the coverage of each base within the exome for each sample was performed by a tool called Depth of coverage (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_coverage_DepthOfCoverage.php, produced by Genome Analysis Toolkit (GATK) (https://software.broadinstitute.org/gatk/) (McKenna et al., 2010). It uses aligned BAM files to check for depth of sequencing per locus, gene or overall. Variants were then called to generate final genomic variant call format (gVCF) (VdAuwera, accessed 31 October 2017) from patients' BAM files. gVCF were merged to allow for joint analysis. Variant calling was performed using GATK best practices (DePristo et al., 2011). This step generates VCF files using Haplotype Caller (https://samtools.github.io/hts-specs/VCFv4.2.pdf) (accessed 31 October 2017). Variants were subsequently filtered using publicly available databases of variants, such as dbSNP142, (https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi?build_id=142), ExAC (http://exac.broadinstitute.org) and gnomAD (http://gnomad.broadinstitute.org) to remove any variants that have minor allele frequency (MAF) of ≥1% and further ≥0.01% in population. After the removal of common polymorphisms, filtered VCFs were subjected to the variant effect predictor (VEP). VEP defines the functional effect of all the

variants, i.e. deletion, insertion, frameshift, nonsense, synonymous, missense, intronic or non-coding (https://github.com/Ensembl/ensembl-vep). Resulting lists were further filtered with in-house perl scripts such as "getFunctionalVariants.pl", "getHetVariants.pl" and "findBiallelic.pl" (available from VCFhacks, https://github.com/gantzgraf/vcfhacks) to select for variants based on their effect on the coding amino acid sequence or the mode of inheritance of the samples. Variants were then ranked on CADD score and annotated using the Ensembl GeneAnnotator from vcfhacks. VCFs were then converted into Excel files for manual filtering. This filtering was based on variants that have passed the quality threshold filter, which relies on Java Expression Language (JEXL) expressions and contains five different expressions which are QualByDepth (QD), FisherStrand (FS), RMSMapping Quality (MQ), MappingQualityRankSumTest (MQRankSum) and ReadPosRankSumTest (ReadPosRankSum) (https://software.broadinstitute.org/gatk/documentation/article.php?id=1255). The JEXL operates in a way that it specifies what annotations to select and the rules that it needs to apply to be able to select the variants of interest. To do this, it uses five expressions to set thresholds and cut offs for both SNPs and indels. Once the variants are above the corresponding cut off values, they were deemed to have passed the quality threshold. Subsequently, VCFs were converted into Excel files. An additional analysis looking for CNVs was performed using the R package, ExomeDepth (https://cran.r-project.org/web/packages/ExomeDepth/index.html) (Plagnol et al., 2012) to identify any large deletions or insertions. ExomeDepth was performed on autosomes and the X chromosome (Section 2.20).

2.19 Online Bioinformatics Tools

2.19.1 Literature searches

All literature searches were made using PubMed (https://www.ncbi.nlm.nih.gov/pubmed). Additional information was obtained from Online Mendelian Inheritance in Man (OMIM) (https://omim.org). OMIM contains information on disease phenotypes related to known genes and their chromosomal location.

2.19.2 Primer design

Primers were designed using the ExonPrimer software available through the UCSC genome browser (https://ihg.helmholtz-muenchen.de/cgi-bin/primer/ExonPrimerUCSC.pl?db=hg38&acc=uc001ont.4). This tool was based on the primer design tool Primer3 but automates the primer design for screening gDNA or cDNA and avoids designing primers over known SNPs.

Standard settings using these tools included:

Minimal distance between primer and exon/intron boundary:	35bp
Primer region:	70bp
Maximal target (exon) size:	500bp
Overlap (for bigger exons):	50bp
Annealing temperature:	60°C
Primer size:	17bp (min), 20bp (optimum), 27bp (max)
GC%:	20% to 80%
Maximum length of a mononucleotide repeat:	4bp

Another tool called AutoPrimer3 (https://github.com/gantzgraf/autoprimer3) was also used. This tool was designed to generate primers to genomic coordinates and genes using primer3 (http://primer3.ut.ee). This tool also detected SNPs underneath primer sequences.

Primer-target buffer:	50bp
Maximal target (exon) size:	500bp
Overlap (for bigger exons):	50bp
Annealing temperature:	60°C
Primer size:	17bp (min), 20bp (optimum), 27bp (max)

Primers to screen FFPE DNA were designed manually keeping to the following parameters: PCR product of between 60-170bp, 20bp primers, 50% GC content and annealing temperature of 55°C - 65°C. Primers for FFPE DNA were checked by BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) to ensure no SNPs were detected under each primer and to ensure that primer pairs are unique using Primer-Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). All primers were manufactured by Sigma and purified by desalting.

2.19.3 UCSC Genome Browser

In 2012, the first version of UCSC genome browser was launched (Kent et al., 2002) (http://genome.ucsc.edu/). It consisted of reference genomes for human, mouse, fruitfly, worm, yeast and others and can be used to search for genomic sequences, predicted protein sequences and mRNA sequences. It also uses various tools to analyse the genome sequences such as BLAT, Table browser and others. Initial search and information regarding specific

genes were found on the UCSC genome browser. In this study, BLAT and ExonPrimer were used.

2.19.4 Variant pathogenicity prediction tools

2.19.4.1 Combined Annotation Dependent Depletion (CADD)

CADD score (https://cadd.gs.washington.edu) is a pre-computed matrix of all possible SNVs in the human genome that is used for scoring the damaging probability of single nucleotide polymorphisms (SNPs) and indel variants in the human genome. A quantitative score called C-score is used for scoring the deleteriousness of a variant (Kircher, 2014). CADD incorporates annotations into one metric by contrasting variants that survived natural selection with non-real mutations.

The CADD score ranges from 1 to 99, based on the rank of each variant with the top 10% of deleterious variants having score 10 and above, the top 1% scores of 20 and above and top 0.1% having scores 30 and above. Version GRCh37-v1.3 was used.

2.19.4.2 PolyPhen-2

PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) uses the comparison between the wild type allele and the corresponding disease-causing allele and its effect on the harbouring protein (Adzhubei et al., 2010). It scores the impact of variants on protein function and structure based on the phylogenetic and structural characteristics of the substitution. The score ranges between 0 to 1. The higher the score the higher the probability that the variant is deleterious.

2.19.4.3 Blosum62

Blosum62 uses a scoring system from substitution matrices that is based on calculating the likely trades of one amino acid with another amino acid. The scoring system ranges between -4 to +3. A negative score suggests a more damaging variant (Henikoff and Henikoff, 1992).

2.19.4.4 Sorting Intolerant from Tolerant (SIFT)

SIFT (http://sift.jcvi.org) predicts if a substitution in a genomic sequence affects the protein function. The scoring system categorises the substitutions as tolerated or deleterious based on sequence homology (Ng and Henikoff, 2001).

2.19.4.5 Protein Variant Effect Analyzer (PROVEAN)

PROVEAN (http://provean.jcvi.org/index.php) calculates the effect of a single amino acid change or indels would be on the protein sequence by obtaining pairwise sequence alignment scores that generated precomputed predictions for all single amino acids in human and mouse (Choi et al., 2012).

2.19.4.6 MutPred 1.2

MutPred (http://mutpred.mutdb.org) uses the protein sequence to describe the changes between the wild type allele and the mutant allele to see if the change is disease-associated. The prediction works by predicting the impact of the variant on over 50 protein properties which allows for the scoring of pathogenicity based on the inference of molecular mechanisms (Li et al., 2009).

2.19.4.7 Database of genomic variation and phenotype in humans using Ensembl resources (DECIPHER)

DECIPHER (https://decipher.sanger.ac.uk) is a database used for interpreting genomic variants by retrieving information from many bioinformatics resources relevant to the variant in question. The variant is displayed as normal and pathogenic variation reported at that locus which allows for the interpretation.

2.19.5 Exome Variant Server (EVS)

Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS/) is a database that records SNPs and indels reported from 6,503 control individuals. EVS was used for removing variants with minor allele frequency of ≥1% and ≥0.01%.

2.19.6 Exome Aggregation Consortium (ExAC)

Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org) provides exome data from 60,706 unrelated individuals without severe congenital disease. ExAC was used for removing variants with minor allele frequency of ≥1% and ≥0.01%.

2.19.7 gnomAD

The Genome Aggregation Database (http://gnomad.broadinstitute.org) incorporates data from ExAC (Section 2.19.6) with additional exome and genome data to a total of 123,136 exomes and 15,496 genomes from unrelated individuals. gnomAD was used for removing variants with minor allele frequency of ≥1% and ≥0.01%.

2.19.8 Homologene

Homologene aligns protein sequences across different species to look at conservation of amino acids. Multiple protein alignments were obtained from NCBI Homologene (https://www.ncbi.nlm.nih.gov/homologene/) (Agarwala et al., 2016).

2.19.9 Simple Modular Architecture Research Tool (SMART)

SMART (Schultz et al., 1998; Letunic et al., 2014) (http://smart.embl-heidelberg.de) was used to identify domains and the domain architectures of proteins of interest.

2.19.10 Integrative Genomics Viewer (IGV)

IGV (Robinson et al., 2011; Thorvaldsdóttir et al., 2013) (http://software.broadinstitute.org/software/igv/) is a visualization tool that allows the interactive exploration of diverse, large genomic datasets. In this study it was used for viewing reads and coverage over patients' exomes and variants identified over specific genes.

2.20 WES copy number variations (CNV) analysis using ExomeDepth

Exome depth (Plagnol et al., 2012) was used to detect large CNVs that are not detected by SNV analysis of WES data (https://cran.r-project.org/web/packages/ExomeDepth/index.html) (Section 2.20). This software compares the exon read depths across an exome with the read depth of a reference set of exomes, typically 5-10. Reference exomes were from

unrelated individuals whose libraries were generated alongside the query sample from the same run. The output files were provided as a .csv (comma-separated values) file and were annotated to flag known CNVs (Conrad et al., 2010). ExomeDepth outputs were viewed in Excel, with predicted CNVs scored according to Bayes factor (BF), to search for CNVs in known disease genes. The BF quantifies the statistical value for each CNV and it represents the log10 of the ratio for the CNV call divided by the normal copy number. The higher the resulting number, the more confident the software is that the CNV is real.

2.21 Molecular cloning

2.21.1 Restriction enzyme digestion

Restriction enzymes were purchased from Promega and used according to manufacturer's instructions. Briefly, each reaction was set up in the following manner: 2μl of restriction enzyme 10X buffer, 0.2μl of 10μg/μl acetylated BSA, 1μl of DNA, 1μg/μl and 16.3μl of deionized H_2O. Reactions were then mixed by pipetting before adding 0.5μl of 10u/μl restriction enzyme and incubated at enzyme's optimum temperature for 1-4 hours. The digestion was visualised by adding the loading buffer to each reaction at 1X final concentration and proceeded to gel electrophoresis.

2.21.2 Site-directed mutagenesis (SDM) using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent)

SDM was carried out using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent) according to the manufacturer's instructions. Mutagenic primers

were designed using the web-based QuikChange Primer Design Program available online (http://www.genomics.agilent.com/primerDesignProgram.jsp). Primers were purified with HPLC or PAGE unless otherwise stated. To prepare a sample reaction, the following components were added into thin-walled PCR tubes: 5μl of 10x reaction buffer, 10ng of dsDNA template, 125ng of oligonucleotide primer #1, 125ng of oligonucleotide primer #2, 1μl of dNTP mix, 3μl of QuikSolution reagent and up to 50μl of double-distilled H_2O (ddH_2O). Once the sample reaction mix was made up, 1μl of PfuUltra HF DNA polymerase (2.5 U/μl) was added to each sample. A control reaction was prepared in the same manner using the pWhitescript 4.5kb control plasmid (5ng/μl) and control primers provided in the kit. Each reaction was cycled using the parameters as follows: 1 minute at 95°C for denaturation, 18 cycles of 95°C for 50 seconds, 60°C for 50 seconds and 68°C for 1 minute per every kb of plasmid length. Final extension step was carried out for 7 minutes at 68°C. PCR products were then digested with Dpn I in the following manner: 1μl of the Dpn I restriction enzyme (10 U/μl) was added to each amplification reaction. Reactions were spun down for 1 minute to collect any fluid and then immediately incubated for 1 hour at 37°C to digest the parental (non-mutated) double stranded DNA. The resulting plasmid was transformed into XL-10 Gold Ultracompetent cells supplied with the kit with the addition of β-Mercaptoethanol (β-ME) which increases the transformation efficiency (Section 2.21.5).

2.21.3 Plasmid building tools

Software such as Serial cloner (https://serial-cloner.en.softonic.com) and Snapgene (http://www.snapgene.com) were used to visualise the plasmids created by Gateway Technology (Section 2.21.4). Reactions such as digestion, BP and LR reactions for Gateway cloning were visualised.

2.21.4 Generation of expression constructs using Gateway® cloning technology (Invitrogen)

Gateway technology is a cloning method that uses site-specific recombination sites of bacteriophage lambda (att sites) to efficiently move DNA sequences into vectors (https://www.thermofisher.com/ch/en/home/life-science/cloning/gateway-cloning.html). For the generation of all expression constructs, manufacturer's instructions were followed. The cDNA sequences were first amplified with attB-tagged primers. The insert was then cloned into pDONR201 using the Gateway® BP Clonase II Enzyme Mix (Invitrogen, Life Technologies) to produce an entry clone (work was carried out by Soriano, 2017). The gene cDNA sequence was then transferred into an expression vector using the Gateway® LR Clonase™ II Enzyme Mix (Invitrogen, Life Technologies). The LR reaction comprised of an entry clone (50ng-150ng), destination vector pcDNA™-pDEST40 (Invitrogen) or pcDNA™-pDEST47 (Invitrogen) (150ng/μl) and TE buffer up to 8μl. To each sample prepared, 2μl of LR Clonase™ II enzyme mix was added and mixed well by vortexing twice. Reactions were centrifuged briefly to collect any liquid from the sides and incubated for 1 hour at 25°C. Each reaction was terminated by adding 1μl of proteinase-K solution and incubated for 10 minutes at 37°C. Transformation

into dH5α α-Select Competent Cells (Bioline) ultra-competent cells was carried out as described in Section 2.21.5. Empty gateway vectors were propagated in One Shot® ccdB Survival™2 T1R chemically competent cells (ThermoFisher Scientific).

2.21.5 Bacterial transformation and culture

DH5α α-Select Competent Cells (Bioline), XL-10 Gold Ultracompetent cells (Agilent) and the One Shot® ccdB Survival™2 T1R chemically competent cells (ThermoFisher Scientific) were gently thawed on ice. 1μl of the construct at 100-150ng from each reaction (control and sample) was transferred to 45μl of the ultracompetent cells. Transformation reactions were gently mixed and incubated on ice for 30 minutes.

NZY+ broth (for SDM, ThermoFisher Scientific) and S.O.C media (Invitrogen) was preheated to 42°C. The reactions were heat-shocked at 42°C for 30 seconds and incubated on ice for 2 minutes. To each reaction, 0.5ml of preheated NZY+ broth or S.O.C media was added and incubated for 1 hour at 37°C with shaking at 225-250rpm. Cells were then spread on LB (Section 2.1) containing the appropriate antibiotic at 250μl of the pWhitescript mutagenesis control and 100μl of mutagenesis sample was plated. Transformation plates were incubated for >16 hours at 37°C.

2.21.6 Plasmid DNA isolation and purification

2.21.6.1 Plasmid MiniPrep using QIAprep® MiniPrep kit (Qiagen)

Plasmid DNA was purified from cultures using the QIAprep® MiniPrep kit (Qiagen) according to manufacturer's instructions. Briefly, 5ml bacterial culture in LB was incubated in a shaking incubator overnight at 37°C. The following day, 4ml of bacterial culture were harvested by centrifugation for 10 minutes at 6,000 x *g* at 4-10°C and the remaining 1ml of the culture was retained for future experiments, such as MaxiPrep (Section 2.21.6.2). The pelleted bacterial cells were re-dissolved in 250μl Buffer P1 (resuspension buffer) and transferred to a 1.5ml microcentrifuge tube. 250μl Buffer P2 (lysis buffer) was added and mixed thoroughly by inverting the microcentrifuge tube 4-6 times. Further 350μl of Buffer N3 (neutralization buffer) was added and immediately mixed by inverting the microcentrifuge tube 4-6 times and centrifuged at 8,000 x *g* for 10 minutes. 800μl of the supernatant was added to a QIAprep 2.0 spin column and centrifuged for 30-60 seconds. The column was then washed by adding 0.5ml Buffer PB (wash buffer) and centrifuged at 8,000 x *g* for 30-60 seconds. A further wash was performed by adding 0.75ml Buffer PE and centrifuging at 8,000 x *g* for 30-60 seconds. After disposing of the flow-through, the column was further centrifuged at 8,000 x *g* for 1 minute to remove any residual wash buffer. The column was then placed in a clean 1.5ml microcentrifuge tube and 50μl of dH_2O was added to the centre of each column. Samples were incubated for 1 minute and centrifuged for 1 minute. Eluted DNA was stored at 4°C for immediate use or at -20°C for storage.

2.21.6.2 Plasmid MaxiPrep using EndoFree® Plasmid Maxi Kit (Qiagen)

A bacterial culture of 100ml was grown overnight before being harvested by centrifugation at 6,000 x *g* for 15 minutes at 4°C. Each pellet was re-dissolved in 10ml of Buffer P1 (resuspension buffer) followed by 10ml of Buffer P2 (lysis buffer) and mixed by inverting the tube 4-6 times. Samples were incubated at 15-25°C for 5 minutes. After the incubation, 10ml of chilled Buffer P3 (neutralization buffer) was added and the samples immediately mixed by inverting tube 4-6 times. The lysate was then poured into a prepared QIAfilter Cartridge and incubated at RT for 10 minutes. After incubation, the cap was removed, and a plunger inserted. The cell lysate was filtered into a clean 50ml falcon. 2.5ml Buffer ER (endotoxin removal buffer) was added to each lysate, inverted 10 times to mix and incubated on ice for 30 minutes. The QIAGEN-tip 500 was equilibrated with 10ml Buffer QBT (equilibration buffer) and once it had emptied, the filtered lysate was applied to the QIAGEN-tip. The QIAGEN-tip was subsequently washed with 2 x 30ml Buffer QC (wash buffer) and then DNA was eluted with 15ml Buffer QN. After elution, DNA was precipitated by adding 10.5ml 100% isopropanol at RT, mixed and centrifuged at ≥15,000 x *g* for 30 minutes at 4°C. The supernatant was discarded, and the pellet was washed with 5ml of endotoxin-free RT 70% ethanol and further centrifuged at ≥15,000 x *g* for 10 minutes. The pellet was air-dried for 5-10 minutes and then dissolved in a suitable volume (for large viscous pellets 300μl and for smaller pellets 150μl was used) in endotoxin-free H_2O. The plasmid DNA was then quantified by NanoDrop (Section 2.13).

2.22 cDNA synthesis

Prior to cDNA synthesis, all surfaces and equipment used were cleaned with RNaseZAP™ (Sigma-Aldrich) solution to prevent degradation of RNA and to remove any residual RNases. cDNA was synthesised from 1μg/μl of RNA from the Total RNA Master Panel II (Clontech) tissue library of multiple human adult and fetal tissues. Reverse-transcriptase (RT) PCR was set up in a siliconized RNase free eppendorf tube (negative control containing all the components apart from the RT enzyme and negative control of H_2O was also set up). Each reaction included 1μl of 100ng random hexamer (ThermoFisher Scientific), 1μl of total RNA (1μg/μl) and 10μl of deionized, diethylpyrocarbonate (DEPC) H_2O. Samples were incubated at 70°C for 10 minutes followed by cooling down on ice and promptly spun in a microcentrifuge. A master mix containing 4μl of 5x Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV) RT buffer (Invitrogen), 2μl of 0.1M DTT, 2μl of 10mM dNTPs and 2μl of 40 u/μl RNAsin (Promega) was added to each sample and the reaction equilibrated at 37°C for 2 minutes. 1μl (100U) of M-MLV RT (Invitrogen) per 1μg of RNA was added to each reaction and incubated for 1 hour at 37°C. To stop the reaction, samples were heated to 95°C for 5 minutes and then placed on ice. cDNA was stored at -20°C.

2.23 Cell culture

Standard sterile cell culture conditions were maintained, and the work carried out in this thesis was covered under the University of Leeds CL1 GMAG licence number 183.

2.23.1 Cell lines and sources

Cell line	Origin	Source
HEK293 STF	7x Super-TOPflash stably transfected HEK293 cells	A gift from J. Nathans (Xu et al., 2004)
HEK293	Human embryonic kidney cells	ATCC
U2OS	Human osteosarcoma cells	European Collection of Cell Cultures (ECACC)
HeLa	Human cervical carcinoma epithelial cells	ATCC

Table 2.1 Cell lines used in this study with full names and sources provided.

2.23.2 Cell culturing conditions

Cells were grown and split in Corning 75cm^2 flasks (Sigma Aldrich). Cells were incubated at 37°C with 5% CO_2 in Sanyo MCO 20AIC incubators. All cell culture work was performed in NuAire Labgard 437 ES Class II Biosafety Cabinets. HEK293 cells, HeLa cells and U2OS cells were grown in Dulbecco's Modified Eagle's Medium (DMEM) (Sigma Aldrich) with 10% fetal calf serum (FCS) (Sigma Aldrich) and 10,000 u/ml penicillin with 10 mg/ml streptomycin in 0.9% NaCl (Sigma Aldrich). HEK293 STF cells were cultured in DMEM with Ham's F12 medium (Gibco®) and 10% FCS. HEK293 STF cells were treated with 100μg/ml of Geneticin® (G418) (Gibco®) for a total of two passages to select for stably transfected cells prior to performing experiments.

All cell lines were passaged when approaching confluency by removing the media and washing with 10ml Dulbecco's PBS (DPBS) (Sigma Aldrich). Cells were then detached by adding 1ml of 1x trypsin/EDTA solution (Sigma Aldrich)

and incubating at 37°C for approximately 5 minutes. The trypsin was deactivated by adding 10ml of fresh media. The re-dissolved cells were collected into a falcon tube and centrifuged at 200 x *g* for 5 minutes. The pellets were re-dissolved in appropriate culture media and split into new flasks in the required ratio (1:10 or 1:5 for HEK293 and HEK293 STF and 1:20 for HeLa and U2OS cell lines).

2.23.3 Cell counting

Re-dissolved cells were diluted 1:1 ratio with 10μl trypan blue dye 0.4% (Life Technologies), loaded onto a countess slide and read using a Countess™ Automated Cell Counter (Life Technologies). Dead cells were visualised as blue and viable cells as white. The number of dead and viable cells was counted, and the viability percentage calculated to facilitate accurate seeding densities.

2.23.4 Cryofreeze storage and recovery

Media was removed, and cells were washed with PBS. Cells were then trypsinised and spun at 200 x *g* for 5 minutes. The pellet was re-dissolved in media supplemented with 10% DMSO. Cells were counted (Section 2.23.3) and 1.0 x 10^6 cells aliquoted per cryovial (Nunc) and stored in a Mr. Frosty™ freezing container (Nalgene) at -80°C. Cells were transferred to liquid nitrogen the following day for long term storage.

To recover frozen cells, a cryovial was thawed in a 37°C water bath. Once nearly defrosted, warmed media was slowly added to the cells. In this manner, cells were transferred from cryovial to a 10ml falcon containing the appropriate

cell media for the cell line. Cells were spun at 200 x *g* and re-dissolved in warmed media and plated out in T75 flask and incubated at 37°C with 5% CO_2.

2.23.5 Transient transfections

Transient transfections were done in 6 or 24-well plates (Corning) or a 4 quadrant dish (Ibidi®) (300,000 cells were seeded for a 6-well plate and the number scaled down appropriately for the smaller sized wells). 250μl of transfection media Opti-MEM™ (Gibco®) was added to an eppendorf with 5μl of Lipofectamine™2000 (Invitrogen). In a second eppendorf, 250μl of Opti-MEM and plasmid DNA (1.2μg) were combined. Both solutions were incubated for 5 minutes at RT before they were combined and incubated for further 20 minutes at RT. The DNA/lipofectamine/Opti-MEM solution was then added to each well in a slow dropwise motion and the cells incubated overnight. When transfecting a fluorescent tag, cells were visualised under EVOS™ microscope (Section 2.29.1).

2.24 Cell lysis and protein extraction

Media was removed from cells prior to washing twice with PBS. A cell scraper was used to collect all the cells into a 15ml falcon tube. Cells were then pelleted and re-dissolved in NP40 lysis buffer, consisting of 20mM Tris HCl pH8, 150mM NaCl, 10% v/v glycerol, 1% v/v NP40, 2mM EDTA (Sigma), 1x protease inhibitor cocktail (Roche) and 0.3M Phenylmethylsulfonyl fluoride (PMSF) (Sigma). Lysates were incubated for 30 minutes on ice, then centrifuged at 15,000 x *g* to collect the cell debris. The supernatant (lysate)

was collected into new tubes and stored at -20°C if used immediately and -80°C for long term storage.

2.25 BCA assay using Pierce BCA Protein Assay Kit (Pierce)

To determine the concentration of protein in cell lysates, the Pierce BCA Protein Assay Kit (Pierce, ThermoFisher Scientific) was used according to manufacturer's instructions. This assay is based on the reduction of Cu^{2+} to Cu^{1+} by protein in alkaline medium and the very selective colorimetric detection of Cu^{1+} by bicinchoninic acid (BCA). Protein concentration was measured at 562nm and compared to a protein standard. Provided standards were prepared as instructed with known concentrations of bovine serum albumin (BSA). 10µl of standards and sample lysates were mixed with 200µl of working reagent (WR) provided in the kit. Samples together with the standards were then plated, covered and incubated at 37°C for 30 minutes. Plate was normalized to RT and absorbance was read at 550nm on a Mitras LB 940 Luminometer (Berthold Technologies).

2.26 Dual luciferase TOPflash assay (Promega)

The TOPflash luciferase reporter assay was carried out using the Dual-luciferase reporter assay system (Promega) according to manufacturer's instructions with the minor modification that the LARII and Stop & Glo reagents were diluted by 50%.

STF cells (HEK293 cells stably transfected with the SuperTOPflash reporter construct) were grown in DMEM-F12 (Gibco®) supplemented with 10% FCS (Sigma) and 1% PenStrep (ThermoFisher Scientific). Cells were treated with

geneticin G418 (Gibco®) for two passages to select for cells containing the TOPflash reporter. Cells were allowed to undergo at least one passage to recover from the antibiotic selection. For the TOPflash experiment, 100,000 cells/well were seeded in 24 well plate. Cells were grown overnight to reach 70% confluency. On day 2, the STF were transiently transfected with components of the Norrin/β-catenin signalling pathway. A total of 400ng of DNA was transfected into each well containing the following: *FZD4* (60ng), *NDP* (60ng), *TSPAN12* (60ng), *LRP5* (100ng) and the Renilla control (pRL-TK) (1ng). Total concentration of 400ng was reached with addition of pDEST40 empty vector. To create a transfection master mix, the transfection reagent Fugene® 6 (Promega) (1.5 μl per well) was mixed with 50 μl Opti-MEM I media and incubated at RT for 5 minutes. The 400ng of DNA, comprised of the expression constructs detailed above, was added and the mixture vortexed, spin pulsed and incubated for 20 minutes at RT. The STF cells had a media change and the 50μl of the transfection master mix was slowly added. 48 hours after transfection, the medium was removed, and each well was washed with PBS twice. The cells were then passively lysed by adding 100 μl of 1x passive lysis buffer (PLB) from Promega Dual-Luciferase Reporter Assay System kit (Promega) and incubated at RT for 15 minutes with rocking. 10μl of this lysate was transferred into an opaque 96well plate (Grenier BioOne). Renilla and Firefly luciferase levels were determined using the Dual Luciferase reporter assay (Promega). The assay was run on a Mitras LB 940 Luminometer (Berthold Technologies) reader using the Microwin2000 program and dual injectors were used. The Firefly signal was normalised to the Renilla signal. The pathway activation levels were expressed as relative luciferase units

(RLU) which were determined by Firefly: Renilla ratio where Renilla served as an internal control. The pRL-TK contains the herpes simplex virus thymidine kinase (HSV-TK) promoter which provides low to moderate levels of Renilla luciferase expression in co-transfected mammalian cells.

2.27 Immuno-techniques

2.27.1 Antibodies

Antibody Name	Species raised in	Stock [mg/ml]	Western Blot dilution	Source
Polyclonal anti-LRP5	Rabbit	0.25	1/1000	Invitrogen
Monoclonal anti-β-actin	Mouse	35.2	1/5000	Sigma Aldrich
Monoclonal anti-6X His	Mouse	1	1/1000	Abcam

Table 2.2 Names, concentrations, dilutions and sources shown for all primary antibodies used.
Polyclonal anti-LRP5 = cat.no. 36-5400, Monoclonal anti-β-actin = cat.no. A2228, Monoclonal anti-6X His = cat.no. ab137839.

Antigen Name	Species raised in	Conjugate	Stock [mg/ml]	Western Blot dilution	Source
Rabbit Immunoglobulin	Goat	Horseradish peroxidase (HRP)	1	1/5000	Dako Cytomation
Mouse Immunoglobulin	Rabbit	Horseradish peroxidase (HRP)	1	1/5000	Dako Cytomation

Table 2.3 Names, concentrations, dilutions and sources shown for all secondary antibodies and their conjugates used.
Rabbit Immunoglobulin = cat.no. P0448, Mouse Immunoglobulin = cat.no. D0314.

2.27.2 Western Blotting

Protein lysates (Section 2.24) were thawed on ice and quantified using BCA assay (Section 2.25). 20μg of protein lysate was mixed with 1x NuPAGE® LDS Sample Buffer (ThermoFisher Scientific) and dH_2O to a final volume of 25μl. Samples were heated to 95°C for 10 minutes in a heating block before being centrifuged to make sure any residual liquid from the sides was returned to the bottom of the tube. Meanwhile, X-Cell SureLock electrophoresis tank (Invitrogen) was set up with a NuPAGE® 4-12% Bis-Tris gel (ThermoFisher Scientific). The gel was removed from the packaging and rinsed in H_2O and the white strip at the bottom of the cassette was removed. The gel was then placed and secured into the gel tank as instructed by the manufacturer. NuPAGE® 1x MES SDS running buffer (Novex® by Life Technologies) (400ml of 1x running buffer) was added to the gel tank to cover the gel. The comb from the gel was taken out and by using a Pasteur pipette or 1ml pipette any settled buffer from within the wells of the 4-12% Bis-Tris gel was removed. 25μl of sample was loaded alongside 10μl of Precision Plus Protein™ All Blue Standard (Bio-Rad). The gel was initially run at 120V for approximately 5 minutes until the samples have moved from the wells and then at 150V for 1 hour or 1 hour 30 minutes depending on the size of the expected protein. Invitrolon™ PVDF Filter paper sandwich with a 0.45μm pore size (Novex® by Life Technologies) was activated in methanol for 30 seconds before rinsing in dH_2O for 5 minutes on a rocker. It was then equilibrated in transfer buffer containing 10% methanol. Sponges were run under H_2O to remove any bubbles and placed into transfer buffer prior to assembling the transfer sandwich. The Gel tank was rinsed with H_2O and the gel removed. The gel

cassette was carefully opened, the wells cut off of the top of the gel with a scalpel and also the excess gel at the bottom. A transfer sandwich was assembled in the following manner: 3 sponges, pre-soaked in transfer buffer, were placed on the negative cathode (-) of the transfer module followed by pre-soaked blotting paper. The gel was slowly placed on top of the blotting paper. A PVDF membrane was placed on top of the gel followed by another pre-soaked blotting paper and a set of 3 sponges. The positive anode (+) was carefully placed on top of the sandwich. After each step, any residual bubbles were removed. The transfer module was held over the gel tank, the sandwich compressed and securely placed into the gel tank, before being filled with transfer buffer. Ice was placed into the gel tank surrounding the transfer module to reduce the risk of overheating. The transfer was run at 30V for 1 hour 30 minutes to 2 hours. Once the transfer was finished, the membrane was placed into 10% blocking solution (2.5g of Marvel milk and 25ml of PBS-T) for 30 minutes on a roller. Primary antibody was diluted 1/1000 in 5% milk/PBS-T (0.25g of Marvel milk and 5ml of PBS-T). The membrane was placed into the primary antibody solution and incubated at 4°C on a roller overnight.

The following day, the membrane was washed 5x with PBS-T for 5 minutes on a rocker. The secondary antibody, conjugated to HRP, was diluted 1/5000 in 5% milk/PBS-T, added to the membrane and incubated at RT for 1 hour. The secondary antibody solution was removed by washing 5x with PBS-T for 5 minutes on a rocker. To image the membrane, Supersignal® West Femto Kit or Supersignal® West Pico Chemiluminescent substrate (ThermoFisher Scientific) was loaded onto the membrane and incubated for 5 minutes. The

membrane was imaged using the Image Lab 6.0. The size standard marker was visualised first followed by the protein.

To blot for the loading control (β-actin), the membrane was stripped with Restore™ Western Blot Stripping buffer (ThermoFisher Scientific) and blocked in 10% blocking solution (2.5g of Marvel milk and 25ml of PBS-T) for 30 minutes on a roller. It was then incubated in primary antibody 1/5000 in 5% milk/PBS-T either at 4°C overnight or 1 hour at RT. The primary antibody solution was removed by washing with 5x PBS-T for 5 minutes on a rocker. A solution of 1/5000 secondary antibody, conjugated to HRP, in 5% milk/PBS-T was added to the membrane and incubated at RT for 1 hour. The secondary antibody solution was washed off with PBS-T 5 x 5 minutes on a rocker. The membrane was stained and imaged as described earlier.

2.28 Live cell imaging (LCI) using the Nikon Biostation IM

HEK293, HeLa and U2OS cells were seeded at 40,000 cells per quadrant in live cell imaging plates (Ibidi®). 24 hours after seeding, cells were transfected (Section 2.23.5). 24 hours post transfection, cells were placed in the BioStation IM (Nikon). The 20X objective of the inverted microscope was used to capture the images. Images were taken every 5 minutes over a 24 hours period. The exposure time was set to 2 seconds, the lamp to 100%, gain (sensitivity) to 3.97 and resolution to 800 x 600 binning. The resulting video and images were analysed in ImageJ.

2.29 Microscopy

All cell lines were visualised for cell viability and growth on an Olympus CKX41 bright field microscope with 4x and 10x objective lenses.

2.29.1 EVOS™ Cell imaging microscope

Cells were imaged using the EVOS™ microscope to visualise transient transfections when constructs were tagged with green fluorescent protein (GFP). A GFP filter cube with 470nm excitation and 525nm emission was used. The manufacturer's instructions were followed for correct visualisation.

3 Chapter: Identification of variants and candidate genes in Coats Disease

3.1 Introduction

The aim of the current study was to expand on the work performed by Black and co-workers (Section 1.5.4.3) by utilising NGS technology, particularly WES, to globally screen the exomes of Coats eyes to uncover new genes causing this phenotype. Seven FFPE Coats eyes were sectioned and an effective DNA extraction and WES library preparation protocols were derived. WES analysis has identified a potentially pathogenic mutation in *LRP5* gene, which was further assessed by functional assay TOPflash. Furthermore, a number of candidate genes and variants were identified in all patients.

3.2 Results

3.2.1 Histological analysis of enucleated FFPE Coats eyes

The basis of this experiment was a collection of enucleated Coats eyes obtained through a collaboration with Professor Alex Levin (Wills Eye Hospital and Thomas Jefferson University, Philadelphia, USA). The collection consisted of eyes from seven male unrelated unilateral Coats disease patients who were all diagnosed and enucleated by Drs Carol and Jerry Shields (Wills Eye Hospital, Philadelphia, USA) over a number of years. All eyes were FFPE for routine histological analysis as part of the patients' clinical care. The remaining

tissue (approximately half an eye globe per case) was sent to Leeds for genetic analysis under ethical approval obtained from the Institutional Review Board of Wills Eye Hospital (#10-024). None of the patients had undergone any genetic testing prior to this study.

In total, eight tissue blocks were analysed; one tissue block for each patient, except for one patient (ID 71208) where two blocks from the same eye were available (M1 and M2). The tissue blocks were sectioned, and H&E stained (Section 2.3) to show the full pathology of each eye and to highlight regions for DNA extraction. The majority of the eyes showed extensive regions of dysplastic retina in the vitreous chamber, apart from 67662ON which only displayed mild retinal dysmorphology (Figure 3.1).

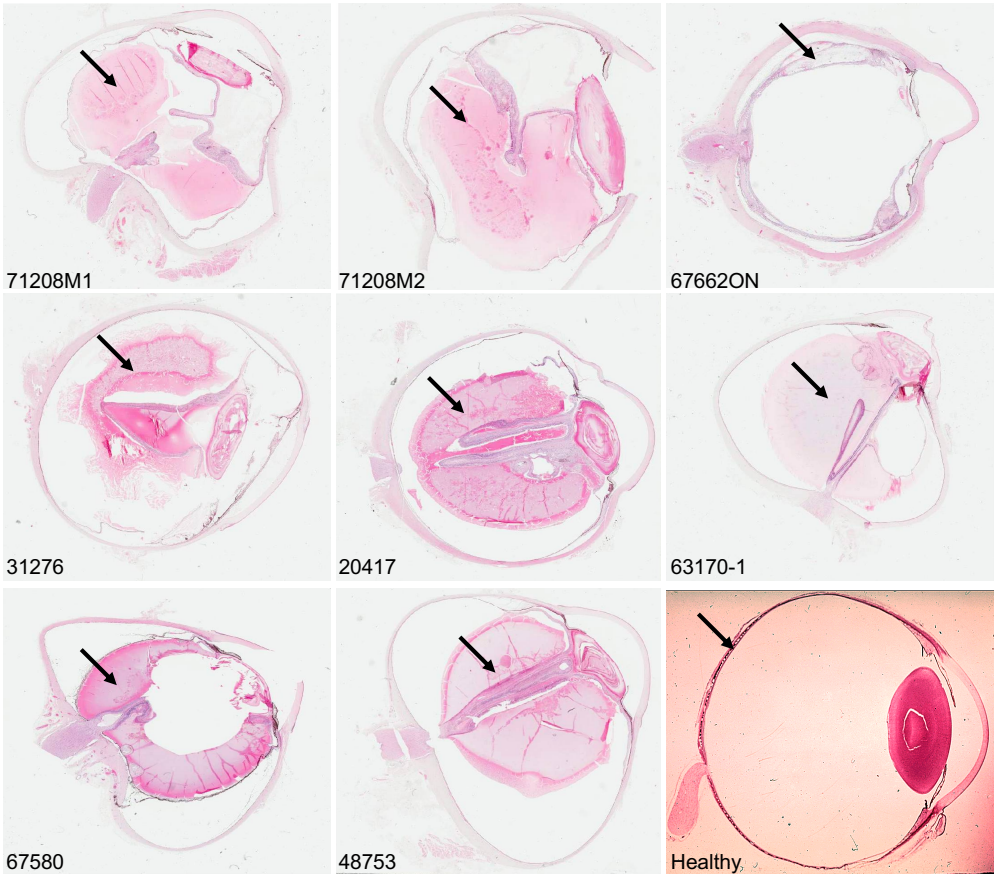


Figure 3.1 Histology of Coats eye.
The FFPE Coats eyes were sectioned and stained with H&E. All eyes are presented with their unique identifier. Sections 71208M1 and 71208M2 are from different blocks created from the same patient eye. A H&E section of a healthy eye is shown in the bottom right hand corner for comparison. Arrows indicate the dysplastic and detached retina accumulating within the vitreous space. Arrow in the 'Healthy' image indicates a normal retina with no detachment. Healthy eye figure provided by Dr Kamron Khan, St James's University Hospital, Leeds. Eyes were sectioned and stained by Mr Mike Shires, St James' University Hospital, Leeds.

3.2.2 Library preparation of DNA extracted from Coats eyes

Each tissue block was freshly sectioned and deparaffinised for DNA extraction (Section 2.3). Initially, attempts were made to macro dissect and extract the DNA from only the retinal tissue (including the dysplastic retinal tissue) and the lens. This was carried out to try and obtain DNA from two different embryonic sources to facilitate the detection of somatic versus germline mutations as lens is derived from surface ectoderm and retina is derived from neural ectoderm

(Section 1.3). However, this was not successful and, therefore, DNA was extracted from ~50 full sections. Only four to five 5-10µm thick sections were processed per extraction column as additional sections resulted in column clogging. The DNA was quantified using a fluorometer and the Qubit™ dsDNA BR and High Sensitivity Assay Kit (Section 2.6) but the concentrations were very low due to the large volume required to elute the DNA from the Qiagen column (Table 3.1). To increase the sample concentrations, the samples were vacuum concentrated (Section 2.7). The samples were re-quantified, and all had sufficient amounts of DNA at a suitable concentration (Table 3.1). The extraction of sample 67580 did not yield enough DNA and, therefore, the whole DNA extraction process was repeated.

Sample ID	DNA (ng/µl)	Vacuum concentrated DNA (ng/µl)
20417	8.38	15.32
48753	6.18	35.2
31276	6.72	43.6
71208M1	16.1	26.8
71208M2	12.56	117.6
67662ON	6.22	42.2
63170-1	14.4	82.6
67580	2.82 & 2.7	14.82 & 8.76

Table 3.1 Concentrations of DNA extracted from enucleated Coats eyes.
DNA concentrations calculated by Qubit analyser from processing of 50 5-10µm thick sections. Samples were vacuum concentrated to reduce the amount of diluent. Sample

67580 contains two values as the DNA extraction protocol was performed twice to achieve enough DNA.

The extracted DNA was used to prepare WES libraries for each of the eight tissue blocks. Briefly, the libraries were prepared using the NEBNext® Ultra DNA Library Prep Kit and SureSelect^{XT} v5 kit using an in-house protocol optimised for FFPE samples with low DNA input created by Laura Crinnion (Leeds Next Generation Sequencing Facility) (Section 2.16) (Table 3.2).

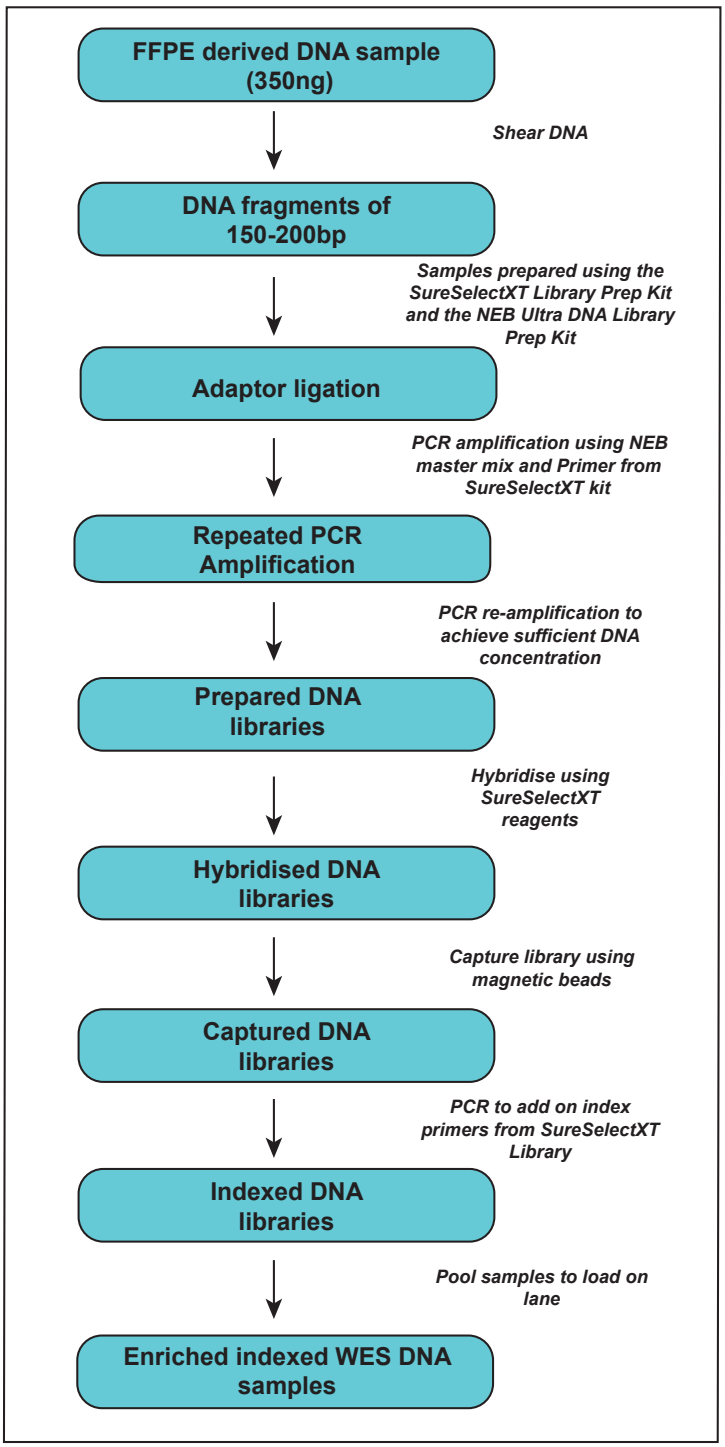


Figure 3.2 Outline of WES library preparation protocol.
Flow diagram of the Modified Protocol for the Preparation of Pre-Capture Libraries for SureSelect^{XT} Target Enrichment System for Paired End Sequencing using the SureSelect^{XT} kit and NEB Ultra DNA Library Prep Kit.

During the library preparation, DNA samples were analysed on a Bioanalyser to assess the concentration and size of the raw FFPE double stranded DNA

before and after the samples were sheared on a Covaris S2 system. The results for unsheared DNA for patient 71208M2 showed an undetectable peak representing the fragment sizes at 150-300bp (Figure 3.3). Results for the rest of the samples (Appendix 7.1) showed a similar trend to the results presented in Figure 3.3. In addition, some samples masked the ladder, so no definite size or quantity of DNA fragments could be determined. This was due to the presence of large quantities of low molecular weight DNA in the samples. The output of the results post shearing showed no significant peak at size of 150-200bp as expected, presumably because the shearing had reduced the fragmented samples to negligible sizes (Figure 3.3). Nevertheless, the library protocol was continued due to the precious nature of the samples.

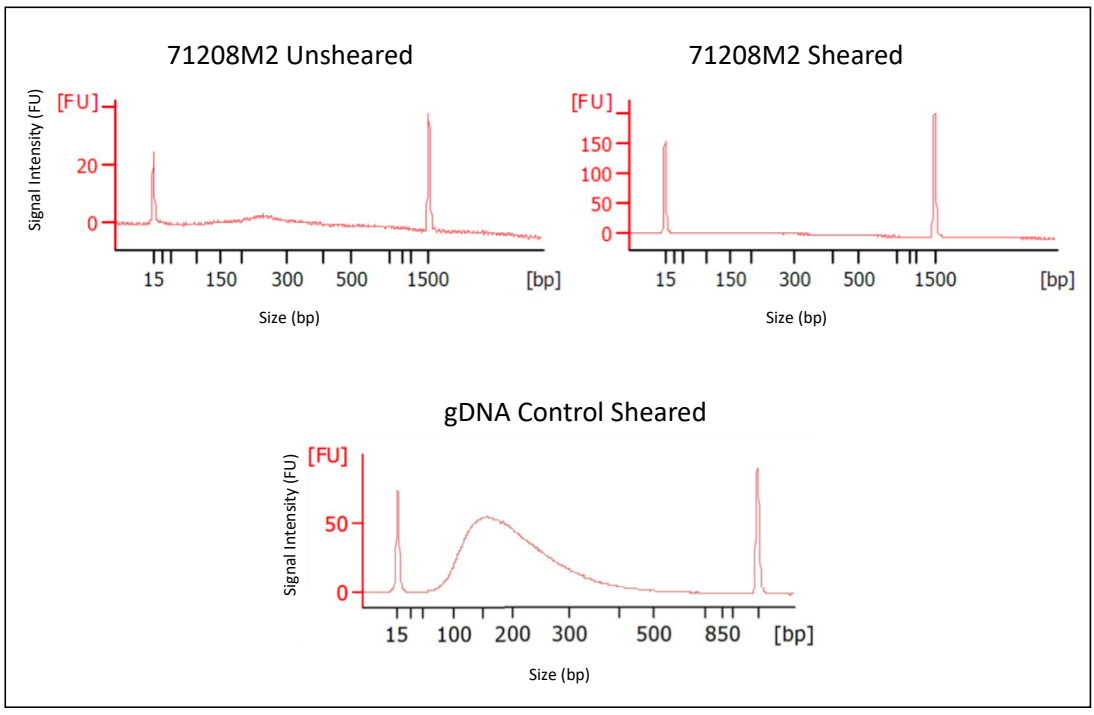


Figure 3.3 QC analysis of the extracted FFPE DNA before and after shearing. An example of FFPE extracted DNA run on the Bioanalyser using the High Sensitivity DNA kit to assess the size of DNA fragments before and after shearing. X-axis shows the size in base pairs (bp) and the y-axis represents the signal intensity of detected DNA. The y-axis differs for the samples as they were not run on the same bioanalyser chip.

Following the PCR amplification of the adapter ligated DNA, the samples were reassessed on the Bioanalyser and fragments of ~200bp were detected (Figure 3.4). Samples 48753 and 20417 required reamplification as there wasn't enough DNA within the samples after the first PCR amplification.

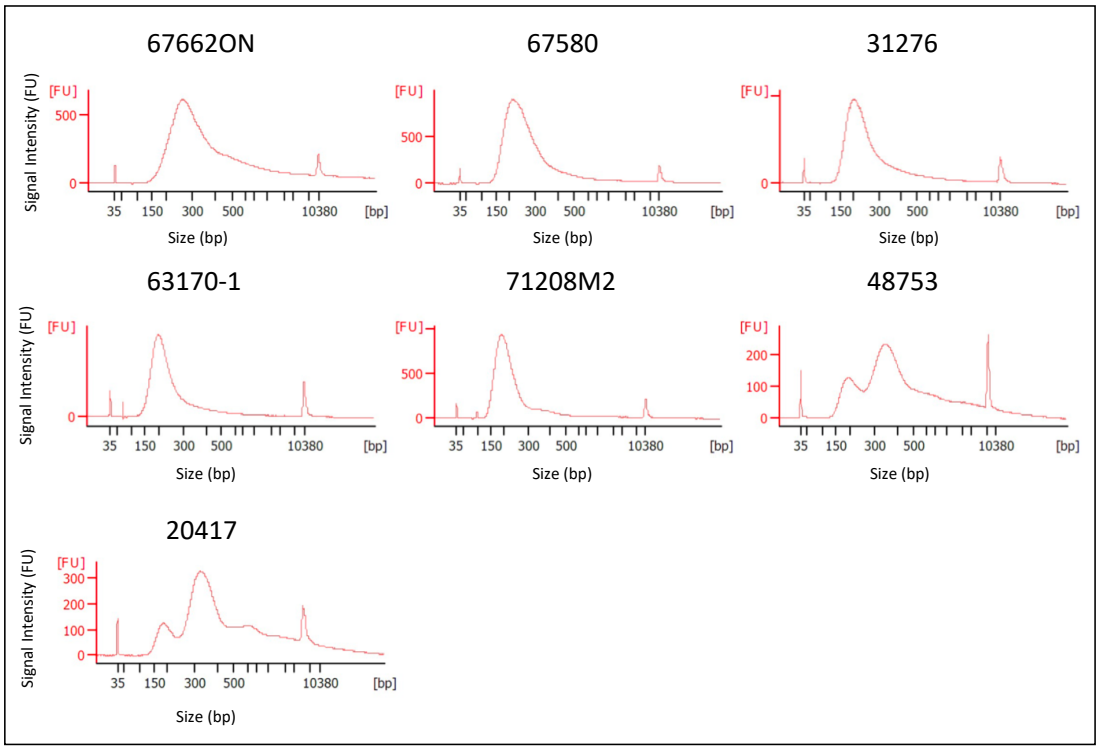


Figure 3.4 QC analysis of the extracted FFPE DNA post PCR amplification. DNA was assessed after PCR amplification using the Bioanalyser High Sensitivity kit to check that the required peak at ~200bp was present for all samples. Samples 20417 and 48753 underwent a second round of amplification to achieve the required concentrations as not enough DNA was detectable after only one amplification. This figure shows 7 samples as only the DNA extracted from 71208M2 was taken forward for library preparation. This was due to higher DNA concentration compared to block M1 from the same eye.

All seven samples were pooled and sequenced as 100bp paired end reads on one lane of an Illumina HiSeq2500 by the University of Leeds Next Generation Sequencing Facility (http://dna.leeds.ac.uk/genomics/).

3.2.3 Bioinformatics analysis of WES data

The bioinformatics analysis was performed to prioritise variants of interest based on their predicted functionality and zygosity (homozygous, heterozygous and compound heterozygous).

Raw fastq files generated by the sequencer were first analysed to assess for data quality. QC data was generated to assess the quality scores across all bases and the quality score distribution over all sequences, base sequence content, GC content, sequence length distribution and number of duplicate sequences. Examples of three of the QC results generated for sample 71208M2 are shown in Figure 3.5. The first result showed the quality scores across all bases (Figure 3.5A). All the base sequence calls generated were quality assessed based on the Phred scoring system. Phred scores above 30 mean that the base call is 99.9% accurate. The average Phred score for sample 71208M2 was >30 showing that the data generated was of excellent quality. The quality of the raw sequencing data was checked for each sample and these are presented in Appendix 7.2. The results showed that all the samples passed the QC as the average Phred scores were >30. The second QC result showed the average quality score per read for all sequencing reads generated per sample (Figure 3.5B). This output again uses the Phred scoring system and shows the number of reads and their Phred score. The majority of the reads had a mean sequence quality score >30 showing that the data was of high quality. The last QC result presented here is the sequence duplication level. This QC was used to establish the percentage of sequences that were duplicated during the PCR amplification steps. This QC is important as the PCR step in the first run of WES had to be repeated for a number of samples.

62.6% of all sequences for sample 71208M1 were shown to be duplicated 2 to 10+ times. The sequence duplication level for all other samples are presented in Appendix 7.2. All patients were deemed to have passed the QC.

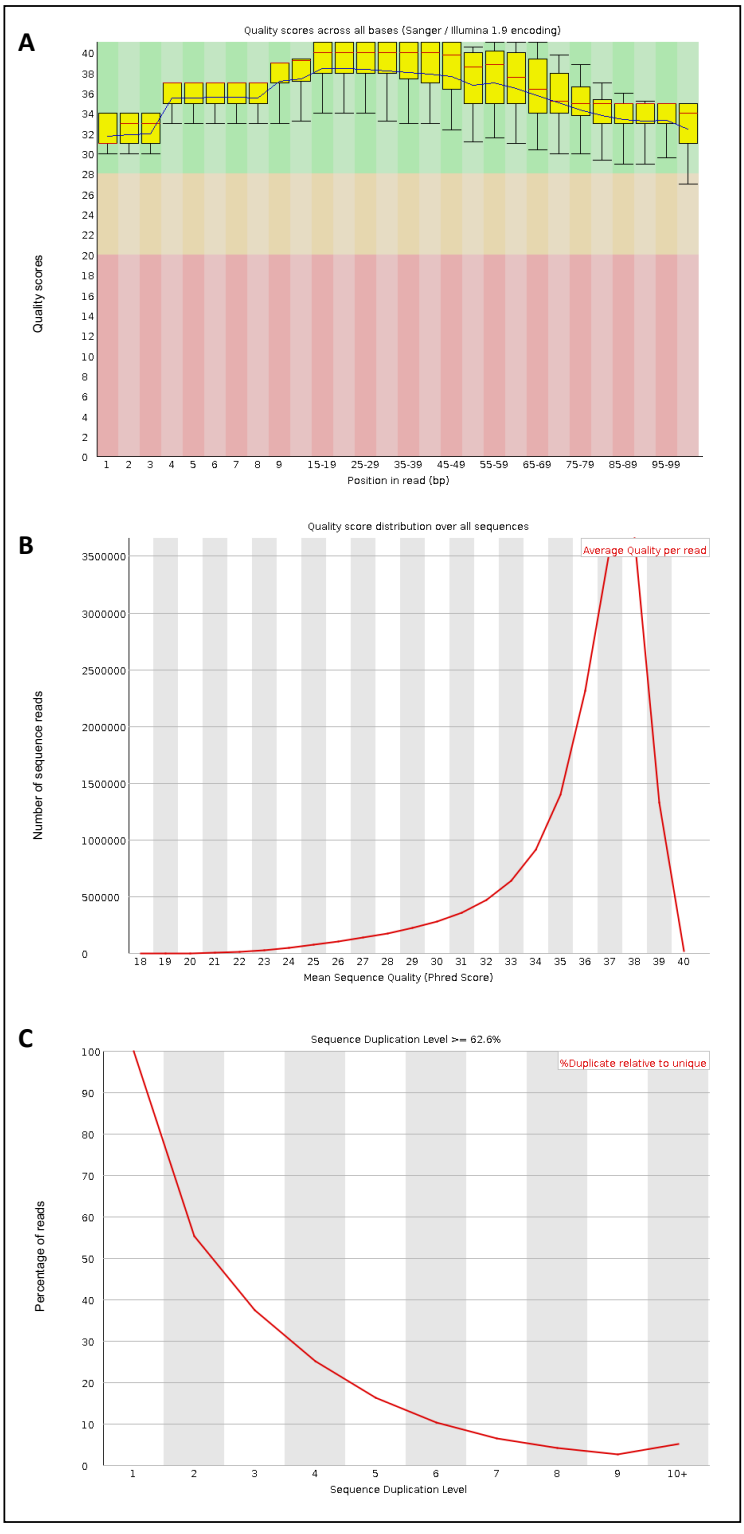


Figure 3.5 QC for WES data for patient ID 71208M2.
(A) A representative of the QC of the raw fastq file for patient 71208M2 using a BoxWhisker plot. The plot shows the quality scores across all 100 bases for each read generated by the Illumina HiSeq 2500. The x-axis represents the position in read in bp and the y-axis shows the quality Phred scores. The scores below 20 represent poor quality calls, scores between 20-28 represent calls of acceptable quality and the scores above 30 show very high-quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality.

(B) Representative example showing the quality score per read distribution over sequences of the fastq file for sample 71208M2. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads. **(C)** Representative example of sequence duplication level for sample 71208M2. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results showed that 62.6% of all sequences were duplicated between 2 to 10+ times. 37.4% of all sequences were only read once.

After carrying out the QC, the bioinformatics analysis was continued (Section 2.18). In summary, the raw fastq files were trimmed to remove the adapters and the trimmed files were aligned to human reference genome GRCh37 (build hg19) using BWA to generate the SAM files. After the alignment, the SAM file was converted to a BAM file.

At this point, another quality check that was essential for the aligned BAM file was the determination of the coverage of each base within the exome and this was performed by a tool called Depth of Coverage (Section 2.18). This analysis was performed after the removal of PCR duplicates using Picard and it provides the user with coverage details for each chromosomal location. The depth of coverage data was calculated for each sample (Table 3.2).

The overall depth of coverage varied between samples. Samples 31276, 67662ON, 63170-1 and 71208M2 showed high percentage of bases covered above 5 reads. The coverage for the same samples above 20 reads was also high. However, samples 20417, 48753, 67580 showed very low coverage above 20 reads but high coverage above 5 reads. This was suitable to proceed with the analysis, but it became evident that the WES needed to be repeated to improve upon the read depth as the FFPE DNA contains high levels of errors and artefacts (Section 3.3.6).

Sample ID	% of bases above 5	% of bases above 10	% of bases above 20
20417	69.1	35.7	6.1
31276	93.4	87.5	72.5
48753	68.1	34.9	7.0
67580	68.5	35.2	6.9
67662ON	99.2	97.7	92.9
63170-1	98.7	95.9	84.7
71208M2	92.6	79.8	45.8

Table 3.2 Depth of coverage for Coats samples WES data.
Sample IDs are provided and corresponding percentage of bases covered above certain read depths are listed.

After the removal of duplicates, indel realignment targets, base recalibration and variant calling were carried out using GATK. Variants were called to generate the gVCF file and then final VCF file using Haplotype Caller was generated. Since the gVCF contains extra information, it was possible to carry out joint analysis of multiple samples. gVCFs for all patients were merged into one joint gVCF file. All variants (SNPs and indels) were called and the variants were filtered using publicly available databases of variants, such as dbSNP142, ExAC and gnomAD (accessed 10 August 2016), to remove any variants that have MAFs of ≥1% and further ≥0.01%. After the removal of common polymorphisms, filtered VCFs were further subjected to VEP to determine the functional effect of all the variants, ie. missense, nonsense, deletion or an insertion. The variants were assessed for their pathogenicity by

PolyPhen, SIFT and CADD scores (Sections 2.19.4.1, 2.19.4.2, 2.19.4.4). Lists were then filtered using in-house perl scripts such as “getFunctionalVariants.pl”, “getHetVariants.pl” and “findBiallelic.pl” commands to select for variants based on their effect on the coding amino acid sequence or their inheritance pattern. The resulting lists were then ranked on CADD scores and annotated using the Ensembl GeneAnnotator from vcfhacks. Subsequently, VCFs were converted into Excel files.

The focus was on protein coding variants, such as frameshifts, stop gain/loss, exon/intron variants and missense variants, as they would have a potential impact on the translation of the protein, its structure and folding. Heterozygous and compound heterozygous variants were of high importance as the hypothesis for this study was that Coats disease is caused by somatic mutations and the chance of the same variant spontaneously occurring on both alleles at the exact same location was highly unlikely, although large heterozygous deletions in trans with point mutations can mimic homozygosity.

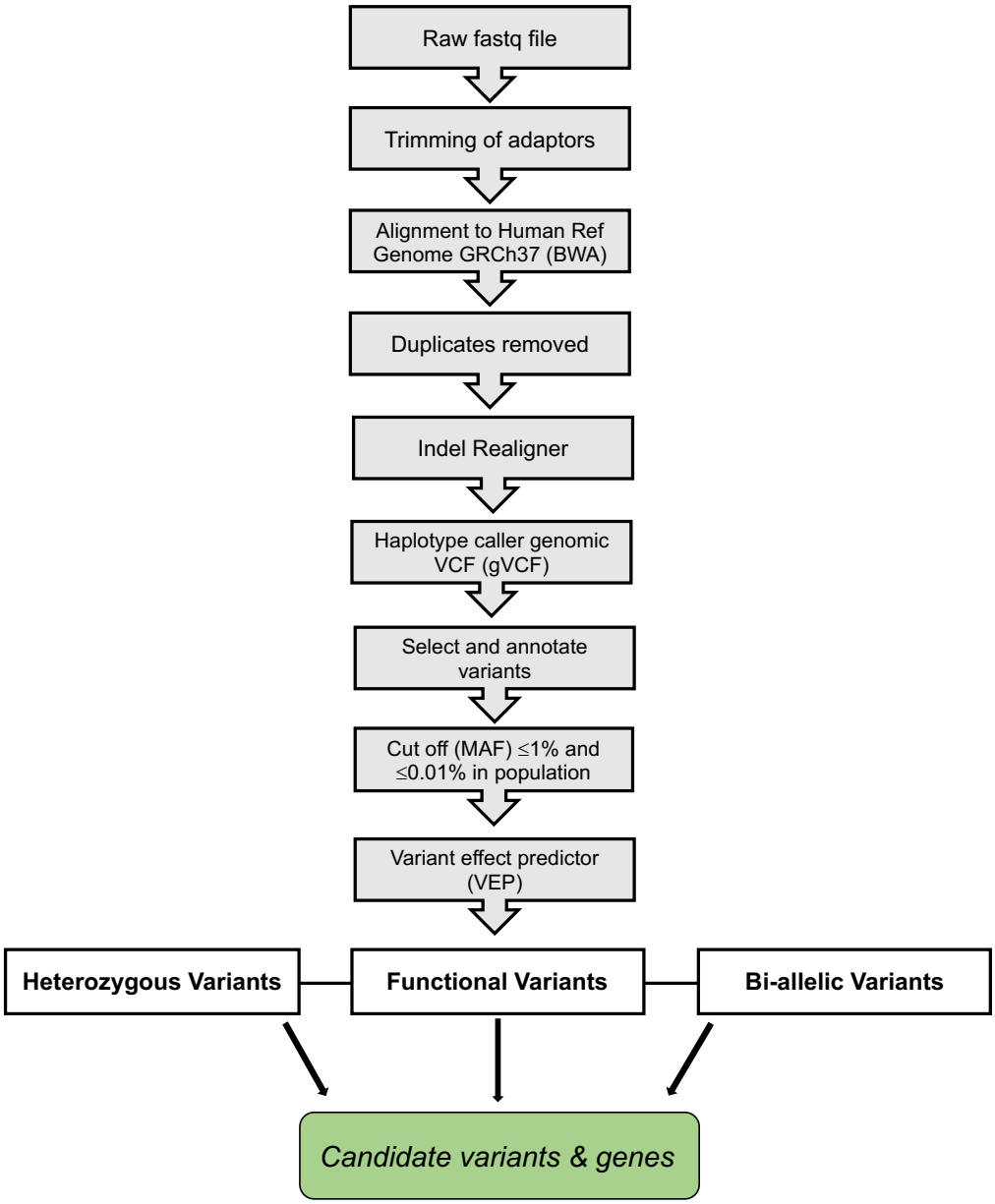


Figure 3.6 Schematic representation of the bioinformatics pipeline.
Flowchart describing the analysis pipeline workflow that was used to analyse WES of all Coats disease patients with the aim to uncover new variants and genes.

3.2.3.1 WES analysis excludes *NDP* mutations from all Coats eyes

The first pass of the WES data generated huge lists of variants and most of these had very low allele depths. Table 3.3 shows how many variants and

genes were generated per sample using each of the described commands to identify functional, heterozygous and bi-allelic variants.

Sample ID	GetFunctional	GetHeterozygous	FindBiallelic
20417	13,091	32,784 (13,067)	7,077 (1,615)
31276	10,913	55,041 (10,833)	6,800 (2,308)
48753	13,012	32,117 (12,999)	6,995 (1,892)
67580	1,550	14,751 (6,161)	2,235 (613)
67662ON	6,182	6,113 (1,480)	882 (467)
63170-1	2,098	8,925 (2,020)	857 (393)
71208M2	3,622	11,235 (3,566)	1,206 (399)

Table 3.3 Number of variants present from WES data analysis using vcfhacks filters.
Number of variants for each patient using the GetFunctional.pl command including all the functional variants such as missense, frameshift, nonsense and splice variants. GetHeterozygous.pl command identifying the overall number of variants in heterozygous state. Number of all identified bi-allelic variants per patient and the number of functional bi-allelic variants using the FindBiallelic.pl command in brackets.

To be able to analyse WES data from FFPE-extracted DNA, the depth of coverage needs to be good to be able to differentiate true variants from FFPE-derived artefacts. Unfortunately, the data generated was not sufficient to fully analyse the Coats eyes in a meaningful way and it became evident that the DNA extraction and WES procedure needed repeating to achieve better coverage. However, the data was amenable for candidate gene analysis. Therefore, the analysis of the WES data focused on only looking at genes

involved in the Norrin-β-catenin signalling pathway or in the pathogenesis of FEVR (Section 1.8).

This analysis first focused on inspecting *NDP* using IGV, as this was the only gene known to be implicated in Coats disease at this point in the study. Although the coverage over the coding exons of *NDP* was not very high, it was sufficient to exclude *NDP* variants from all cases. A read depth of 5x was deemed to be enough to exclude variants as *NDP* is on the X chromosome and all of the eyes in this study were from males. *NDP* was suitably covered in all of the samples indicating that the Coats eyes harboured mutations in new genes.

The analysis then focused on looking at the remaining genes known to be mutated in FEVR: *FZD4*, *LRP5*, *TSPAN12*, *KIF11*, *ATOH7* and *ZNF408*. The genes were filtered out using a script which specifically pulled out variants based on their chromosomal location within the reference genome (Appendix 7.5). This analysis was carried out after the removal of PCR duplicates and creating of the vcf file but before any filtering took place, such as selecting for only homozygous or heterozygous variants. The MAF used for this filtering was 0.01 (<1% in population). Low quality variants were removed and variants with high CADD scores were selected and are presented in Table 3.4.

Patient ID	Gene ID	cDNA position	Protein position	Transcript ID	Allele Depth (WT/Mut)	CADD score	ExAC	Allele frequency
71208M2	*LRP5*	c.2951A>G	p.(Tyr984Cys)	ENST00000294304.7	20/19	29.2	2/121,146	1.651e-05
48753	*ZNF408*	c.1934C>T	p.(Ser645Phe)	ENST00000311764.2	4/2	25.6	1/119124*	8.395e-06
48753	*ZNF408*	c.1925C>T	p.(Ser642Phe)	ENST00000311764.2	2/2	23.8	1/119,112	8.395e-06
20417	*LRP5*	c.2827C>T	p.(Pro943Ser)	ENST00000294304.7	9/5	22.4	8/107,644	7.432e-05
20417 63170-1	*TSPAN12*	c.170T>C	p.(Leu57Ser)	ENST00000441017.1	3/4 41/4	22.3	1/120,676*	8.287e-06
31276	*FZD4*	c.100G>A	p.(Ala34Thr)	ENST00000531380.1	41/8	21.5	6/112,998*	5.31e-05
31276	*FZD4*	c.81G>C	p.(Leu27Phe)	ENST00000531380.1	32/6	21.3	1/113,532* or 29/112,910*	8.808e-06 or 2.568e-04

Table 3.4 Variants present in known FEVR genes with a CADD score >20 from the first run of WES.
Variants in known FEVR genes are presented with cDNA and protein position. CADD, variant quality and patient numbers are displayed. Low quality variants and low allele depth variants were removed. * = number of alleles for the nearest SNP.

This analysis revealed a convincing missense heterozygous variant in exon 13 of *LRP5* (NM_002335.3) in sample 71208M2, GRCh37, Chr 11:68183919 , c.2951A>G p.(Tyr984Cys). The read depth presented as wild type base A=20 and the alternative (mutant) base G=19. This variant is described further in the following Section 3.2.4.

None of the other variants highlighted in Table 3.4 were taken forward. This was because they either had a low depth of coverage or because the proportion of variant reads was low compared to the WT allele. Given the high number of artefacts generated in FFPE-extracted DNA, this data could not be trusted. Similarly, the depth of coverage over the autosomal FEVR genes was not sufficient to convincingly exclude the presence of heterozygous variants in these genes. Therefore, no further analysis was undertaken on this WES data and the DNA extraction and WES experiment was repeated to get better quality data (Section 3.2.7).

3.2.4 Analysis of the *LRP5* variant c.2951A>G, p.(Tyr984Cys) identified in 71208M2

The c.2951A>G p.(Tyr984Cys) variant identified in 71208M2 by WES was visualised in the BAM file using IGV (Figure 3.7).

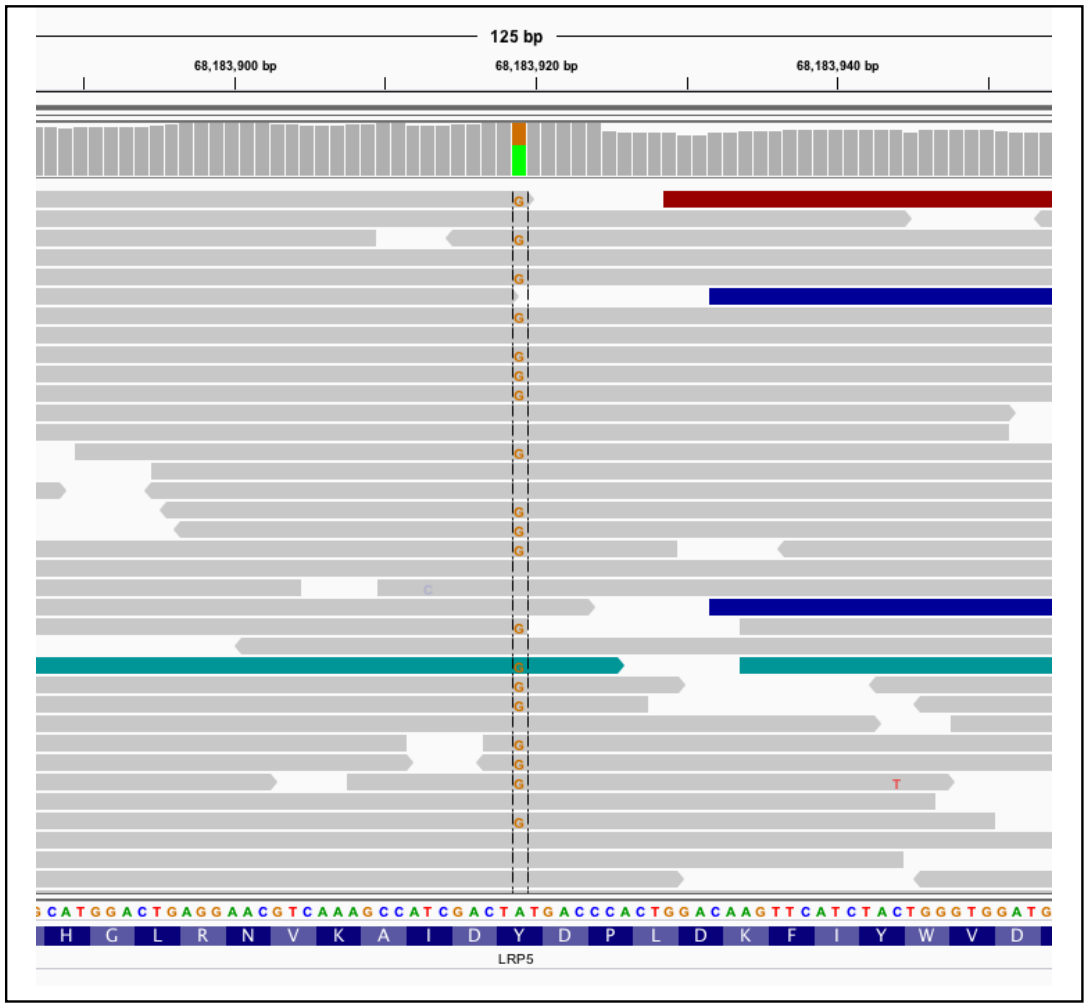


Figure 3.7 IGV snapshot of *LRP5* variant for patient 71208M2.
IGV screenshot of the *LRP5* c.2951A>G, p.(Tyr984Cys) variant. Orange colored Gs represent the reads with the alternative (mutant) base. Wild type base (A) is shown in green in the reference sequence below. Grey bars in both images represent reads above the zoomed in region of *LRP5*.

The presence of this variant in various online databases was investigated to determine its frequency. The databases included EVS (accessed 01/11/2017), ExAC (accessed 01/11/2017) and gnomAD (accessed 01/11/2017). The c.2951A>G p.(Tyr984Cys) variant was reported in a heterozygous state in 2 individuals in the ExAC database (1 African and 1 Latino) and in gnomAD in 5 individuals (2 African, 2 European Non-Finnish and 1 Latino) (Table 3.5 and Table 3.6). The variant was not reported in the EVS database and it was not reported in a homozygous state in any of the databases.

EVS	ExAC	gnomAD
Not reported (0/6,503)	2 heterozygous (2/121,146)	5 heterozygous (5/246,238)

Table 3.5 Frequency of c.2951A>G p.(Tyr984Cys) variant in variant databases. EVS (http://evs.gs.washington.edu/EVS/) does not show this variant in its database. It has been reported in 5 out of 246,238 alleles in the gnomAD database (http://gnomad.broadinstitute.org). This variant was reported in the ExAC database (http://exac.broadinstitute.org) in 2 alleles out of 121,146 alleles.

The detailed population frequencies were available within the gnomAD database and the different ethnicities and their frequencies are provided in Table 3.6.

Population	Allele Count	Allele Number	No. of Homozygotes	Allele Frequency
African	2	15302	0	0.0001307
Latino	1	33582	0	0.00002978
European (Non-Finnish)	2	111700	0	0.00001791
Ashkenazi Jewish	0	9850	0	0.000
East Asian	0	17246	0	0.000
European (Finnish)	0	22294	0	0.000
Other	0	5484	0	0.000
South Asian	0	30780	0	0.000
Total	**5**	**246238**	**0**	**0.00002031**

Table 3.6 Population frequencies generated by gnomAD database for the *LRP5* c.2951A>G, p.(Tyr984Cys).
The table shows the frequency for this variant by population. Number of heterozygotes (allele count) and number of homozygotes are shown. Number of alleles this variant was screened for is clearly indicated along with the calculated frequencies for each population. The *LRP5* variant is present but rare in African, Latino and European (Non-Finnish) population.

The *LRP5* variant was then subjected to a number of prediction tools to assess its pathogenicity (Section 2.19.4). Table 3.7 shows the pathogenicity scores of the *LRP5* variant from six pathogenicity prediction software tools. All six tools predicted the *LRP5* variant to be deleterious. CADD ranked the variant to be in the top 1% of deleterious variants with a score of 29.2. PolyPhen2 predicted the variant to be probably damaging with the value of 0.989, very close to the cut off value of 1 which predicts a deleterious variant. SIFT and PROVEAN

scored the variant deleterious. Similarly, the Blosum matrix value of -2 predicts this variant to be damaging and MutPred's value of 0.488 showed the variant to be near the cut off value of deleteriousness of ≥0.50.

CADD*	PolyPhen2	SIFT	Blosum62*	PROVEAN	MutPred*
29.2	Probably damaging (0.989)	Deleterious (0)	-2	Deleterious (-7.053)	0.488

Table 3.7 Pathogenicity software prediction scores for *LRP5* c.2951A>G p.(Tyr984Cys).
URLs: CADD, (Kircher, 2014); PolyPhen-2, (Adzhubei et al., 2010); SIFT, (Ng and Henikoff, 2001); Blosum62, https://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt (Henikoff and Henikoff, 1992); PROVEAN, http://provean.jcvi.org/index.php (Choi et al., 2012); MutPred, http://mutpred.mutdb.org (Li et al., 2009). The outputs of CADD, PolyPhen2 and SIFT were included in the output of the WES analysis. *Blosum62 scores range between +3 to -3 and negative scores are more likely to be damaging. MutPred score is damaging if ≥ 0.50. CADD score ranges from 1 to 99, based on the rank of each variant with the top 10% of deleterious variants having score 10 and above, the top 1% scores of 20 and above and top 0.1% having scores 30 and above.

The conservation of the p.Tyr984 residue across a number of species was investigated using HomoloGene from NCBI (Section 2.19.8). This analysis was performed to establish the importance of the affected residue across a number of species within the LRP5 protein. The amino acid is highly conserved in the majority of species down to frog, with only zebrafish differing at the p.Tyr984 residue (Figure 3.8).

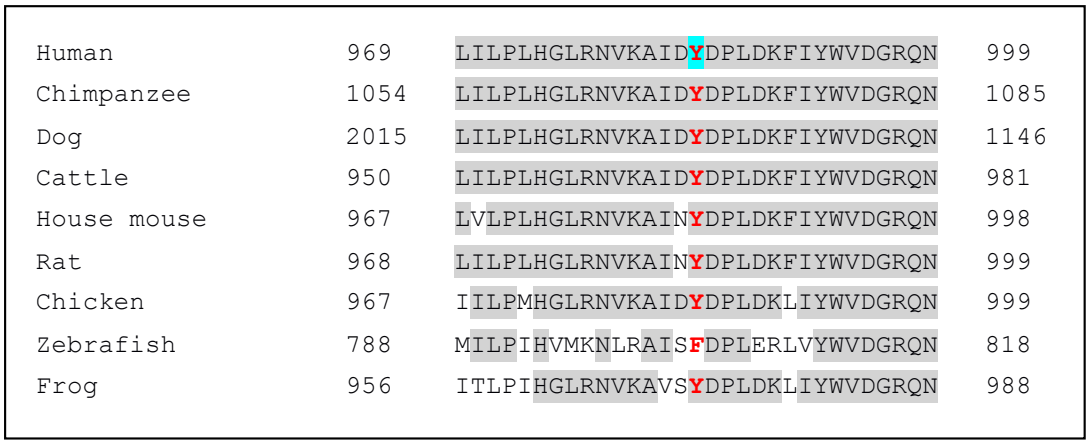


Figure 3.8 Protein sequence alignment of human LRP5 with its orthologues for the p.(Tyr984Cys) variant.
Alignments were calculated with HomoloGene (NCBI). Human NP_002326 (Homo sapiens); Chimpanzee XP_508605 (Pan troglodytes); Dog XP_003432463 (Canis lupus familiaris); Cattle XP_002699451 (Bos Taurus); House mouse NP_032539 (Mus musculus); Rat NP_001099791 (Rattus norvegicus); Chicken NP_001012915 (Gallus gallus); Zebrafish NP_001170929 (Danio rerio); Frog XP_002941689 (Xenopus tropicalis). 15 amino acids on either side of the variant are shown. The mutated amino acid residue is highlighted in cyan and in red in the orthologues.

To confirm that this variant was not an artefact, FFPE extracted DNA was PCR amplified using primers designed to amplify *LRP5* exon 13. Due to the fragmented nature of the DNA, the PCR primers were designed to amplify only 120bp surrounding the variant (Appendix 7.4). Initial attempts failed to amplify a PCR product, so WGA was performed on the extracted DNA (Section 2.12). This time, the PCR was successful, and the resulting PCR product was purified (Section 2.11.1) and used as the template for Sanger sequencing using the same primers as those used for the PCR. This analysis confirmed the heterozygous presence of c.2951A>G in patient 71208M2 (Figure 3.9).

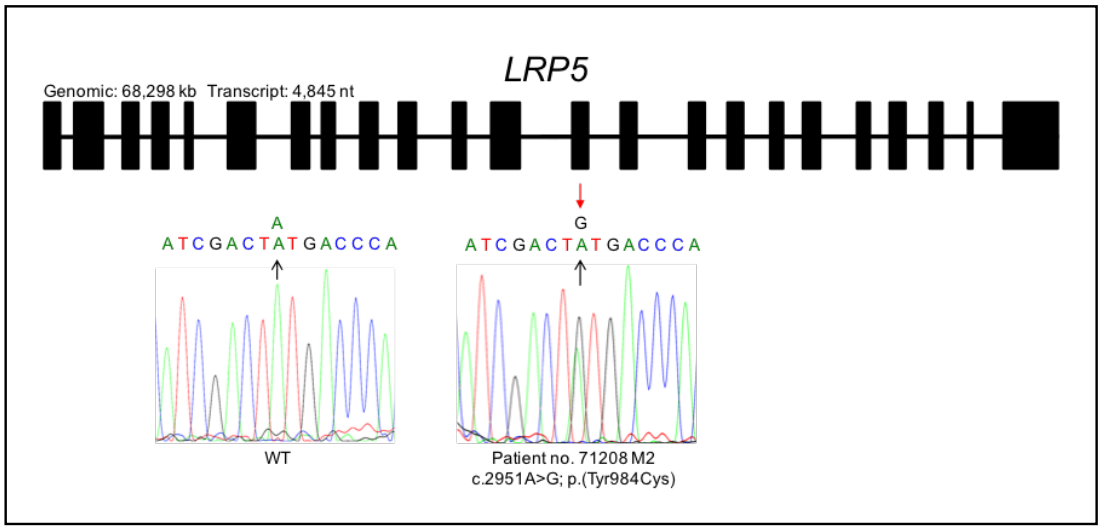


Figure 3.9 *LRP5* mutation identified in Coats patient 71208M2.
Schematic representation of the *LRP5* genomic structure. Location and sequence traces of the *LRP5* mutation c.2951A>G in exon 13 identified and the corresponding wild-type (WT) allele. The gene is presented in 5' to 3' orientation on the positive strand.

The variant c.2951A>G, p.(Tyr984Cys) was shown to be pathogenic by a number of prediction tools and was confirmed within the somatic DNA from patient tissue by Sanger sequencing. This variant was, therefore, taken forward to be assessed for its pathogenicity in functional assays (Section 3.2.6).

3.2.5 Clinical phenotype for patient ID 71208M2

This genetic finding from case 71208M2 was reported back to the recruiting clinicians; Professor Alex V. Levin, MD, MHSc (Wills Eye Hospital, Philadelphia, Pennsylvania, USA & Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania, USA). Drs J. A. Shields and C. L. Shields (Wills Eye Hospital, Philadelphia, Pennsylvania, USA). They provided the following clinical information:

"The case was of a 16-year old boy diagnosed with amblyopia along with depressive disorder and attention deficit disorder. There was no family ocular

history reported. At 16 year of age, retinal examination of the right eye showed peripheral telangiectasias of the retinal vessels extending from 7:30 to 12:00 o'clock with exudative retinal detachment to the foveal region with large areas of subretinal exudate (Figure 3.10). Optical coherence tomography (OCT) showed cystoid macular edema and some degree of macular traction within the same eye (data not shown). IVFA showed peripheral telangiectasias with typical "light bulb" configuration, leakage and avascular retina temporally and superiotemporally. Fundus examination, IVFA and OCT of the left eye were normal. The eye was treated by cryotherapy and intravitreal bevacizumab. Six months later the patient was shown to have neovascularization of the iris and a total retinal detachment with neovascular glaucoma. Enucleation was performed. Microscopic examination of the enucleated right eye disclosed florid iris neovascularization and early secondary angle closure. The retina was completely detached, had abnormal vessels and was focally thickened by exudates. The findings were consistent with Coats disease (Figure 3.10)."

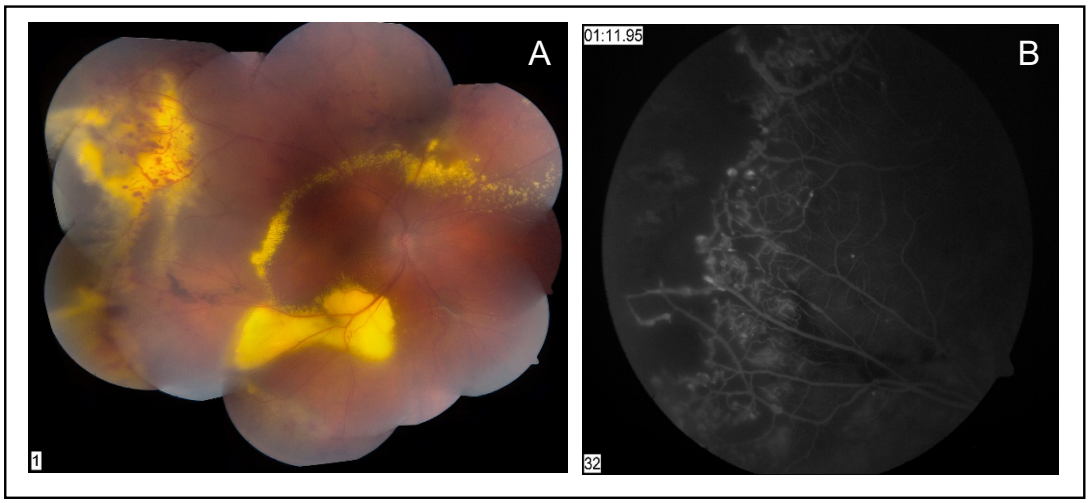


Figure 3.10 Clinical images of affected eye from Coats patient 71208M1/M2.
A: Fundus photography of the right eye showing peripheral telangiectasias of the retinal vessels extending from 7:30 to 12:00 o'clock with exudative retinal detachment to the foveal region and large areas of subretinal exudates. **B**: Intravenous fluorescein angiography (IVFA) showing peripheral telangiectasias, deep retinal vascular leakage and peripheral nonperfusion.

3.2.6 Functional analysis of LRP5 p.(Tyr984Cys)

The aim of this experiment was to functionally assess the pathogenicity of the *LRP5* missense variant p.(Tyr984Cys) by performing functional assays which included the TOPflash assay (Section 3.2.6.3) and LCI (Section 3.2.6.4).

The schematic of the protein with the location of the variant is presented in Figure 3.11).

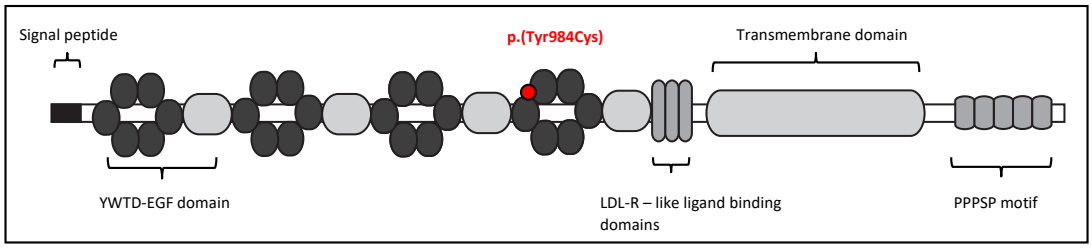


Figure 3.11 Schematic of the LRP5 protein structure.
Individual domains are labelled. Red circle denotes the p.(Tyr984Cys) missense variant found in patient 71208M2. The LRP5 structure was adapted from SMART tool (http://smart.embl-heidelberg.de) and (He et al., 2004). The YWTD domain – tyrosine (Y), tryptophan (W), threonine (T) and aspartic acid (D) β-propeller domain; EGF domain – epidermal growth factor; LDL-R-like domain – low-density-lipoprotein receptor–like (LDL-R–like) ligand-binding; PPPSP motif – proline (P), serine (S).

3.2.6.1 Creation of LRP5 expression constructs using Gateway® cloning technology

The full length pENTR_LRP5 was created by Soriano (2017). pDEST40_LRP5 and pDEST47_LRP5 expression constructs were generated in this study using Gateway® cloning technology (Figure 3.12) (Section 2.21.4). pcDNA™-pDEST40 contains a C-terminal V5 epitope tag followed by 6x-His tag and pcDNA™-pDEST47 contains a C-terminal green fluorescent protein (GFP) tag containing a CMV promoter for mammalian expression and SV40 promoter for the bacterial expression (maps of all expression constructs are presented in

Appendix 7.12). The tags included in both expression vectors aided in the detection of the fusion protein.

The entry vector was streak purified from a stock and propagated by transformation. The DNA was subsequently isolated, and sequence verified. The entry vector pENTR_LRP5 was further subjected to SDM (Section 2.21.2) (primers for SDM are in Appendix 7.11) to introduce the *LRP5* variant c.2951A>G p.(Tyr984Cys). The pENTR_LRP5 WT and p.(Tyr984Cys) insert was then transferred into the pcDNA™-DEST40 and 47 vectors using the Gateway® LR Clonase II (Figure 3.12). The expression constructs were sequenced to confirm the *LRP5* change in the cDNA and to exclude any other variants that may have been introduced during SDM (primers to sequence the constructs are in Appendix 7.10). Endotoxin-free Maxipreps were generated for the clones ready to use in mammalian cells (Section 2.21.6.2).

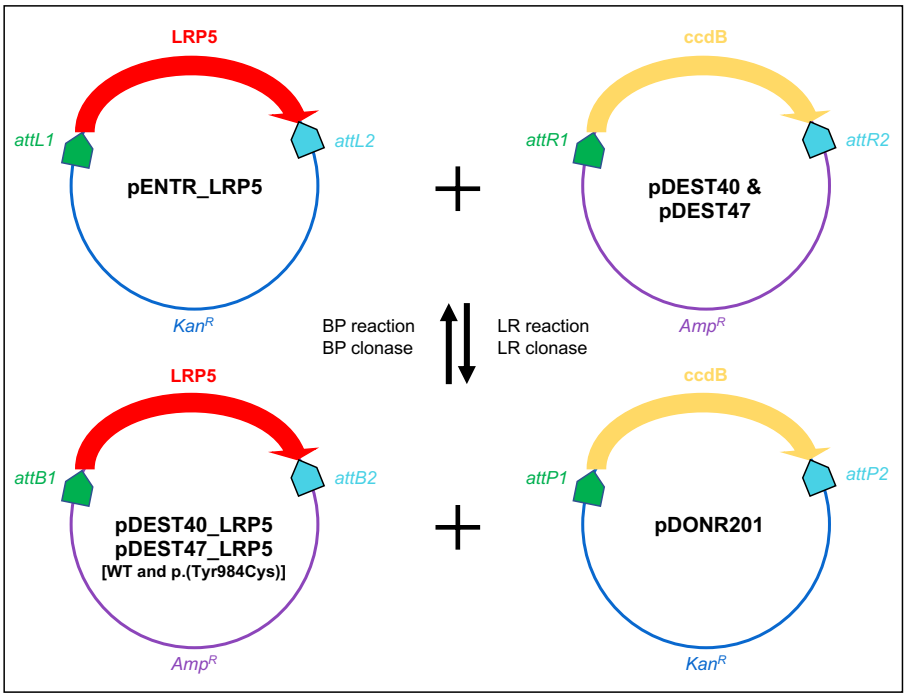


Figure 3.12 Schematic of the workflow of Gateway® Technology to create constructs used in this study.
The schematic outlines the types of plasmids that are used in Gateway® cloning together with the enzyme mixes to facilitate the reactions. The donor vector contains the ccdB gene flanked by the attP sites. The entry vector containing the attL sites is created by attB and attP recombination using the BP clonase enzyme. The attB expression clone is created by recombination of the entry vector attL sites and destination vector attR sites by the LR clonase enzyme.

3.2.6.2 Validation of LRP5 expression constructs' expression in HEK293 cell line

The aim of this experiment was to verify the protein expression from the LRP5 constructs created in this study. HEK293 cells were transiently transfected with each construct and the expression of the fusion proteins verified by western blot (WB) (Section 2.27.2). The expression constructs tested were pDEST40_LRP5 WT, pDEST40_LRP5 p.(Tyr984Cys), pDEST47_LRP5 WT and pDEST47_LRP5 p.(Tyr984Cys) along with the empty vectors pDEST40 and pDEST47 as controls. Cells treated with transfection reagent (Lipofectamine® 2000) only were used as a transfection control. 48 hours after

transfection, the cells were lysed, and protein extraction was performed (Section 2.24). Total protein concentrations for the cell lysates were determined using the BCA assay (Figure 3.13, Section 2.25) and 20μg of protein was used per sample for WB analysis.

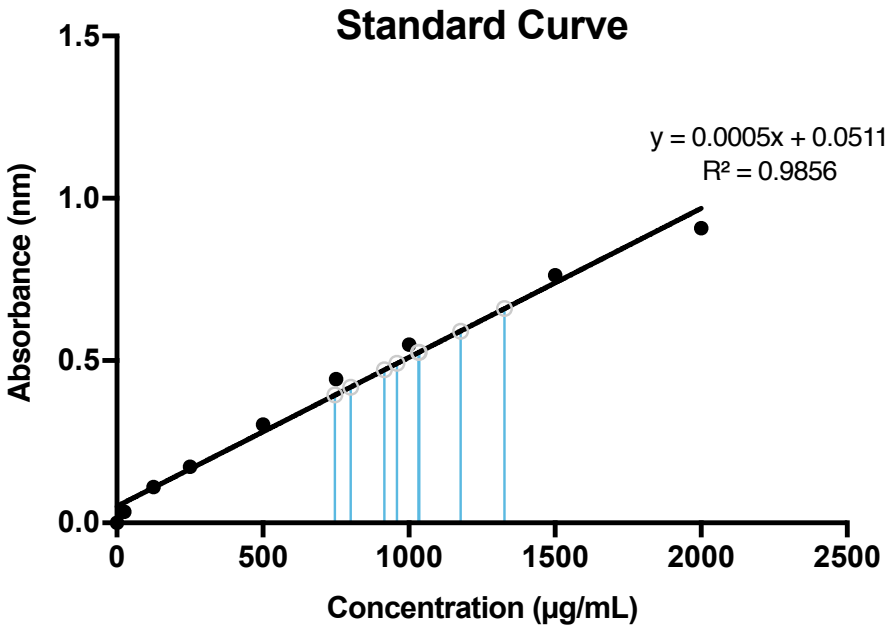


Figure 3.13 Standard curve for the bicinchoninic acid (BCA) assay to determine the quantity of protein in the cell lysates.
BCA assay was performed with protein standards to quantify the protein expression levels within the transfected cell samples with the expression of vectors pDEST40_LRP5 and pDEST47_LRP5 and their corresponding empty vectors. Black line with black circles indicates the standard curve whilst the blue line with grey empty circles indicate the calculated concentration of each unknown sample.

The LRP5 fusion proteins generated by the pDEST40 and pDEST47 constructs were detected using rabbit α-LRP5 antibody. The primary antibody was then detected by goat α-rabbit HRP antibody and visualised by Femto SuperSignal West Reagent (Figure 3.14, Section 2.27.1). The molecular weight of native LRP5 (NP_002326.2) is 179 kDa determined by the ExPASy ProtParam tool (http://web.expasy.org/cgi-bin/protparam/protparam). The molecular weight of pDEST40_LRP5 including the C-terminus tag of V5

epitope tag followed by 6x-His tag in pDEST40_LRP5 is 180 kDa. The molecular weight of pDEST47_LRP5 including the C-teminus GFP tag is 205 kDa. The blot was stripped with Restore™ Western Blot Stripping buffer and re-probed with mouse α-β-actin antibody, followed by rabbit α-mouse HRP conjugated antibody to detect the loading control β-actin (molecular weight 42kDa) (Section 2.27.1). This analysis confirmed that all the LRP5 constructs were producing fusion proteins of the expected size and at comparable levels (Figure 3.14).

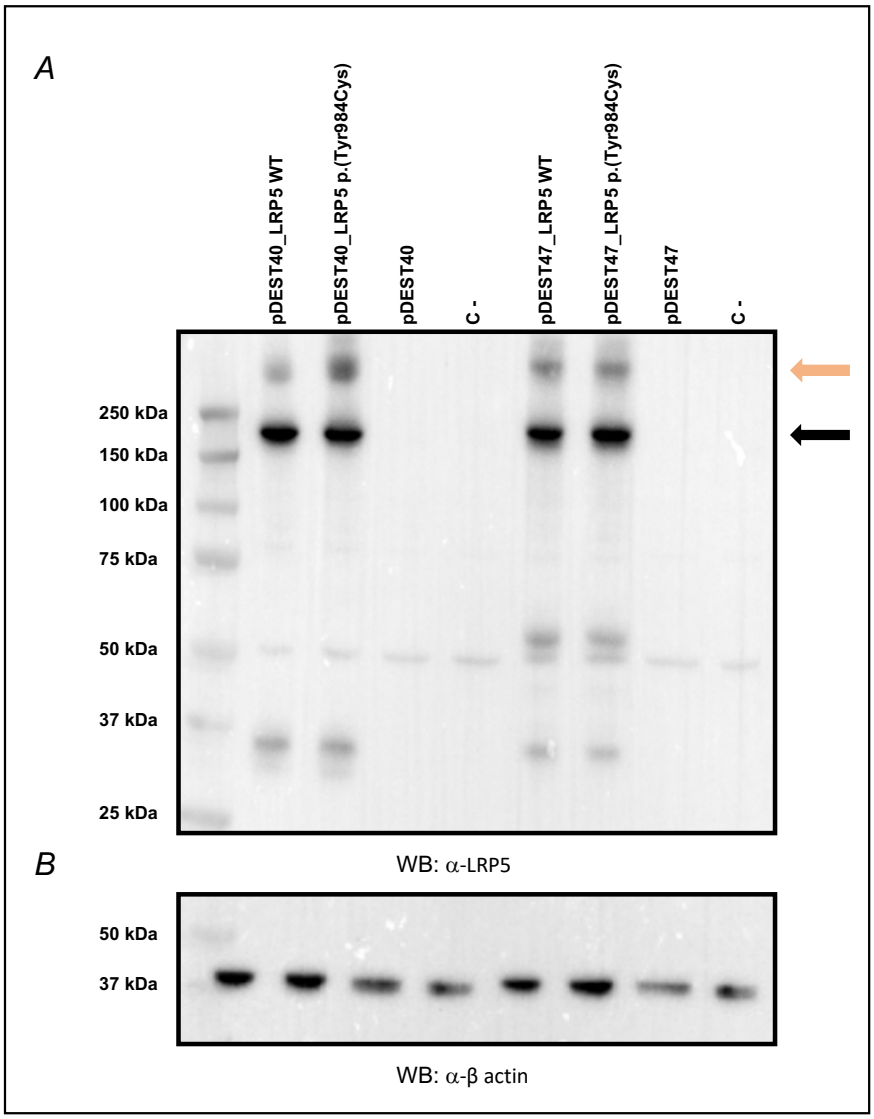


Figure 3.14 WB to verify the LRP5 expression constructs.
HEK293 cells were transiently transfected with pDEST40_LRP5 and pDEST47_LRP5 WT, p.(Tyr984Cys) mutants and corresponding empty vectors and the presence of LRP5 in the cell lysates was determined by WB. ***A:*** WB of pDEST40_LRP5 WT and p.(Tyr984Cys), pDEST40 empty and transfection only cell lysates (C-) incubated with α-LRP5 antibody (Invitrogen, cat no. 36-5400). ***B:*** The same blot probed with α-β actin antibody (Cat no. A2228) after being stripped to show equal loading. Black arrow corresponds to the expression of fusion proteins and the orange arrow shows the expected dimers formed by LRP5. The ladder is Precision Plus Protein™ All Blue Standard (Bio-Rad). The molecular weight of endogenous LRP5 is 179 kDa, which is not seen in the blot due to low exposure, pDEST40_LRP5 is 180 kDa including the C-terminal tag (V5 and 6x His tag). pDEST47_LRP5 construct is 205 kDa including the C-terminal tag of GFP. β-actin showed the predicted molecular weight of 42kDa.

3.2.6.3 Investigating the effect of the LRP5 p.(Tyr984Cys) variant on Norrin-β-catenin signalling using the TOPflash assay

The aim of this experiment was to investigate whether the LRP5 p.(Tyr984Cys) variant had any effect on the activation of the Norrin-β-catenin signalling pathway. The receptor complex of the Norrin-β-catenin pathway composes of FZD4, LRP5 and TSPAN12. The ligand protein activating the pathway is Norrin (Section 1.8.2.2). The TOPflash assay was created to assess the activation of the Wnt-signalling pathway (Molenaar et al., 1996) but it is widely used today to also evaluate the activation of Norrin-β-catenin signalling. The assay consists of a TOPflash reporter construct which encompasses the Firefly luciferase gene downstream of seven copies of the TCF/LEF binding sites known as the SuperTOPflash (Xu et al., 2004). The cells used for the TOPflash assay were STF cells, HEK293 cells stably transfected with the SuperTOPflash (STF) reporter construct (cell line was a kind gift from Prof Jeremy Nathans, John Hopkins University, USA) (Xu et al., 2004). Upon the binding of the Norrin ligand and activation of the pathway, β-catenin translocates into the nucleus where it binds to the TCF sites on the TOPflash construct and allows for the activation of the *Firefly* luciferase gene (Figure 3.15). The activation of β-catenin signalling is, therefore, directly proportional to the amount of luminescence signal being emitted.

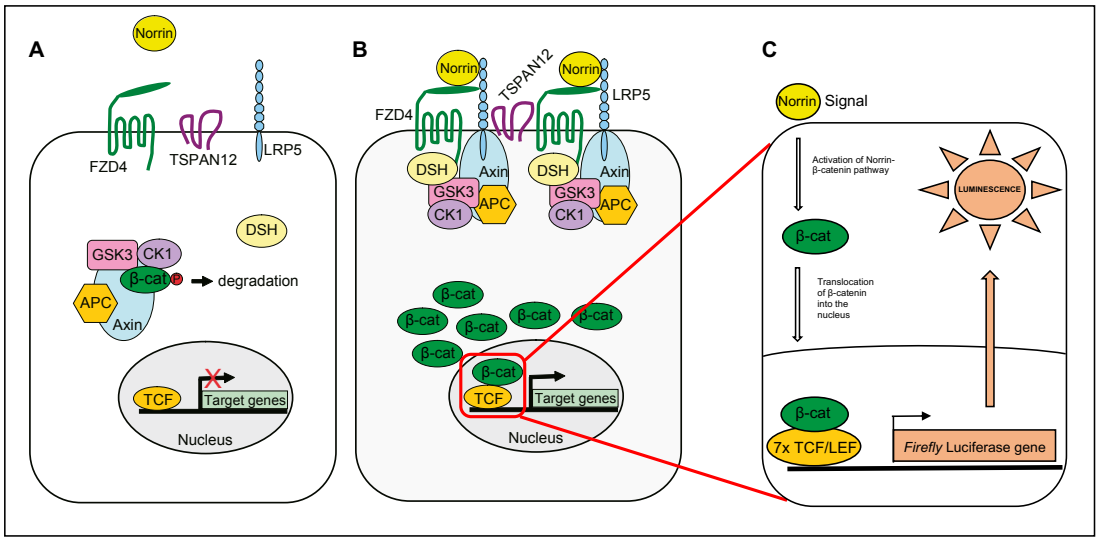


Figure 3.15 Schematic of the TOPflash reporter assay.
A: Schematic of the Norrin-β-catenin signalling pathway when the ligand (Norrin) is not bound to the receptor complex composed of FZD4, LRP5 and TSPAN12. β-catenin is phosphorylated and subsequently undergoes degradation and target genes are not transcribed. **B:** Norrin binds to the receptor complex and sets of a signalling cascade which inhibits the degradation of β-catenin allowing it to translocate into the nucleus. β-catenin binds to the TCF sites and activates target genes. **C:** The TOPflash reporter is activated when β-catenin translocates into the nucleus and interacts with TCF/LEF and binds to the 7x TCF/LEF binding sites leading to the activation of the luciferase reporter gene. This allows for the luminescence values to be measured. Figure adapted with permission from Dr Carmel Toomes.

For the assay, the receptor complex components (LRP5, FZD4 and TSPAN12) and the ligand (Norrin) of the Norrin-β-catenin signalling pathway were co-transfected into the STF cells. This method was adapted from that described in Xu et al. (2004), but was modified by adding the receptor TSPAN12, which was identified as a component of the receptor complex in 2009 (Junge et al., 2009). The pDEST40_FZD4 and pDEST40_TSPAN12 constructs were created and verified by Soriano (2017). The AP-3myc-Norrin construct was a gift from Prof Jeremy Nathans (John Hopkins University, USA) which contained the alkaline phosphatase (AP) tag along with 3x myc tag at the N-terminus of the protein (Xu et al., 2004).

The TOPflash assay was carried out in 24-well plates containing STF cells grown to 70-80% confluency. The pDEST40_LRP5 WT, pDEST40_LRP5 p.(Tyr984Cys) or empty pDEST40 constructs were transiently co-transfected with pDEST40_FZD4, pDEST40_TSPAN12 and AP-3myc-Norrin constructs with Fugene® 6 (Section 2.23.5). A transfection control vector *Renilla* (pRL-TK) was also co-transfected with the expression constructs. Control experiments transfecting in only empty pDEST40 and pRL-TK and a mock transfection (transfection reagent but no DNA) were also performed. To ensure that the amount of DNA transfected into cells remained constant (400ng) and to control for DNA-related transfection toxicity, the quantity of empty pDEST40 vector was varied where necessary. Each condition was performed in triplicate. Cells were lysed 48 hours post transfection and the levels of *Firefly* and *Renilla* luciferase were determined in triplicates using the Dual-luciferase reporter assay system (Section 2.26). The *Firefly* signal was normalised to the *Renilla* luciferase levels. Pathway activation was expressed as relative luciferase units (RLU) using the *Firefly/Renilla* ratio. The data was subjected to statistical analysis using GraphPad PRISM 7.0 software and the statistical test used was one-way ANOVA. All individual experimental values were pooled together, presented as a single data set with bars representing the standard error of the mean (Schagat et al., 2007).

The assay was independently repeated six times to compensate for the variability in co-transfecting four different constructs. The results showed that the TOPflash assay was working as expected as there was a statistically significant increase in reported activation between cells transfected with no LRP5 and WT LRP5 (Figure 3.16). The results presented a statistically

significant (p value ≤0.05) reduction of 18% in the TOPflash reporter activation in cells transfected with the p.(Tyr984Cys) mutant form of LRP5 compared to the WT.

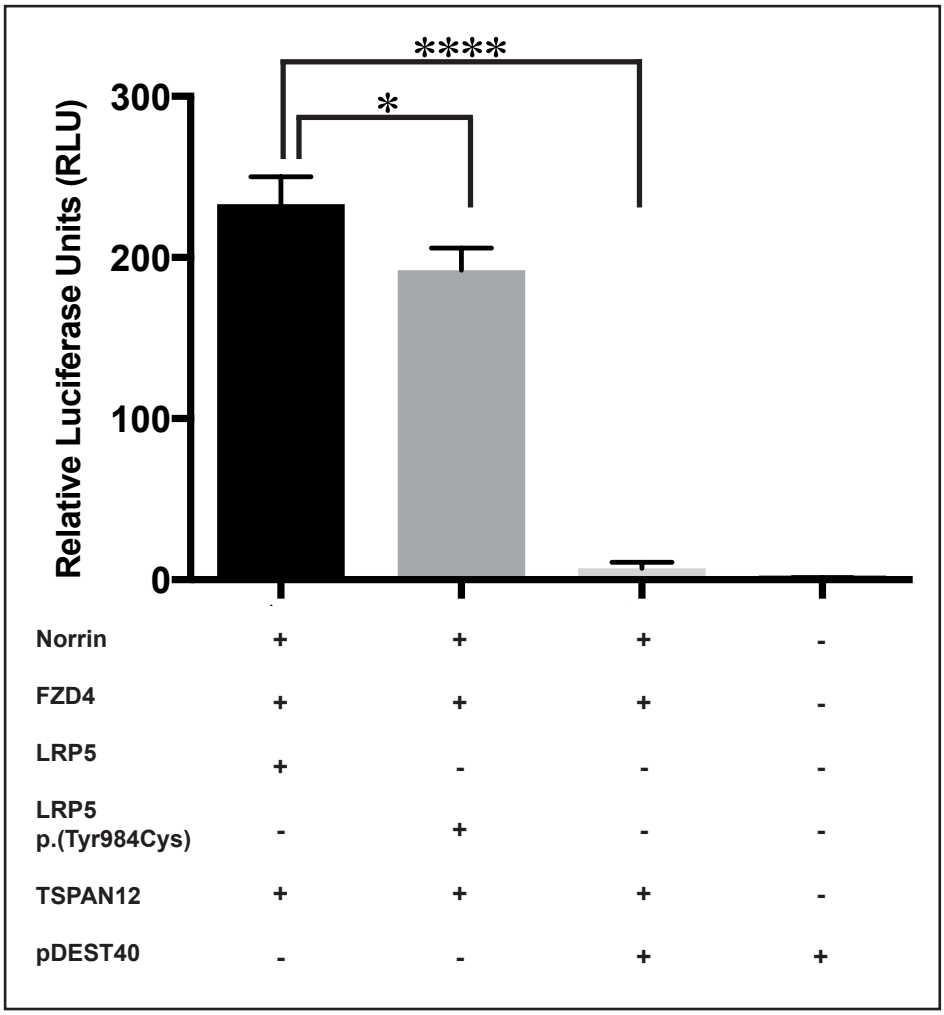


Figure 3.16 TOPflash luciferase reporter assay to assess the effect of the LRP5 p.(Tyr984Cys) variant.

HEK293 STF cells transfected with the components of the Norrin-β-catenin signalling pathway to compare the WT and mutant LRP5. Control of no LRP5 (column 3 from left) and control of empty vector (column 4 from left) were also included. The TOPflash activation levels were measured 48 hours post transfection, calculated as *Firefly/Renilla* ratio and reported as relative luciferase units (RLU). ∗: p value ≤0.05; ∗∗∗∗: p value ≤0.0001. Error bars represent the standard error of the mean. A one-way ANOVA multiple comparison test was performed. Comparison test between no LRP5 (column three from left) was made to the WT LRP5 (column one from left) to evaluate the assay functionality. Comparison test between the WT LRP5 and LRP5 p.(Tyr984Cys) was made. Experiments were performed in triplicates and the assay was independently repeated six times.

3.2.6.4 Localisation of LRP5 p.(Tyr984Cys) using LCI

The aim of this experiment was to investigate whether LRP5 p.(Tyr984Cys) had any localisation differences compared to the WT protein. The reduction in signalling in the TOPflash assay (Section 2.26) could indicate that the mutant protein's expression is lower in comparison to the WT at the cell surface or due to the changes of the mutant protein in receptor recycling. For this experiment, time-lapse imaging, in particular the Nikon BioStation IM live cell imaging, was used to visualise fluorescently tagged LRP5 created using the expression constructs pDEST47_LRP5 WT and pDEST47_LRP5 p.(Tyr984Cys). HEK293 cells were chosen for this experiment as this created a link to the TOPflash assay. HEK293 cells were seeded into a quadrant plate (µ-Dish 35mm Quad, Ibidi®) suitable for the chamber in the BioStation. Cells were transiently transfected with either pDEST47_LRP5 WT, pDEST47_LRP5 p.(Tyr984Cys) or empty pDEST47. The fourth quadrant was used as a mock transfection control (no DNA). pDEST47 empty vector served as a DNA control as it does not produce a GFP protein. Cells were first grown to 70% confluency and then transiently transfected with 300ng of each construct using Lipofectamine®2000 and were then visualised from 6 to 12 hours post transfection. Time-lapse images were taken at 5 minutes intervals and at 6 different points within the quadrant per condition. Movies were collected over a 24 hours period.

Unfortunately, the experiment could not be optimised to allow for the visualisation of single cells due to cell overgrowth. A number of different seeding densities were tested. However, at lower densities, the HEK293 cells detached and died. To resolve this issue, human cervical carcinoma epithelial

(HeLa) and human osteosarcoma (U2OS) cell lines were used instead. Both are carcinoma cell lines that do not possess adherent issues (Gey et al., 1952). The experiment was performed as described previously for the HEK293 cell line. However, upon transfection of the pDEST47 GFP tagged constructs, neither of these two cell lines had sufficient survival rates for the experiment. The transfection of the LRP5 WT and p.(Tyr984Cys) constructs resulted in the lowest survival rate with ~50% of cells dying, whereas transfection with the empty pDEST47 vector showed ~30% cell death. The transfection reagent only control did not show any significant cell death. This pattern of cell death was never recorded when the HEK293 cells were used for transfections. The surviving cells didn't look healthy and they were not suitable for analysis. Attempts were made to optimise the experiment by enhancing cell survival by feeding the cells post transfection with 30% FCS supplemented media. This boosted the cell survival; therefore, the cells were imagined on the Biostation. However, the cells started dying within the first 12 hours of imaging and no valuable results were obtained. This experiment was not fully optimised and due to the issues encountered the LCI results were not deemed to be reliable and the assessment of the *LRP5* missense variant is still ongoing.

3.2.7 Second run of WES

A second attempt at DNA extractions and library preparation was undertaken as the first run of WES did not produce the required read depth to warrant in-depth analysis of the data (Section 3.2.3). All the samples were analysed again but this time, block 71208M1 was substituted for block 71208M2 (as both blocks contain tissue from the same patient eye).

For this second experiment, the dewaxing and DNA extraction protocols were revisited and optimised to produce not only a higher yield of DNA, but also better DNA quality. This time, the deparaffinisation was performed on individual sections mounted on slides rather than in an eppendorf tube containing multiple sections. The period of deparaffinisation was also extended from 2 to 5 minutes incubations in xylene and graded ethanol. The Proteinase K incubation was also extended from a period of 1 hour to 72 hours, with the enzyme re-added after the first 24 hours. The rest of the DNA extraction steps were carried out as described previously (Section 2.5).

In both DNA extraction attempts, the same number of sections was used (48-50 sections at 5µm thick) allowing for a valid comparison between the two extraction methods. The improvement in DNA yield between samples was clearly visible with an average of 186.6ng/µl compared to DNA yield from the first set of DNA extractions of 47.26ng/µl (Table 3.1and Table 3.8).

3.2.7.1 Quality control of extracted FFPE DNA

Although DNA yields for the first run of WES were sufficient for the WES protocol to be performed on all samples, the poor quality data generated was influenced by the poor quality input DNA, which was not fully amplifiable. Therefore, an additional QC step was undertaken to ensure the quality of the newly extracted DNA was optimal for WES prior to the library preparation. This was performed using the NGS FFPE QC kit (Section 2.8) which used a real-time PCR method to measure the quantity and quality of amplifiable DNA in the extracted FFPE DNA (Figure 3.17).

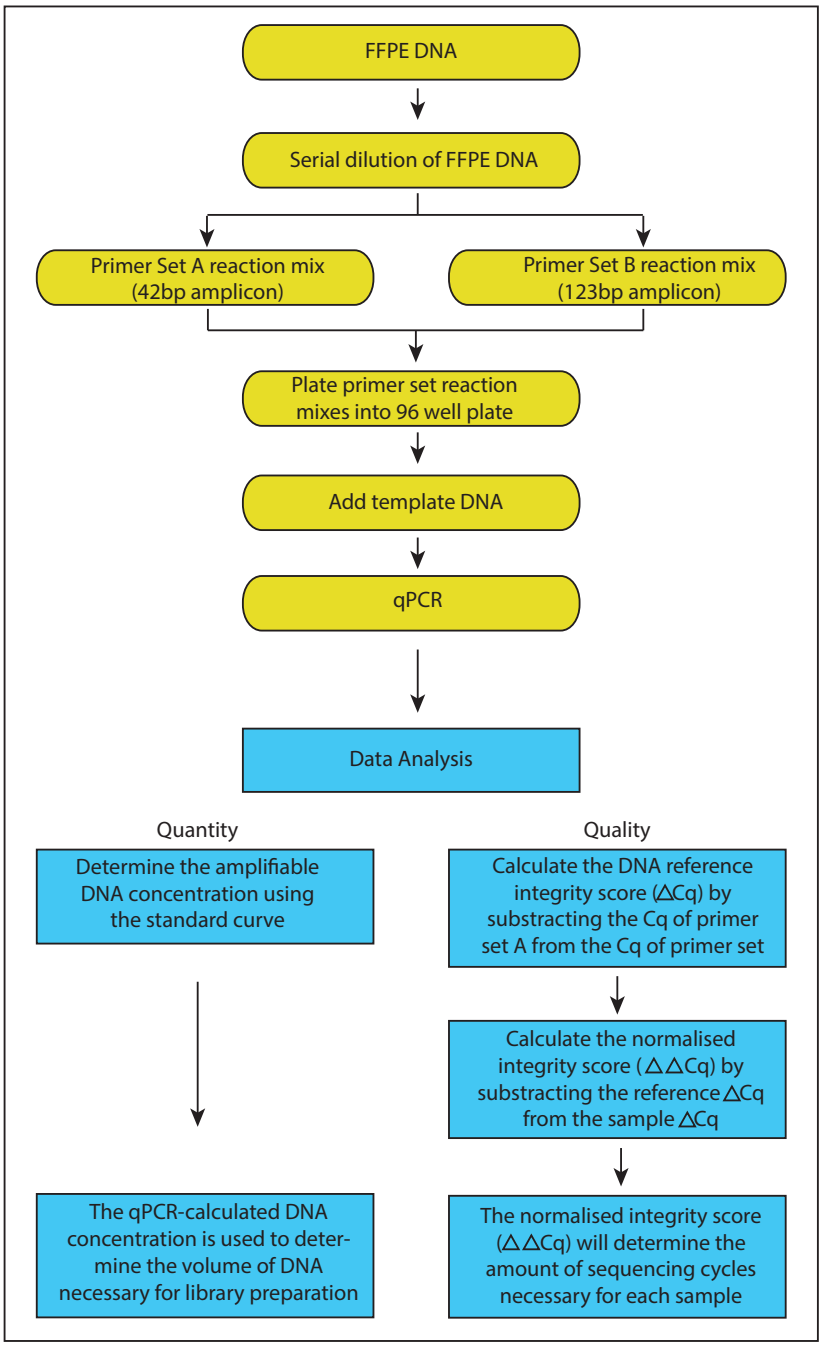


Figure 3.17 Schematic representation of the NGS FFPE QC Kit.
DNA is prepared in serial dilutions and added to the reaction mixes of primer set A (42bp amplicon) and primer set B (123bp amplicon). Data analysis provides the quantity of amplifiable DNA that is determined by the standard curve analysis. This analysis is then used for the determination of the amount of DNA to be used for library preparation. Quality of DNA is based on the DNA integrity (ΔCq) score which determines the cycling conditions for each sample during library preparation. Image adapted from Agilent Technologies.

Briefly, the FFPE DNA is diluted and added to the primer reaction mixes that target two different amplicons (42bp and 123bp) within the human genome and

the qPCR is performed. The analysis determines the amount of quantifiable DNA within the sample using a standard curve analysis and this was then used to establish the necessary amount of DNA to use as an input in the WES library preparation (Table 3.8). The data showed that on average only 14.6% (12.4% - 17.5%) of the extracted DNA was able to be amplified and, therefore, usable in downstream applications.

Sample ID	DNA (ng/µl)	Amplifiable DNA (ng/µl)	Percentage of amplifiable DNA
20417	72	12.6	17.5%
48753	294	45.6	15.5%
31276	178.8	24.7	13.8%
71208 M1	296	42.3	14.3%
67662 ON	35.6	4.4	12.4%
63170-1	372	53.5	14.4%
67580	57.8	8.5	14.7%

Table 3.8 Concentrations and percentage of amplifiable DNA of Coats patients. Extracted DNA concentration calculated by Qubit fluorometer (ng/µl) is shown in the second column post vacuum concentration. Using the Agilent NGS FFPE QC Kit (Agilent Technologies), the concentration (ng/µl) of amplifiable DNA was determined by standard curve analysis and percentages subsequently calculated.

The quality of the FFPE DNA was assessed by firstly determining the DNA reference integrity score (ΔCq_{ref}). The DNA integrity score of each sample (ΔCq_{sample}) was produced for each sample from the qPCR values of primer sets A and B. This value was then normalised to the ΔCq_{ref} to generate the

normalised DNA integrity score (ΔΔCq). An example of the calculation for patient 71208M1 to determine the DNA integrity number is shown in Figure 3.18 and the DNA integrity scores for each patient are shown in Table 3.9.

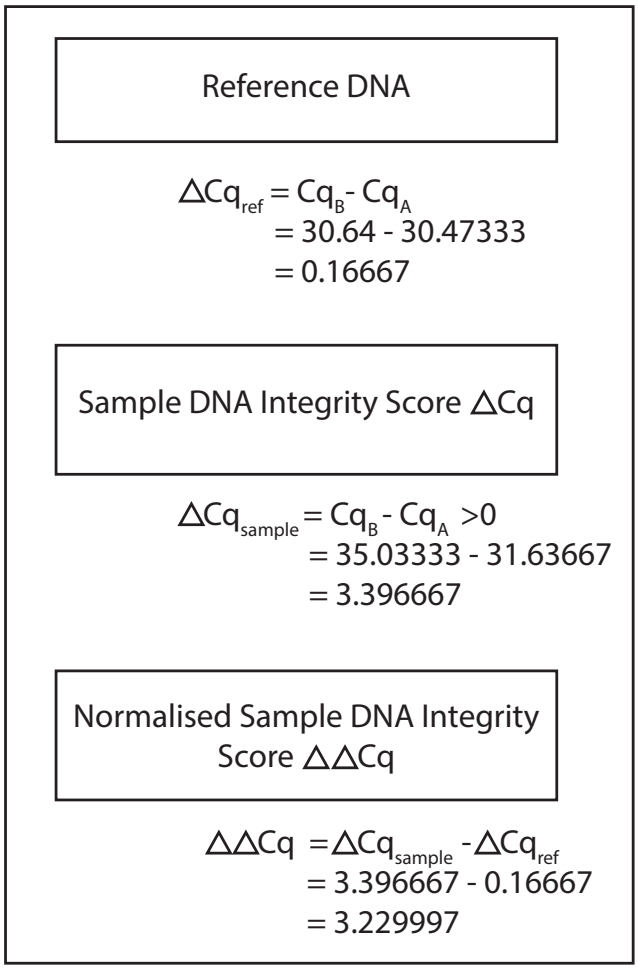


Figure 3.18 An example of the calculation of DNA integrity score for patient 71208M1.

Sample DNA integrity score (ΔCq) was determined by subtracting the quantification cycle (Cq) values of primer set B of the larger amplicon (123bp) from Cq values of primer set A of the smaller amplicon (42bp). The normalised DNA integrity score (ΔΔCq) was determined by subtracting the ΔCq of the sample from the reference ΔCq. This value then determined the number of cycles needed for amplification during WES library preparation (Table 3.10).

Sample ID	DNA Integrity Score ($\Delta Cq_{Sample} = Cq_B - Cq_A > 0$)	Normalised DNA Integrity Score ($\Delta\Delta Cq = \Delta Cq_{Sample} - \Delta Cq_{Ref}$)
20417	4.63	4.46
48753	2.33	2.17
31276	3.05	2.89
71208 M1	3.39	3.23
67662 ON	1.06	0.89
63170-1	5.30	5.13
67580	2.57	2.40

Table 3.9 Calculated DNA integrity and normalised DNA integrity score for all Coats disease samples.
Integrity scores were calculated according to manufacturer's instructions using the NGS FFPE QC Kit (Agilent).

The normalised DNA integrity number for each sample established the number of cycles required for the amplification step during the library preparation (Table 3.10). Sample 67662ON had a different method of quantification in comparison to the rest of the samples as its normalised integrity score ($\Delta\Delta Cq$) was <1.3.

Sample ID	ΔΔCq	No. of cycles	Method of quantification
20417	4.46	13 cycles	qPCR
48753	2.17	10 cycles	qPCR
31276	2.89	10 cycles	qPCR
71208M1	3.23	10 cycles	qPCR
67662ON	0.89	10 cycles	Qubit
63170-1	5.13	13 cycles	qPCR
67580	2.4	10 cycles	qPCR

Table 3.10 DNA integrity scores (ΔΔCq) for all samples.
Cycles and quantification method determined based on the guidelines for different DNA integrity scores (ΔΔCq) according to instructions from SureSelect^{XT} kit. Qubit was used for sample 67662ON as its integrity number was ΔΔCq ≤1.

3.2.7.2 Library preparation for second run of WES

A second run of WES was undertaken using the input DNA guidelines from the Agilent NGS FFPE QC kit (Section 2.8). For each sample, 200ng of amplifiable DNA was used, apart from sample 67662ON where only 125ng was used. Similarly, the ΔΔCq scores for each sample determined the number of cycles needed for the pre-capture PCR amplification step during the library preparation (Table 3.10). This prevented over amplification during PCR and the introduction of more artefacts. If primers become limited during the PCR (using too many cycles), the templates can anneal to themselves and form a "loop". This artefact runs at high molecular weight and interferes with library preparation (personal communication, Sally Fairweather NGS sequencing

facility). Other modifications were also made to the WES library preparation protocol, all of which were recommended by the Agilent NGS FFPE QC kit instructions and these modifications are detailed in Section 2.8.

Protocol Step & Parameter	Non-FFPE samples	FFPE samples		
		ΔΔCq≤1*	**ΔΔCq between 1-4**	**ΔΔCq>4**
DNA input for Library preparation	200ng, based on Qubit Assay	200ng, based on Qubit Assay	100ng-200ng of amplifiable DNA, based on qPCR quantification	100ng-200ng of amplifiable DNA, based on qPCR quantification
Pre-capture PCR cycle number	10 cycles	10 cycles	10 cycles	13 cycles

Table 3.11 DNA integrity (ΔΔCq) modification table for WES library preparation. Quantification method and number of cycles breakdown based on the ΔΔCq score for library preparation for FFPE samples according to SureSelect^{XT} Target Enrichment System for Illumina Paired End Multiplexed Sequencing Library kit (Agilent Technologies) protocol. **If FFPE samples had ΔΔCq ≤1, they were treated as non-FFPE for the above steps. DNA concentrations were used from Qubit assay to calculate the required 200ng.*

Libraries were prepared using the SureSelect^{XT} 200ng protocol and v6 SureSelect^{XT} Capture library target enrichment baits with modifications to the protocol outlined in Section 2.17 (Figure 3.19). The SureSelect^{XT} Human All Exon v6 capture library is an updated version of the v5 that was used in the initial run of WES.

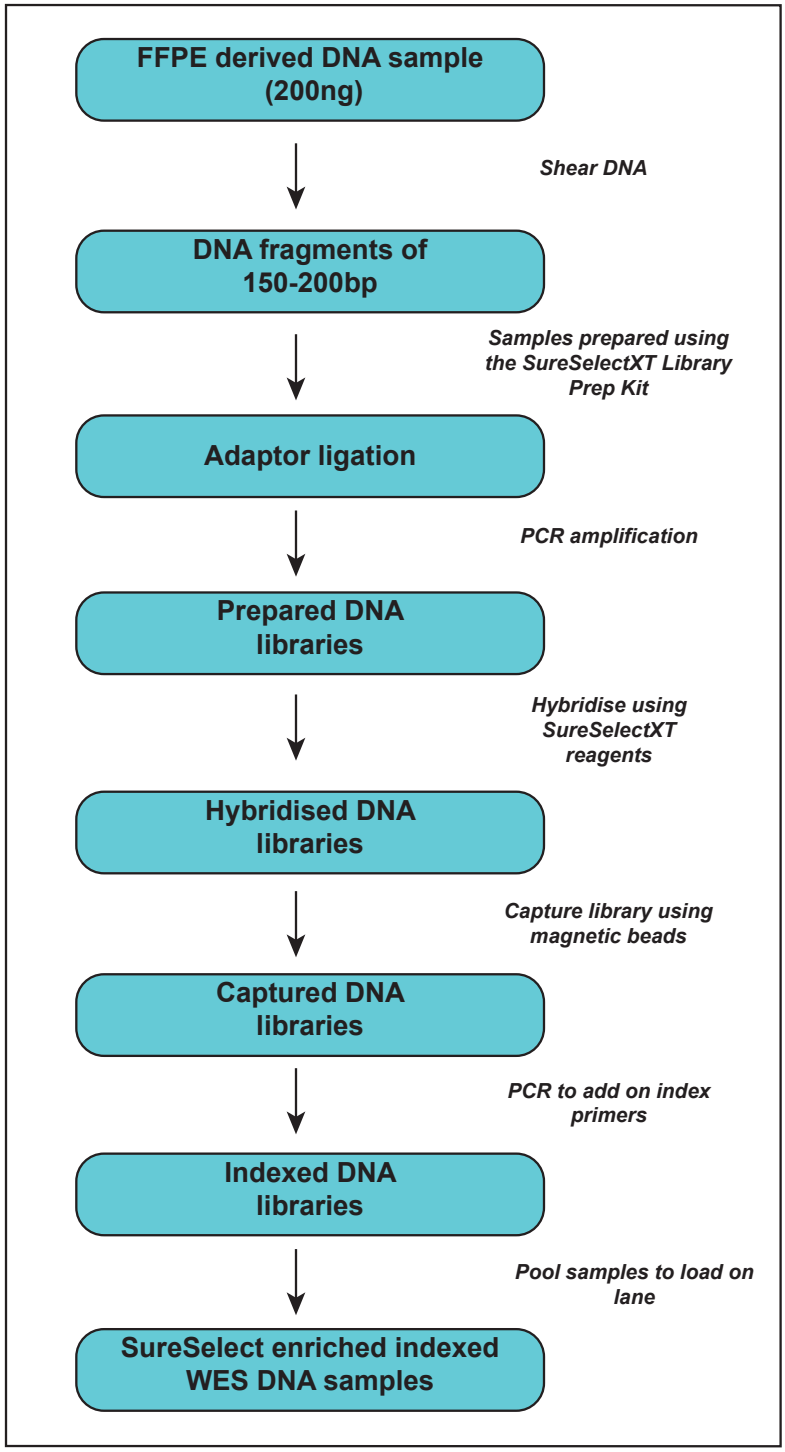


Figure 3.19 Outline of 2nd run of WES library preparation protocol.
Flow diagram of the SureSelect^{XT} Target Enrichment System for Illumina Paired End Multiplexed Sequencing Library kit (Agilent) using the low DNA input of 200ng for each sample.

This protocol included shearing of the samples with the Covaris S2 system as described previously. However, during the second run of library preparations,

the samples were assessed for their fragment size using the 2200 TapeStation with the High sensitivity D1000 DNA screen tape instead of the Bioanalyser The protocol was then followed as detailed in Section 2.15. The results of the sheared samples with the ligated adapters are presented in Figure 3.20 and all samples produced the desired fragments of ~300bp.

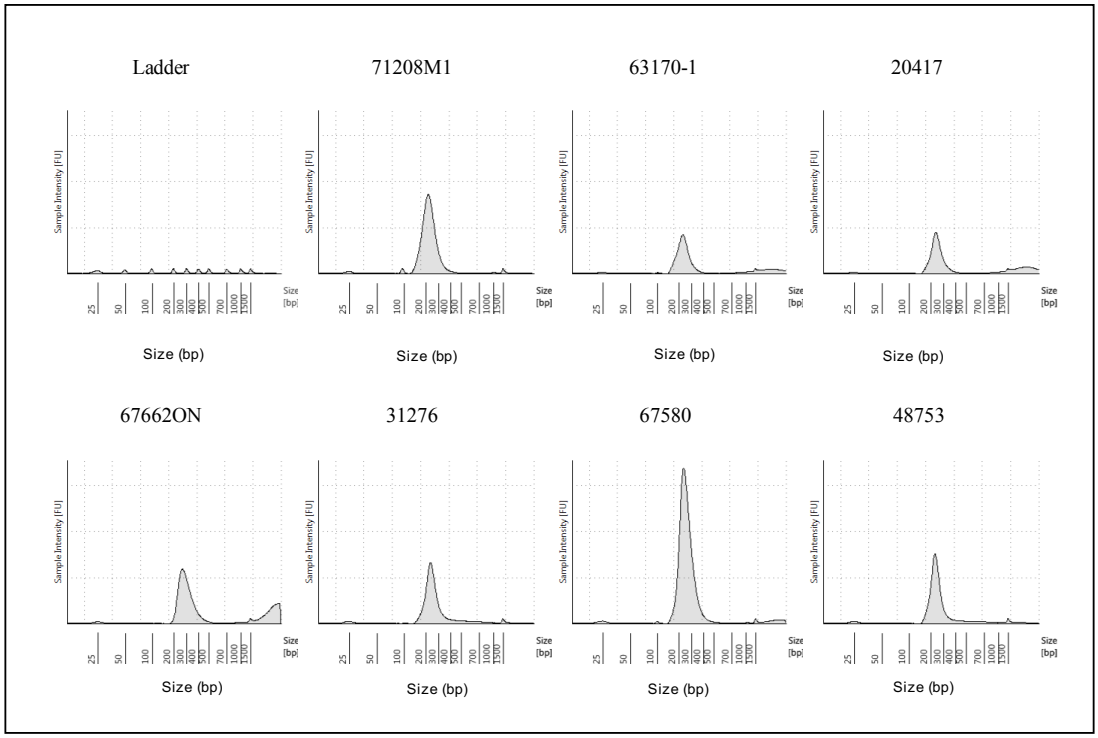


Figure 3.20 Tape station analysis of Coats patients' DNA prior to library prep. Sheared samples with ligated adapters of seven Coats disease patients assessed for size of DNA fragments. Fragments of 300bp were detected for all samples. X-axis shows the size in base pairs (bp) and the y-axis represents the signal intensity of detected DNA.

Samples were pooled and sequenced as 150bp paired end reads on one lane on the HiSeq3000 (Illumina) sequencing platform by the University of Leeds Next Generation Sequencing Facility (http://dna.leeds.ac.uk/genomics/).

3.2.7.3 Bioinformatics analysis of 2nd run of WES

The bioinformatics analysis for the second set of WES data was analysed with the same pipeline as described in Section 3.2.3. The same QC analysis was

undertaken for the second set of WES data (Section 3.2.3). This showed that the quality of scores across all bases and the average quality per read has dramatically increased in quality in comparison to the first run of WES (Figure 3.21A+B). The number of duplicates has also decreased in the second run of WES (Figure 3.21C). The duplication numbers for the rest of the samples are shown in Appendix 7.2. On average, the duplication levels between the first and second run of WES has decreased from 69.89% to 55.65%.

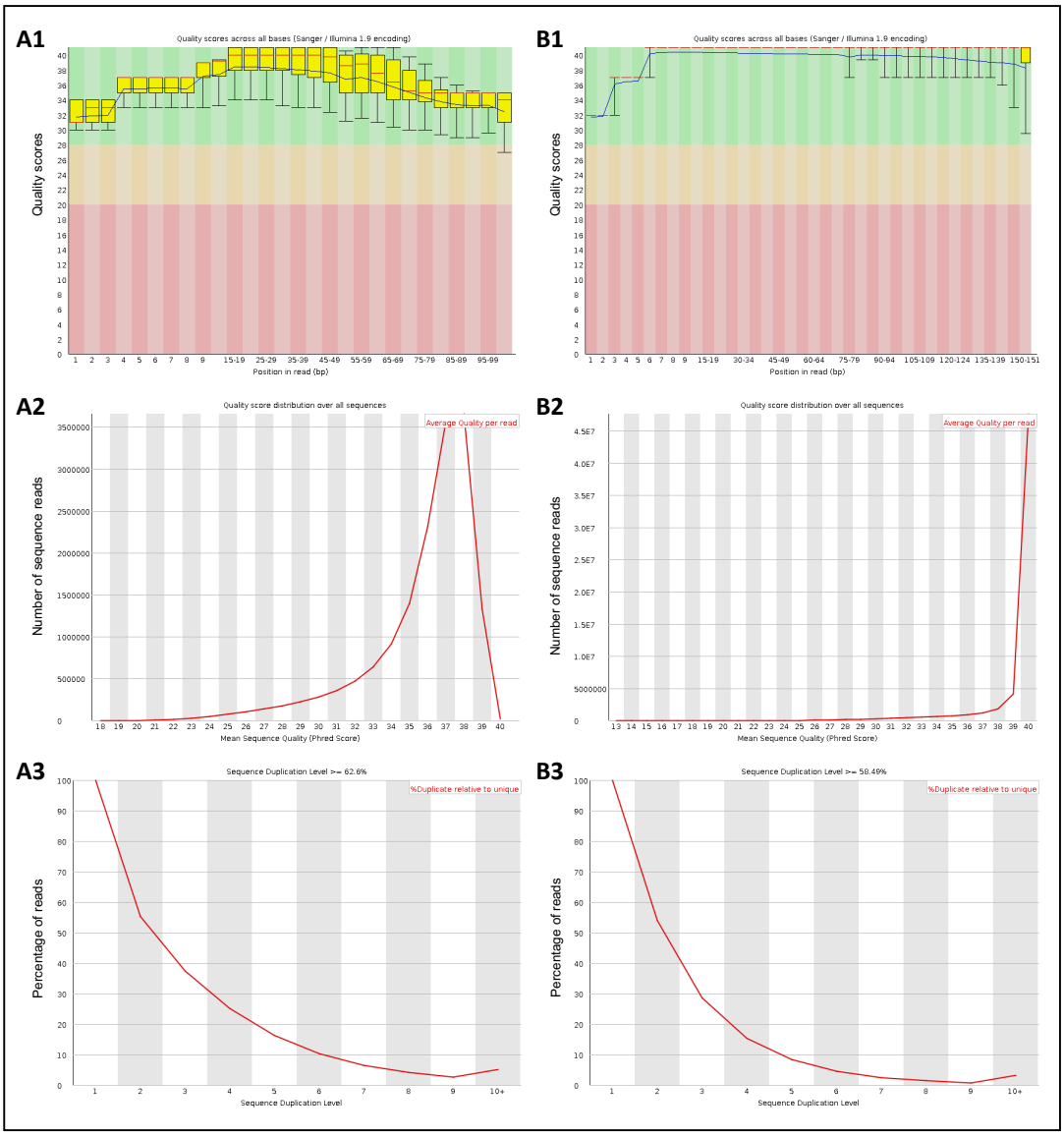


Figure 3.21 Comparison of the QC between first (A) and second (B) run of WES for patient 71208M2.
(A1) & (B1) A representative of the QC of the raw fastq file for patient 71208M2 using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 71208M2. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 62.6% of all sequences were duplicated between 2 to 10+ times. 37.4% of all sequences were only read once. However, from the second run of WES (B3), the percentage of duplicates was improved to 58.49% between 2 to 10+ times. 41.51% of all sequences were only read once.

The depth of coverage data for the second run of WES is summarised in Table 3.12. The results showed that the new WES data was much higher quality than the previous batch as a minimum of 91% of bases in each sample were covered by over 20 reads compared to only one sample achieving this level of coverage in the first data set and many having substantially lower values of 6-7% (Table 3.2) (Section 3.2.3).

Sample ID	% of bases above 5	% of bases above 10	% of bases above 20
20417	98.8	97.7	94.6
48753	97.6	95.9	91.6
31276	98.6	97.5	94.9
71208 M1	99.7	99.3	97.9
67662 ON	99.7	99.4	98.3
63170-1	98.5	97.2	93.5
67580	99.6	99.1	97.6

Table 3.12 Depth of coverage for Coats samples for second run of WES.
Sample IDs are provided and corresponding percentage of bases covered above certain read depths are listed.

An additional analysis that was performed with the second batch of WES data was ExomeDepth to identify any large deletions or insertions (Section 2.20) (Plagnol et al., 2012). ExomeDepth was performed on the autosomes as well as the X chromosome. The reference datasets used were all the other FFPE DNA Coats samples and the software automatically decides which ones it

selects for the analysis (Table 3.13). Variants in the known FEVR genes were checked and no convincing variants were highlighted.

Sample ID	Controls used
20417	31276, 67580, 71208M1, 67662ON, 48753
48753	31276
31276	20417, 48753, 71208M1
71208 M1	67580, 20417, 67662ON, 31276, 63170-1
67662 ON	67580, 71208M1
63170-1	31276, 71208M1, 48753
67580	67662ON, 20417, 71208M1

Table 3.13 ExomeDepth analysis of Coats disease patients from the second run of WES.
Sample IDs and control IDs for the Exomedepth analysis of individual patients.

The detailed analysis of the entire data set was then carried out following the same pipeline as in Figure 3.6 with modifications that are described later in this section. However, prior to this, the data was put through the exact same analysis as the first batch of WES data, using each of the described commands to identify functional, heterozygous and bi-allelic variants, to allow a comparison between the two data sets. The results of the number of variants per command are summarised in Table 3.14. The number of variants significantly decreased for the second set of WES data in comparison to the first (Table 3.3). This showed that the better quality DNA used for generating

the sequence had produced higher quality data, which in turn facilitated distinguishing between real variants and artefacts and, therefore, allowed for more confidence in the identified variants.

Sample ID	GetFunctional	GetHeterozygous	FindBiallelic
20417	1,474	5,184 (1,002)	944 (385)
31276	1,399	4,511 (901)	915 (361)
48753	1,419	4,866 (960)	922 (401)
67580	1,516	5,041 (1,025)	999 (429)
67662ON	1,510	4,595 (970)	1,043 (419)
63170-1	1,503	5,188 (1,038)	938 (405)
71208M1	1,331	4,269 (852)	911 (389)

Table 3.14 Number of variants present from WES data analysis using vcfhacks filters.
Overall number of variants for each patient using the GetFunctional.pl command including all the functional variants such as missense, frameshift, nonsense and splice variants. GetHeterozygous.pl command identifying the overall number of variants in a heterozygous state. Number of all identified bi-allelic variants per patient and the number of functional bi-allelic variants using the FindBiallelic.pl command in brackets.

The subsequent bioinformatics analysis differed to the initial WES analysis (Section 3.2.3) in a number of ways. The general pipeline in Figure 3.6 was followed until the Haplotype caller step where a gVCF (VdAuwera, accessed 31 October 2017) was generated. During the second analysis, data generated from gDNA of 57 unrelated control individuals was merged with the Coats' samples data. The controls were patients who were not diagnosed with any eye related diseases and whose WES data was generated in the same

manner. For this analysis, a script was used to extract variants from individual Coats' samples and filter them against the controls to remove anything that they may have in common to aid in removing SNPs and artefacts (Appendix 7.6).

Another step in the analysis that differed was the comparison of patients to each other to search for any shared variants. This filtering was performed after the removal of duplicates and common variants from variant databases and controls but prior to any strategic filtering based on function or zygosity. This command can be found in Appendix 7.6. The analysis was performed to check whether any common variants were shared between at least two or all patients (n=2 or n=7). This analysis was carried out to identify whether a gene could be the germline cause of Coats disease as somatic mutations are extremely unlikely to occur in the exact same location in two patients. However, there were no variants shared between all of the patients and no variants of interest between 2 patients. Further analysis was then performed to see if any of the patients had different variants in the same gene. A list of the top shared genes and their variants are summarised in Table 3.15. The variants presented were selected based on passing the quality threshold control filter that was used in the bioinformatic analysis, having an allele frequency of <0.01% in the general population, having a high allele depth and a CADD score >10.

SHARED GENES							
Gene ID	**cDNA position**	**Protein Position**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**	**ExAC**	**No. of patients**
MUC4	c.11721_11724delCTAT;	p.(Asp3909Leufs*349);	ENST00000479406.1	274/34; 38/140; 242/148	31	2/10,318*	3
	c.11721A>T	p.(Asp3909Val)	ENST00000466475	69/172; 38/140; 2/2; 242/148	12.31	2/10,318*	4
FOXO3	c.1140_1141insG;	p.(Leu382Alafs*3);	ENST00000343882.6	239/104; 236/26	31	1/120,980*	2
	c.1517T>A	p.(Arg506His)	ENST00000406360.1	179/155	23.8	1/116,362*	1
GRID2IP	c.1814C>T;	p.(Pro605Leu);	ENST00000452113.1	2/2	23.5	67/5,752*	1
	c.1407C>T	p.(Arg469Cys)	ENST00000452113	8/2; 6/4	15.66	3/13,594*	2

Table 3.15 Shared genes with any coding variants present in at least 2 patients.
MAF of 0.0001 (<0.01% in population) and all mutations presented have an effect on the coding amino acid. The genes presented were the top candidates based on the PASS filter, CADD score >10 and high allele depth and not present in ExAC. *Number of alleles to the closest SNP.

3.2.8 Candidate variants identified in individual Coats patients by WES

Bioinformatic analysis was also done for each patient individually based on the pipeline and filtering as described earlier. The resulting candidate lists are documented in separate tables for each patient (Table 3.16 - Table 3.22). The analysis of the second run of WES confirmed the presence of the *LRP5* variant c.2951A>G, p.(Tyr984Cys) in patient ID 71208M2.

20417					
Gene ID	**cDNA change**	**Protein change**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**
MDGA2	c.391C>T	p.(Arg131Trp)	ENST00000486952.2	42/60	34
KMT2C	c.2740delT	p.(Ser914Valfs*42)	ENST00000355193.2	10/4	32
OIP5	c.546C>T	p.(Ala182Val)	ENST00000220514	45/23	32
SSU72	c.182C>G	p.(Tyr61Cys)	ENST00000359060.4	64/53	27.7
EHBP1L1	c.1660C>T	p.(Ile554Phe)	ENST00000533237.1	48/73	27.1
FAM211A	c.850C>T	p.(Ser284Cys)	ENST00000470794.1	62/41	26.7

Table 3.16 List of putative candidate variants from second run of WES for Coats disease patient ID 20417.
List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls.

31276					
Gene ID	**cDNA change**	**Protein change**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**
UPF1	c.3127_3131delCTTGC	p.(Asp1043Glyfs*41)	ENST00000262803.5	301/121	36
SHPRH	c.816G>T	p.(Val272Phe)	ENST00000519632	179/137	33
CARD11	c.1157G>A	p.(Arg386Gln)	ENST00000396946.4	91/86	32
MUC4	c.11721_11724delCTAT	p.(Asp3909Leufs*349)	ENST00000479406.1	274/34	31
CDC25B	c.1537C>T	p.(Ser513Cys)	ENST00000245960.5	283/244	26.8
MLXIPL	c.72C>A	p.(Asp24Glu)	ENST00000429400	32/50	26.5

Table 3.17 List of putative candidate variants from second run of WES for Coats disease patient ID 31276.
List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls.

48753					
Gene ID	**cDNA change**	**Protein change**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**
PROP1	c.334A>T	p.(Arg112*)	ENST00000308304.2	83/57	36
MACF1	c.3898G>A	p.(Gln1300Lys)	ENST00000524432.1	39/44	28.4
PLEKHH1	c.379A>G	p.(Lys127Glu)	ENST00000329153.5	50/44	28.1
WAPAL	c.1825G>C	p.(Val609Leu)	ENST00000298767.5	93/70	24.2
AKAP3	c.1271T>C	p.(Met424Thr)	ENST00000545990.2	34/30	22.6

Table 3.18 List of putative candidate variants from second run of WES for Coats disease patient ID 48753.
List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls.

63170-1					
Gene ID	**cDNA change**	**Protein change**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**
SETD9	c.29T>A	p.(Trp10*)	ENST00000285947.2	134/141	36
KCTD6	c.609delG	p.(Phe203Leufs *11)	ENST00000490264.1	70/63	34
CEP95	c.1625T>A	p.(Arg542Lys)	ENST00000553412.1	33/22	33
IARS	c.2333A>G	p.Tyr778Cys	ENST00000447699.2	89/86	29.2
CLCN3	c.2165A>T	p.(Pro722Leu)	ENST00000504131.2	124/121	27.9
GLI3	c.2704C>A	p.(Ala902Thr)	ENST00000395925.3	139/108	26.9

Table 3.19 List of putative candidate variants from second run of WES for Coats disease patient ID 63170-1.

List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls.

67580					
Gene ID	**cDNA change**	**Protein change**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**
CTD-3088G3.8	c.1098T>A	p.(Trp366*)	ENST00000344649.3	92/107	36
OTUD7A	c.1066C>T	p.(Val356Phe)	ENST00000382902.1	83/69	33
ABI3BP	c.360C>T	p.(Arg120Cys)	ENST00000382902	95/121	32
PLD3	c.84C>A	p.(Ser29Asn)	ENST00000596470.1	71/77	29
LRP1B	c.2040T>C	p.(Trp680Arg)	ENST00000389484	119/90	26.1
ACAD11	c.2269A>C	p.(Asp757His)	ENST00000264990.6	98/93	25.1

Table 3.20 List of putative candidate variants from second run of WES for Coats disease patient ID 67580.
List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls.

67662ON					
Gene ID	**cDNA change**	**Protein Change**	**Transcript ID**	**Allele Depth (WT/Mut)**	**CADD score**
PKD1L2	c.3793A>T	p.(Glu1265*)	ENST00000533478.1	98/113	46
NELL1	c.865T>A	p.(Gly289Ser)	ENST00000357134.5	169/154	25.6
ADAMTS8	c.388G>A	p.(Gly130Ser)	ENST00000257359.6	14/20	25.3
IL17C	c.388C>A	p.(Leu130Met)	ENST00000244241.4	46/52	24.8
CSNK1A1L	c.856A>G	p.(Arg286Gly)	ENST00000379800.3	98/104	23.7
NYAP2	c.1472A>T	p.(Ala491Val)	ENST00000272907.6	118/117	23.4

Table 3.21 List of putative candidate variants from second run of WES for Coats disease patient ID 67662ON.
List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls.

71208M1					
Gene ID	**cDNA change**	**Protein change**	**Transcript ID**	**Allele depth (WT/Mut)**	**CADD score**
ISPD	c.206_207delGC	p.(Cys69Serfs*46)	ENST00000399310.3	140/113	34
RAD21L1	c.607T>A	p.(Asp203Asn)	ENST00000409241.1	120/82	32
**LRP5*	c.2951A>G	p.(Tyr984Cys)	ENST00000294304.7	200/180	29.2
TTC37	c.2689T>C	p.(Cys897Arg)	ENST00000358746.2	49/43	28.5
SCN10A	c.700A>T	p.(Val234Phe)	ENST00000449082.2	23/11	28.4
GABRG3	c.542C>T	p.(Asp181Val)	ENST00000554696.1	8/8	28.1

Table 3.22 List of putative candidate variants from second run of WES for Coats disease patient ID 71208M2.
List of each variant including the consequence, cDNA position, protein position, transcript number, allele depth and the CADD score. Criteria included: <0.01% in population, >20 CADD and high allele depth, heterozygous variants, not present in ExAC database and compared to unrelated controls. **LRP5* variant identified in patient 71208M1/M2 and described in detail in Section 3.2.4.

Apart from the *LRP5* variant, two more variants were investigated further. A heterozygous frameshift variant in *ATOH7* (NM_145178) c.46delG; p.(Ala16Hisfs*71) was one of the potential candidates in patient ID 20417. Allele depth for this variant was GC=7/G=2 with a CADD score of 23.9. This variant was not present in ExAC or gnomAD databases (accessed August 2017). *ATOH7* was a good candidate gene as it has been documented to be mutated in arFEVR cases (Khan et al., 2012; Panagiotou, 2018). Unfortunately, this variant was not covered in the first set of WES data so to confirm if it was real Sanger sequencing was performed. The patient's FFPE

DNA was used initially but it failed to amplify and produce any product even when the PCR product size was as low as 86bp. Therefore, the DNA was WGA (Section 2.12). The amplification was attempted numerous times with a selection of different primer pairs as *ATOH7* is a single exon gene with a very high GC content (all primer sequences are provided in Appendix 7.18). Eventually, a product was generated with primer set F/R, but this change was not present in the Sanger sequencing.

Another candidate gene identified in the WES data was *ISPD.* ISPD has a known function in the eye and mutations in *ISPD* are known to cause Walker-Warburg syndrome which is characterised by muscle, brain and eye abnormalities (Roscioli et al., 2012). The eye phenotype includes microphthalmia, cataracts, enlarged eye balls due to increased intraocular pressure, and the condition also affects the optic nerve. A heterozygous frameshift variant in *ISPD* c.206_207delGC; p.(Cys69Serfs*46) was identified in patient ID 71208M2. This variant was in the same Coats patient as the already described *LRP5* variant c.2951A>G; p.(Tyr984Cys). The amplification of the patients' DNA was not successful after multiple attempts. The WGA DNA was also unable to be amplified. The amplification was halted due to the value of DNA.

All the variants in the above tables were not manually verified by sequencing. Given the difficulties encountered amplifying *ATOH7* and *ISPD*, an alternative strategy for confirming the variants was to repeat the DNA extraction and perform a third round of WES. This work is still ongoing.

3.3 Discussion

In this chapter, FFPE eye tissue from seven unrelated Coats disease patients were subjected to DNA extraction, WES library preparation and bioinformatics analysis to identify variants and genes that underlie this disorder. The analysis allowed the exclusion of variants in *NDP*, identified a convincing mutation in *LRP5* (which was verified using a cell-based functional assay), and provided a list of potential genes which may play a role in Coats disease.

3.3.1 Challenges of working with DNA from archival FFPE tissues

Key to the success of this study was the extraction of good quality DNA from FFPE tissues. There are many challenges that are associated with working with DNA extracted from FFPE samples ranging from the time of fixation, fixation procedure, storage of specimens, degradation of DNA due to fixation and the subsequent downstream analysis, whether it is for PCR or NGS technologies.

3.3.1.1 Issues associated with fixing specimens with formaldehyde

Tissues are routinely fixed with formalin, a saturated formaldehyde solution of 37%. Formaldehyde creates cross-links between nucleic acids and proteins, and crosslinks proteins together and is, therefore, well-suited for preserving cellular architecture for histological analysis, especially as it is fast acting, has short cross-linker lengths and can permeate tissues well. Unfortunately, formaldehyde is also capable of inducing DNA strand breakage, loss of bases and deamination of cytosine, and these processes cause problems when DNA

is analysed from FFPE tissue (Koshiba et al., 1993; Do & Dobrovic, 2015) (Figure 3.22).

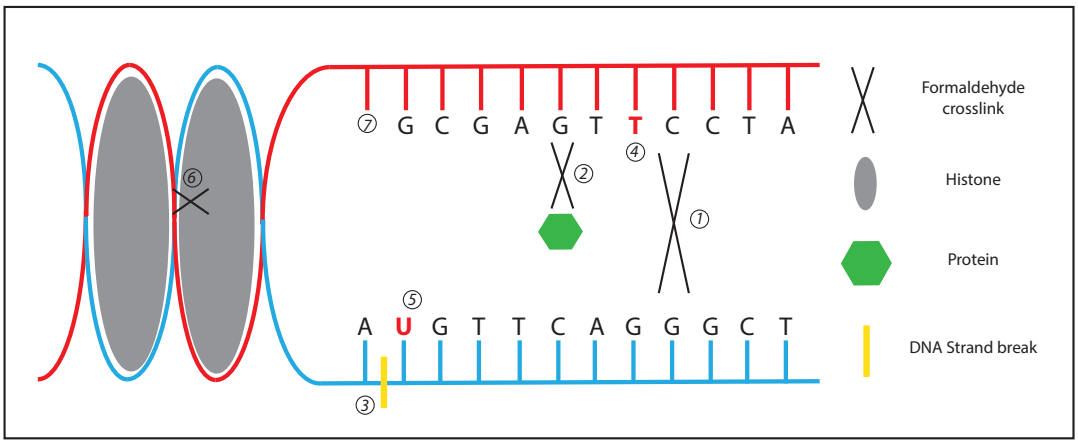


Figure 3.22 Schematic representation of DNA damage within FFPE samples. The fixation of samples can produce a number of damages to the DNA. These include: (1) DNA-DNA crosslinks, (2) DNA-protein crosslinks, (3) DNA strand-break leading to DNA fragmentation, (4) Thymine and (5) Uracil resulting from deamination of Cytosine, (6) histone-DNA crosslinks and (7) abasic sites where the nucleotides are lost. Figure adapted from Do & Dobrovic (2015).

DNA fragmentation is a common form of damage caused by fixing tissues with formalin (Wong et al., 2013). Studies by Ludyga et al. (2012) and Wong et al. (2014) showed that the amount of fragmentation depends on the storage time and also on the acidic pH of the formalin used in the fixing process. However, every FFPE tissue sample shows some degree of DNA fragmentation (Wong et al., 2014). It is therefore crucial to be aware of this complication when designing experiments (Do and Dobrovic, 2015). The short-length reads associated with NGS technologies make FFPE extracted DNA very amenable to this methodology, however, confirmation of the variants by Sanger sequencing proved to be problematic. The Coat samples used for the current study were from a historical archive and were approximately 25 years old. In a study analysing samples of a similar age, PCR analysis of 365 FFPE-extracted DNAs found that the longest amplifiable fragments were 152bp in 69% of

samples, 268bp in 17% of samples, 676bp in only 5% of samples and in 31% no fragments of any length were successfully amplified (Gillio-Tos et al., 2007). Similar results to these were found with the FFPE-extracted DNA from this study, with successfully amplified products only ranging from 69bp to a maximum of 170bp (Section 3.2.4).

Another issue related to formaldehyde treated samples is the difficulty in extracting good quality DNA (Crisan and Mattson, 1993) due to the formaldehyde cross-linking DNA and proteins. Although reversible, it is often impossible to remove all of the cross-links and it has been shown that as little as 2.5% of cross-links need to remain to affect the PCR amplification of products over 200bp (Karlsen et al., 1994). Non-amplifiable DNA was a major problem in this study, especially in the first round of DNA extraction. However, extended digestion with proteinase K has been shown to increase the breakage of cross-links to yield amplifiable DNA (Jackson et al., 1990). In this study, for the second round of DNA extractions, the proteinase K digestion time was increased and may have contributed to the better quality DNA extracted (Section 3.2.7).

A further issue associated with formaldehyde fixation of tissue is the creation of abasic bases, which are locations in DNA that have neither a purine or pyrimidine base (Do and Dobrovic, 2015). Abasic bases are caused by the hydrolysis of the *N*-glycosidic bonds of the purine bases to the sugar backbone (Suzuki et al., 1994). The hydrolysis of the bonds is caused by formalin which can’t maintain a neutral pH level. Abasic bases can also cause destabilisation

of the DNA double helix which leads to further degradation of DNA (Vesnaver et al., 1989).

The deamination of cytosine to uracil is another issue caused by using formaldehyde as a fixative. The unavailability of the repair mechanism to replace the incorrect uracil base means that they are amplified in the PCR reaction and cause many sequencing artefacts (Do and Dobrovic, 2015). The clonal nature of NGS makes it easier to spot these artefacts as they usually only affect a small proportion of the reads. However, there are many regions of the exome that are difficult to sequence, such as CpG islands, and these tend to have lower depth of coverage in WES data. In these situations, it is often difficult to determine if a variant is an artefact or real, as evidenced by the *ATOH7* artefact identified in this study (Section 3.2.8).

The preparation and storage of tissues prior to fixation also needs to be considered as this can further contribute to DNA damage (Hedegaard et al., 2014). The delay in fixation of samples is a huge factor that damages DNA. This time is known as the perioperative ischemic time and it can cause serious damage to freshly harvested tissues. Clearly, the preparation and fixation process play an important role in preserving the integrity of the DNA within a tissue. These processes should, therefore, be performed according to highly optimised and widely used protocols and done in an appropriate time frame. The cancer diagnostics field routinely extracts DNA from FFPE tissues and have, therefore, optimised tissue harvesting and fixation protocols for this reason. Unfortunately, due to the samples in this study being a historical

archive and fixed purely for histological reasons, the careful and efficient procedure of fixing was not considered, and this greatly impacted on the quality of the DNA extracted. As a result, a great deal of optimising was needed to establish a suitable DNA extraction protocol that provided good quality amplifiable DNA suitable for successful WES.

3.3.2 Optimising the FFPE DNA extraction protocol

There are many ways of recovering DNA from FFPE samples and the approaches range from standard methods using phenol-chloroform to a variety of commercial kits (Hedegaard et al., 2014; Heydt et al., 2014; Janecka et al., 2015; Ludyga et al., 2012; Pikor et al., 2011; Schweiger et al., 2009; Sengüven et al., 2014; Skage & Schander, 2007; Wong et al., 2014; Yost et al., 2012). The kit for DNA extraction used in this study was the QIAamp DNA FFPE Tissue Kit (Qiagen). This is a commercially available kit that has been widely reported in the literature to yield high DNA concentrations (Hedegaard et al., 2014; Janecka et al., 2015; Ludyga et al., 2012; Sengüven et al., 2014). Whilst the use of this kit successfully generated enough DNA for the initial WES experiment using manufacturer's instructions, modifying the protocol by prolonged deparaffinisation and extending the proteinase K incubation times, not only allowed for greater quantities of DNA to be extracted from the blocks, but increased the quality of the extracted DNA (Turashvili et al., 2012; Sengüven et al., 2014; Janecka et al., 2015). This facilitated in the generation of good quality DNA suitable for generating high quality WES data. Unfortunately, the DNA was still heavily fragmented, and this hampered attempts to verify variants by Sanger sequencing.

3.3.3 PCR amplification of FFPE-extracted DNA

PCR amplification of fragmented DNA has been shown to be difficult in many studies. In research presented by Dietrich et al. (2013), it was shown that problems with amplification arise not only from the inability of DNA polymerase to amplify the DNA due to fragmentation but also from the low number of available templates that are amplifiable. Dietrich and colleagues overcame these difficulties by increasing the concentrations of the polymerase, dNTPs and additionally, the time of the elongation step during amplification. Some of these steps were applied in this study but with limited success, especially as increasing the amount of input DNA was often problematic due to its limited supply.

To overcome this problem, WGA was used to increase the amounts of template DNA. In this study the GenomiPhi V2 kit (G.E Healthcare) was used with some success. However, there are newer kits available which are aimed at amplifying small samples such as the REPLI-g kit (Qiagen) and the GenomePlex® kit (Sigma). Although both of these kits have been successfully used for WGA for NGS technologies, the GenomePlex® kit has been shown to perform the best (Mendez et al. 2017).

Optimising PCR primers and reaction conditions can use up a great deal of DNA and as this was in limited supply for this project, attempts to verify variants using Sanger sequencing were abandoned. As an alternative, a further extraction of FFPE-DNA will be performed and subjected to WES. The

artefacts are unlikely to be replicated so this will allow the true variants to be verified.

Recently, there have been a number of DNA repair enzymes that have widely been used in the literature to repair FFPE-extracted DNA. During the next DNA extractions, a repair enzyme step will be included to establish whether it improves the amplification of the FFPE-DNA. One of these repair enzymes is uracil-DNA glycosylase (UDG), which removes the incorrect uracil base when the deamination of cytosine to uracil occurs naturally in living cells. This repair mechanism does not occur in fixed cells (Do and Dobrovic, 2015) but they can be treated with UDG prior to PCR amplification to reduce the number of transitions (Do & Dobrovic, 2012; Do & Dobrovic, 2015). Another example is Restorase™ DNA polymerase (Sigma-Aldrich), which consists of AccuTaq DNA polymerase and a DNA repair enzyme. Restorase™ corrects strand nicks, depurinated sites and base modifications, allowing the amplification of damaged DNA (Skage and Schander, 2007). The Restorase™ repair enzyme is reported to simulate the nucleotide excision repair (NER) mechanism seen in living cells (Costa et al., 2003). Similarly, a kit called PreCR Repair Mix (NEB) has been used for the same types of DNA damage (Marosy et al., 2017). Another repair enzyme called uracil-specific excision reagent (USER) (NEB) comprises UDG, which excises the incorrect uracil base creating an abasic site (Lindahl et al., 1977), and endonuclease VIII, which breaks the phosphodiester backbone at the 3’ and 5’ sites of the abasic base and releases the base-free deoxyribose leaving the complementary base exposed (Jiang et al., 1997). This allows for the repair and addition of the correct nucleotide.

3.3.4 Optimising WES library preparation with FFPE DNA

Creating sequencing libraries from FFPE samples is another challenge that comes with preserving fresh tissues with formalin. The standard WES protocol requires 3ug of starting DNA (Agilent). However, for difficult or precious samples such as FFPE, Illumina has devised a protocol for low input samples using only 200ng. This was the amount of DNA used to generate the WES libraries in this study, but publications have reported successful preparations with as little as 16ng of FFPE-extracted DNA (van Allen et al., 2014).

For the first round of WES, 150ng-1000ng of DNA was added but with hindsight it is clear that not all of this DNA was amplifiable. The incorporation of the NGS QC FFPE Kit made a huge impact on the quality of the second run of WES (Section 3.2.7.1). The QC FFPE kit results showed that only 12-17% of DNA was amplifiable (Table 3.8) and this probably accounts for the lack of peaks visible in the post shearing bioanalyser results in the first WES library preparations (Figure 3.3) and contributed to the poor quality WES data.

Although not detailed in this thesis, additional attempts were made at preparing the WES libraries with the FFPE-extracted DNA but these have failed. This variability in success has been seen in other similar studies. For example, in a study by Hedegaard et al. (2014), 29.5% of FFPE samples prepared for WES were successful but the remaining 70.5% of samples failed the library preparation.

Another modification made to the final WES library preparation protocol was not to combine the NEBNext® Ultra DNA Library Prep Kit with the SureSelect^{XT} v5 kit, as originally recommended by the Leeds Next Generation Sequencing Facility (Section 3.2.2). During the first library preparation, the two kits were merged at the PCR amplification step, and the NEB master mix and the SureSelect^{XT} primers were combined. This is likely to have contributed to the under-amplification of the samples. The second WES library was created using all the components of the commercial kit SureSelect^{XT} 200ng kit (Agilent) without interchanging reagents (Section 3.2.7.2).

3.3.5 The comparison between the sequencing quality of FFPE exomes

The QC results from the first and second WES experiments have shown some significant differences. The results presented in Figure 3.21 for both library preparations show that the quality scores and the number of sequence reads were of higher quality for the second run of WES in comparison to the first; the quality scores were above 40 Phred score on average for the second run, in comparison to 30 for the first. The sequence duplication level was also improved upon between the first and second WES runs with a 4% reduction in the number of detected duplicates for sample 71208 and up to 15% for other samples (Appendix 7.2). Nearly 42% of all sequences were only read once during the second WES. Overall, this QC data shows the significant improvement upon the input DNA and the library preparation for the second run of WES.

When comparing the depth of coverage between the sets of sequencing data (Table 3.2 and Table 3.12), the improvement in percentage of bases covered 5, 10 or 20 reads was substantial. All samples had above 91% of their bases covered at least 20 reads in the final data set, whereas in the first the average was only 45% and three samples had less than 10% of bases with 20x coverage. This final WES data showed an excellent depth of coverage that was comparable with the non-FFPE exome (Chapter 4, Section 4.2.1).

In this study, seven FFPE samples were run on the same Illumina sequencing lane. This is a lower number of exomes than usual, as for non-FFPE exomes 10-16 samples are routinely run on the same lane. Increasing the sequencing coverage for FFPE exomes has been shown to be a good way to remove any low covered artefacts and minimise the number of false positives (Schweiger et al., 2009). Running as little as three exomes over two lanes has been shown to increase the data quality (Hedegaard et al., 2014). In this study, running the seven samples over more lanes may be beneficial, but this would cost a great deal. Furthermore, the overall depth of coverage will be increased by performing the third round of WES on newly extracted DNA, with the added benefit of reducing further artefacts.

3.3.6 The bioinformatics analysis of FFPE exomes

The sequencing data were analysed with an in-house pipeline that is routinely used for non-FFPE exomes (El-Asrag et al., 2015; Panagiotou et al., 2017). Since there was no experience in analysing FFPE exomes and accounting for the common sequencing artefacts, it was a very promising route of analysis

that needed to be optimised. The full in-house pipelines (standard and fully optimised) are provided in Appendix 7.6.

Both sets of sequencing data were analysed with the same core pipeline, however, there were a number of differences to make the analysis easier and more efficient when considering FFPE samples (Section 3.2.7.3). The differences that were introduced into the second analysis were to remove as many artefacts caused by fixation and to provide more stringent filtering of variants. The changes included lowering the MAF from ≤1% in population to ≤0.01%. This was justified by the fact that Coats disease is a rare disorder (Morris et al., 2010). Variants were also selected based on having a high CADD score (Kircher, 2014) (Section 2.19.4.1). Variants were filtered to select those with a CADD score of 10 or above. High allele depth was another criterion. This was especially important for FFPE samples as there is a high percentage of sequencing artefacts as already described. There was no clear cut-off set for the allele depth as this is very difficult to establish with FFPE samples. However, allele depth >10 per allele was mainly considered. For heterozygous alleles, an allele depth ratio in the region of 50:50 was considered optimal. Further criteria applied included being rare or absent in online variant databases. If the variant was rare, all of the loss of function (LoF) variants across that particular gene were checked for in the databases. The criteria required limited or no LoF variants across the candidate gene to strengthen the deleteriousness of the chosen variant. This analysis was very subjective and difficult as some genes contain known LoF pathogenic variants

in the databases (for carriers of recessive alleles and late onset disorders or mild dominant disorders).

Increasing the number of unrelated controls to 57 was another filtering step that aided in reducing the common SNPs and artefacts introduced by the library preparation. It is also important to note that the controls were not FFPE and were treated as normal samples. The data is documented per patient in Table 3.16- Table 3.22 (Section 3.2.8).

Analysis of shared variants was also carried out to narrow the candidate list. However, this investigation did not yield any potential candidates so the analysis for shared genes with different variants was performed. This analysis produced a number of candidate genes out of which the candidates with the highest CADD scores were selected (Table 3.15, Section 3.2.8). In addition, variant lists were also drawn up for each individual case (Table 3.16 - Table 3.22).

In the analysis of the variants, the primary criterion was to prioritise heterozygous variants. As Coats disease is thought to be caused by somatic mutations (Black et al., 1999), it would have been extremely unlikely that the same random change would have occurred in the same position on both alleles. However, there is the possibility that Coats mutations can be germline, as evidenced in the *RCBTB1* study (Wu et al., 2016) and discussed in more detail in Section 5.2, so homozygous variants were also investigated but there were no candidates.

The above steps were taken to improve upon the analysis of WES FFPE exomes. This advanced analysis has aided in easier identification of unwanted artefacts and improved the search of potential candidate variants and genes.

In this study, no special scripts were used to try to remove the low-level sequence artefacts commonly found in FFPE DNA, which can lead to false-positives (Wong et al., 2014). However, other teams have developed methods to do this. For example, Yost et al. (2012) devised a filtering method that concentrated on the removal of false positives by comparing the global nucleotide mismatch rates and the local mismatch rates. Using this method, whole-genome sequencing (WGS) was performed on FFPE cancer tumours >11 years of age and somatic mutations were identified (Yost et al., 2012). To achieve this, the study used an automated set of custom post-alignment filters to remove false positives somatic variants. The filters included the removal of genomic positions that were of low quality or coverage, removal of previously documented variants in online databases, removal of variants that were present in the corresponding germline sample, the removal of duplicates that result in false positives and the final filter included the removal of variants that had elevated local mismatch rate. By using this strategy, the pipeline removed around 76% of variants that were deemed to be artefacts or false positives. Their filtering method does not use a threshold for substitutions but instead uses the mismatch error rate across the genome of the given sample (Yost et al., 2012).

Other studies have also shown that a more stringent alignment can aid in removing unwanted artefacts that were generated from the DNA damage. This also carries risks as it could potentially remove any true mutations with a lower depth of coverage (Xuan et al., 2013). Stringent alignment was not used in this study to prevent removing any potential variants, but more stringent filtering was used instead (Section 3.2.7.3).

3.3.7 Analysis of candidate Coats disease variants

As the known FEVR genes were excluded from each Coats patient, the resulting lists of variants are likely to contain novel Coats disease genes. Some of the variants presented in Table 3.16 - Table 3.22 were put forward to be confirmed by Sanger sequencing. However, the amplification of the patients' DNA was shown to be very difficult and was often unsuccessful due to its poor-quality.

The two variants that were selected for Sanger sequencing included *ATOH7* c.46delG; p.(Ala16Hisfs*71) and *ISPD* c.206_207delGC; p.(Cys69Serfs*46). Both of these variants appeared to be heterozygous and were found in single patients.

ATOH7 was a good candidate as it is a known arFEVR gene and the encoded protein is involved in the development of the eye, specifically in the development of the optic nerve and RGCs (Brown, et al., 2002; Khan et al., 2012). In 2012, *Atoh7* mutant mice were described to lack the intrinsic retinal vasculature and the regression of foetal blood vessels was halted due to the absence of trophic factors that are secreted by RGCs (Prasov et al., 2012). Kondo et al. (2016) identified mutations in *ATOH7* in patients with

nonsyndromic congenital retinal nonattachment and FEVR. These findings make *ATOH7* a good candidate for Coats disease. Unfortunately, Sanger sequencing failed to confirm this variant (Section 3.2.8). However, given that the template DNA was derived from WGA DNA and the GC-rich nature of *ATOH7* makes it hard to amplify gDNA, it is possible that both alleles of this gene were not sequenced so this variant warrants further analysis before it is completely excluded.

The second candidate chosen for verification was *ISPD* as mutations in this gene are known to cause Walker-Warburg syndrome, which is characterised by muscle, brain and eye abnormalities. The eye conditions include microphthalmia, cataracts and enlarged eye balls due to increased intraocular pressure (Roscioli et al., 2012; Willer et al., 2012; Trkova et al., 2015). Ispd knockdown in zebrafish showed the same phenotypic features as Walker-Warburg syndrome, including hydrocephalus, reduced eye size and muscle degeneration (Roscioli et al., 2012). Microphthalmia can be present in severe FEVR-related retinopathies and cataract is common in Coats disease patients in advanced stages of the disease (Stage 5, Figure 1.7) (Shields et al., 2001a). Sanger sequencing of this variant was not possible, due to amplification problems, so at the moment it remains unconfirmed. However, the allele depth over this variant was very high and the mutation is very convincing (140/113) so it remains a compelling candidate.

During the course of this study, heterozygous germline mutations in *RCBTB1* were identified in patients with Coats disease and FEVR (Wu et al., 2016) and

subsequently bi-allelic mutations were identified in patients with isolated and syndromic retinal dystrophy (Coppieters et al., 2016) (Section 1.8.2.9). Based on these findings, *RCBTB1* was checked to see if it was present in any of the mutations lists and in the CNV lists. However, no variants were found in this gene.

The confirmation of the other variants awaits the third round of WES on freshly extracted DNA. This will allow for all the candidate variants and genes to be confirmed and further investigated with functional assays to elucidate their role in the retinal vasculature and Coats disease. Matched gDNA for the patients is currently being collected, with two samples already received. These genomic samples will also undergo WES in the future and this will allow somatic variants to be distinguished from germline variants which will greatly aid the interpretation of the candidate variants.

Finally, as all three genes implicated in Coats disease (*NDP*, *RCBTB1* and *LRP5*) to date are known FEVR genes, a cohort of FEVR patients was sequenced to try and identify novel FEVR genes which could be used to help pick candidates out of the existing Coats disease lists (Chapter 4).

3.3.8 Identification of *LRP5* as a new Coats disease gene

In this study, a heterozygous missense variant in *LRP5*, c.2951A>G; p.(Tyr984Cys) was identified in the eye tissue of one Coats disease patient. The variant was very rare in online variant databases, such as ExAC and gnomAD (Table 3.5) and was conserved down to zebrafish. It was also deemed deleterious by pathogenicity prediction tools (Table 3.7). To add

further evidence for the pathogenicity of this variant, a cell-based functional assay (TOPflash) was performed. This confirmed that LRP5 harbouring the missense mutation was unable to activate the reporter for Norrin-β-catenin signalling to the same level as WT LRP5 (Section 3.2.6.3). These findings argue strongly that this *LRP5* mutation is pathogenic.

LRP5 is a key component of the Norrin-β-catenin signalling pathway. Although FZD4 alone is sufficient to activate the signalling pathway, the addition of the co-receptors LRP5 and TSPAN12 leads to greatly increased signalling levels (Ke et al., 2013). Furthermore, mutations in all three of these receptors, and the ligand Norrin, lead to a spectrum of overlapping eye disorders termed "FEVR-related retinopathies" (Section 1.8) showing that all components of this ligand-receptor complex are essential for normal eye development. Mutations in *NDP*, which encodes Norrin, have previously been shown to underlie Coats disease (Black et al., 1999). The identification of an *LRP5* mutation in the current study cements the importance of this pathway in Coats disease and confirms that Coats disease is genetically related to the FEVR-related retinopathies and is caused by defective angiogenesis (Richter et al., 1998; Robitaille et al., 2002; Kato et al., 2002; Junge et al., 2009).

LRP5, and the closely related LRP6, belong to the family of low-density lipoprotein receptors (LDLR) which are cell surface receptors responsible for endocytosis by binding and internalising ligands (Willnow et al., 1999). LRP5/6 are single-span transmembrane proteins with an extracellular region composed of four β-propeller domains, each consisting of six YWTD β-

propeller repeats (Springer, 1998). An epidermal growth factor-like (EGF) domain connects the β-propellers. The YWTD-EGF repeats contain the binding sites for many ligands (Davis et al., 1987) including Norrin, Wnt, Dickkopf-1 (Dkk1), SOST and WISE (Itasaki, 2003; Liu et al., 2009; Mao et al., 2001; Semënov et al., 2005; Semënov et al., 2001; Xu et al., 2004). These are followed by three LDLR type A domains. The role of these LDLR repeats within the Norrin-β-catenin signalling pathway remains unknown (Macdonald and He, 2012) but they have been reported to facilitate the dimerization of LRP6 to aid Wnt-β-catenin signalling (Chen et al., 2014). The intracellular segment of LRP5/6 consists of PPPSP motifs which act as docking sites for Axin after phosphorylation (Tamai et al., 2004).

There have been many mutations described in *LRP5* over the years causing predominantly eye and/or bone phenotypes (Gong et al., 2001; Boyden et al., 2002; Toomes et al., 2004b; Jiao et al., 2004; Qin et al., 2005; Semenov and He, 2006; Narumi et al., 2010; Fei et al., 2014; Kramer et al., 2015; Seo et al., 2015; Pefkianaki et al., 2016; Zhang et al., 2016; Ergun et al., 2017; Rao et al., 2017; Huang et al., 2017; Tang et al., 2017; Lin et al., 2018). The LRP5 variant identified in this study, p.(Tyr984Cys), is located in the 4th YWTD β-propeller domain. Published mutations in this domain include a missense mutation in an FEVR patient (p.Tyr1168His) that was predicted to cause destabilisation of the protein from molecular modelling of the YWTD-EGF domain (Rudenko, 2002; Toomes et al., 2004a). Another mutation in this domain was reported in a case of primary osteoporosis (p.Arg1036Gln) (Hartikka et al., 2005). In another study, WES revealed a mutation (p.Arg1188Trp) in a patient with polycystic

liver disease. The change of arginine for tryptophan was predicted to disturb the whole YWTD β-propeller structure due to the loss of the hydrogen bonds and ionic interactions within the β-propeller created by the arginine amino acid (Cnossen et al., 2014).

There are a few potential ways in which the variant identified in this study could disrupt protein function. For example, it could disrupt the correct folding or trafficking of the protein or alter the binding of ligands or LRP5's interactions with co-receptors. The p.(Tyr984Cys) changes a hydrophobic aromatic tyrosine for a neutral, sulphur containing cysteine. Cysteine is a small, neutral (as it is partly hydrophobic and partly polar) amino acid. Its role within a protein is to form disulphide bonds, which in turn stabilises the protein and its structure (Betts and Russell, 2003). The newly introduced cysteine residue could try to create a disulphide bond, which could change the structure and/or the stability of the protein and hence its function.

Alternatively, the change of a hydrophobic tyrosine to a neutral cysteine could potentially affect the role that the domain may have in binding proteins. MESD, an LRP5/6 chaperone, binds to the 4th YWTD-EGF domain to ensure the correct folding of LRP5/6's β-propellers-EGF domains (Liu et al., 2009). Without MESD, LRP5/6 would not localise to the cell membrane and would be retained within the ER (Hsieh et al., 2003). Similarly, the Wnt antagonist Dkk1 is known to bind to the 4th YWTD-EGF domain (Ahn et al., 2011; Boyden et al., 2002; Semenov & He, 2006). The antagonistic effect of this ligand is mediated by Dkk1 simultaneously binding LRP5 and Kremen and triggering endocytosis

of LRP5. It is reasonable to assume that such an action is also likely to affect Norrin signaling too. The vast majority of other ligands which interact with the extracellular domain of LRP5 are not reported to physically interact with the 4th YWTD-EGF domain of LRP5. These include SOST (Semenov and He, 2006), WISE (Itasaki, 2003) and the majority of the Wnts (Xu et al., 2004). However, Wnt3a is reported to bind to the 3rd and 4th YWTD-EGF domain of LRP6 (Bourhis et al., 2010) and may therefore bind to the same region of LRP5.

Potentially the most relevant interaction to this study is that between Norrin and LRP5. This interaction was previously described to be crucial for the Norrin-β-catenin signalling pathway (Xu et al., 2004; Ke et al., 2013). However, no direct interaction has been documented between alkaline phosphatase (AP)-tagged Norrin and Lrp5/6 in the study by Xu et al. (2004). Later, in a study by Ye et al. (2010), the binding between Norrin and Lrp5 was still not confirmed. Eventually, in a study by Ke et al. (2013), the structure-based mutagenesis of Norrin showed that Norrin contains separate binding sites for Fzd4 and Lrp5/6. Maltose binding protein (MBP)-tagged Norrin was shown to increase the signalling activity from 2 to 6-fold when Lrp5 was co-transfected along with Fzd4, compared to Fzd4 acting with Norrin alone. MBP-Norrin was shown to interact with YWTD-EGF domains 1 and 2 and this interaction was not disrupted by the binding of Fzd4 to Norrin. The different binding sites for Fzd4 and Lrp5/6 were confirmed by introducing mutations into Norrin (C55A and C110A) which affected the interaction of Norrin to β-propeller 1-2 (Ke et al., 2013). The next study that elucidated the binding between Norrin and LRP5 was that by Chang et al. (2015), who suggested putative binding sites for

Lrp5/6 at the following Norrin residues Lys54, Arg90, Arg97, Gly112, and Arg121. These putative binding sites showed a small overlap with the binding site predicted by Ke et al. (2013). If all this data is correct, the missense mutation identified in this study is not predicted to have any direct effect on Norrin binding but must be influencing this pathway through some other mechanism.

The cell-based assay used to investigate the *LRP5* missense mutation was the TOPflash reporter assay developed by Molenaar et al. (1996). This assay measures β-catenin signalling by using a luciferase reporter construct and has been widely used in many studies to assess the deleteriousness of variants in the Wnt-β-catenin and Norrin-β-catenin signalling pathways (Chang et al., 2015; Fei et al., 2014; Junge et al., 2009; Qin et al., 2008; Xu et al., 2004; Smallwood et al., 2007). In this study, the TOPflash assay was performed using the same methodology described by Xu et al. (2004) with the addition of TSPAN12 (Junge et al., 2009). The assay was performed in a cell line with the TOPflash construct stably integrated (STF cells). However, the experiment still relied on the successful co-transfection of five different expression constructs and this complexity led to variability in the results. These variations are routinely seen in similar studies (Junge et al., 2009; Qin et al., 2008; Soriano, 2017). However, all of the biological replicates generated similar trends and the cell viability was consistent. The highest TOPflash activity was obtained when all the wild-type Norrin-β-catenin pathway components were transfected in to the cells (Figure 3.16, first column). This was consistent with results found in other studies (Junge et al., 2009; Ke et al., 2013). Very little pathway

activation was detected when none of the components were transfected which served as a control for quantification of background levels of TOPflash activation by the endogenous proteins (Figure 3.16, fourth column). The *Renilla* construct served as internal transfection control. When no LRP5 (WT or mutant) was transfected into the STF cells, there was a significant decrease in the pathway activation (p value $\leq$0.0001) in comparison to all WT components transfected (Figure 3.16, third column). The results of other studies showed the same findings (Qin et al., 2008; Junge et al., 2009; Ke et al., 2013). Upon transfection of the LRP5 mutant p.(Tyr984Cys), a significant decrease in the activation of the Norrin-β-catenin pathway was detected (p value $\leq$0.05) (Figure 3.16, second column). To overcome the variability of this assay, numerous repetitions were performed, and all individual values were compiled and presented as a single data set. It's worth pointing out that this cell-based assay has limitation and might not detect the full consequences of certain types of mutation.

The effects of missense mutations on *LRP5* function using TOPflash assays have been performed before. In a study by Qin et al. (2008), six missense variants (2 mutants and 4 SNPS) were assessed with the overexpression of WT Norrin and FZD4. TSPAN12 was not transfected as it wasn't identified as part of the receptor complex until 2009 (Junge et al., 2009). The two mutants have shown a 45% and 26% reduction in the Norrin-β-catenin pathway activation for p.R444C and p.A522T respectively. These reductions were significant with a p value $\leq$0.05. This was similar to the results in this study, where an 18% reduction was noted in the mutant comparison to the WT

(Section 3.2.6.3). In a similar study, five LRP5 missense mutants, and two double mutants, were seen to significantly reduce TOPflash activation in comparison to the WT. Six of the seven mutants showed a decrease with a p value ≤ 0.05. However, one mutant did not significantly decrease the signalling activity illustrating the limitations of this assay (Fei et al., 2014).

LCI was attempted to investigate the effect of the p.(Tyr984Cys) variant on the cellular localisation of LRP5 (Section 3.2.6.4). Unfortunately, numerous difficulties were encountered including toxicity from the GFP fusion protein and cell attachment issues. Attempts to overcome these issues were made but time-limitations meant that this experiment was not completed.

An alternative method that can be used in the future is immunohistochemistry, and the constructs made in this study will be applicable to this method. Another assay that can be used in the future to assess the effect of the mutation on the localisation or interactions of LRP5 is the proximity ligation assay (Fredriksson et al., 2002). Commercial forms of this assay are available such as the Duolink® system (MERK). This assay is based on incubating fixed cells with two primary antibodies (raised in different species) which are either both specific to LRP5 or specific to LRP5 and one of its interactants. The secondary antibodies are fused to oligonucleotide probes. If the two antibodies are in close proximity, the oligonucleotide probes can be ligated to a complimentary connector oligonucleotide to form a circular DNA template. Upon the addition of a polymerase, the ligated circular template will be amplified to generate a product which can be detected, and if necessary, amplified by a

complementary labelled probe. This assay has been widely used in studies (Jiang et al., 2014; Lai et al., 2017; Söderberg et al., 2006).

Any data generated using a cell-based assay with over-expressing proteins is always going to be compromised, especially experiments like proximity ligase assays. It's always better to investigate endogenous proteins. Traditionally, these have only really been available if patient cells are available. However, with the availability of CRISPR/Cas9-based technologies it is now possible to recreate mutations in WT endogenous proteins (Komor et al., 2017). This will allow very precise analysis of the effects of this LRP5 mutation, and any other Coats genes identified in the future.

3.4 Summary

In this study, a protocol was optimised to extract quality DNA and generate high-quality exome data for a historical archive of FFPE enucleated Coats eyes. WES analysis was undertaken on eyes from seven unrelated Coats patients and mutations in the known Coats gene, *NDP*, were excluded. Candidate variants were identified for all eyes, but no convincing common candidate gene was identified. In one eye, a heterozygous missense variant in the *LRP5* gene c.2951A>G; p.(Tyr984Cys) was identified. This variant was predicted to be disease causing, it was conserved and rare. Functional assessment of this variant using the TOPflash assay showed it reduced the activation of the Norrin-β-catenin signalling reporter. Therefore, the evidence suggests this variant is likely to be a real mutation and identifies *LRP5* as a new gene underlying Coats disease.

4 Chapter: Identification of variants and candidate genes in familial exudative vitreoretinopathy (FEVR)

4.1 Introduction

When this work was initiated, the following genes involved in the pathogenesis of FEVR were identified: *NDP* (Chen et al., 1993a), *FZD4* (Robitaille et al., 2002), *LRP5* (Toomes et al., 2004b & Jiao et al., 2004), *TSPAN12* (Nikopoulos, et al., 2010; Poulter et al., 2010), *ATOH7* (Khan et al., 2012), *ZNF408* (Collin et al., 2013), *KIF11* (Robitaille et al., 2014) and *RCBTB1* (Wu et al., 2016). Out of the established FEVR genes, *LRP5*, *FZD4*, *NDP*, and *TSPAN12* encode proteins which have all been found to be involved in the Norrin-β-catenin signalling pathway (Section 1.8). FEVR is phenotypically variable and genetically heterogeneous with autosomal dominant, autosomal recessive and X-linked inheritance. Previous screening studies of FEVR patients have shown that only around 50% of cases have detectable mutations in the known FEVR genes (Salvo et al., 2015). Therefore, in this study, a panel of FEVR patients was analysed using WES with the aim to identify novel mutations and genes implicated in this disorder. Furthermore, as all the genes implicated in Coats disease to date are known FEVR genes, the aim was to also use this data to aid in the identification of new Coats genes from the data generated in Chapter 3.

4.2 Results

This study focused on 20 unrelated FEVR patients without a molecular diagnosis. The majority of patients presented with “classic” FEVR or disorders within the FEVR phenotypic spectrum such as retinal detachment or vitreoretinal dysplasia. Two patients presented with additional features that do not belong to the FEVR phenotypic spectrum; one with FEVR plus a cleft palate and another with FEVR and congenital skeletal malformations and a ventricular septal defect. One patient was diagnosed with Coats disease; they presented with bilateral disease with the left eye containing exudates and abnormal vasculature resembling Coats disease while the right eye showed FEVR-like peripheral avascular regions. Two patients were diagnosed with FEVR/ROP. Both of these cases were born prematurely (33/34 weeks), but not very earlier as seen in most cases of ROP (<28 weeks), and both have severe disease and came from consanguineous parents. The majority of patients had undergone some Sanger-based pre-screening of the known FEVR genes as detailed in Table 4.1 (*NDP*, *FZD4*, *LRP5*, *TSPAN12*, *KIF11*, *CTNNB1*) (Toomes et al., 2004b; Toomes et al., 2004c; Nikopoulos et al., 2010; Poulter et al., 2010, 2012; Robitaille et al., 2011; Khan et al., 2012; Collin et al., 2013).

Patient ID	Ethnicity/ Country of Origin	Familial/ Sporadic	Phenotype	Pre-screening
F1401	Northern European, UK	S	FEVR	Clinical exome – ve by Yorkshire Regional Genetics
F1394	Northern European, Czech Republic	F (affected mother and three cousins)	Retinal Detachment	None
F1343	Northern American, USA	S	Coats Disease/FEVR	None
F1308	Northern European, Danish	S	FEVR	*TSPAN12, KIF11, CTNNB1*
F1268	Northern European, UK	F (affected father)	FEVR	*NDP, FZD4, LRP5, TSPAN12*
F1261	South Asian ethnicity, UK	S	FEVR/ROP	*TSPAN12*
F1260	South Asian ethnicity, UK	S	FEVR/ROP, cleft palate	*KIF11, TSPAN12*
F1250	Northern European, UK	S	FEVR	*KIF11, TSPAN12*
F1233	Ethnicity unknown, New Zealand	S	FEVR	*TSPAN12* and partly screened for *FZD4* and *NDP*
F1235	Ethnicity unknown, USA	F (affected mother and maternal grandmother)	FEVR, congenital skeletal malformations and ventricular septal defect. Mother had detached retina.	*TSPAN12*
F1225	Northern European, UK	S	FEVR	*LRP5, FZD4, NDP, TSPAN12*
F1213	South Asian ethnicity, UK	S	FEVR	*TSPAN12, FZD4, CTNNB1,* partial *LRP5* and *NDP*
F1057_1072	Northern European, UK	S	FEVR	*TSPAN12, FZD4, NDP*

F1051	Northern European, Australia	S	FEVR	*LRP5, TSPAN12, FZD4, NDP*
F1039	Northern European, Danish	S	FEVR	*LRP5, TSPAN12, FZD4, NDP*
F1020	Northern European, UK	S	FEVR	*LRP5, TSPAN12, FZD4, NDP*
F164	Northern European, UK	S	Vitreoretinal dysplasia	*LRP5, TSPAN12, FZD4, NDP*
F1342*	Northern European, Belgium	F (no family data provided)	FEVR	*CTNNB1*
F1251*	Northern European, UK	S	FEVR	*TSPAN12*
F1375*	Northern European, Belgium	S	FEVR	*CTNNB1*

Table 4.1 Clinical details of the 20 unsolved FEVR cases.
Ethnicity, origin, phenotype and genes pre-screened are detailed for each individual. FEVR – Familial Exudative Vitreoretinopathy; ROP – Retinopathy of Prematurity. *Referring lab may have screened as part of a panel before being referred to Leeds.

4.2.1 WES and Bioinformatic analysis

The WES library preparation and the sequencing of the samples was carried out by the NGS Sequencing facility at the University of Leeds. The library preparation was carried using the SureSelect^{XT} Human All Exon v6 kit (Agilent) according to the manufacturer's instructions with a starting input of 200ng of DNA (Section 2.17). The 20 library preps were sequenced on an Illumina HiSeq 3000 as 150bp pair end reads, with 10 samples pooled per lane. The raw fastq files generated by the sequencer were first assessed for quality as described for the Coats disease samples in Chapter 3 (Section 3.2.3). All 20 patients passed the QC and the samples were of very high quality. An example of the QC for one patient is presented in Figure 4.1. The Phred score across all bases was >30 (Figure 4.1A) and the majority of sequencing reads had an average Phred quality score over 38 (Figure 4.1B). The remaining QCs for patients who had identified mutations are presented in Appendix 7.3.

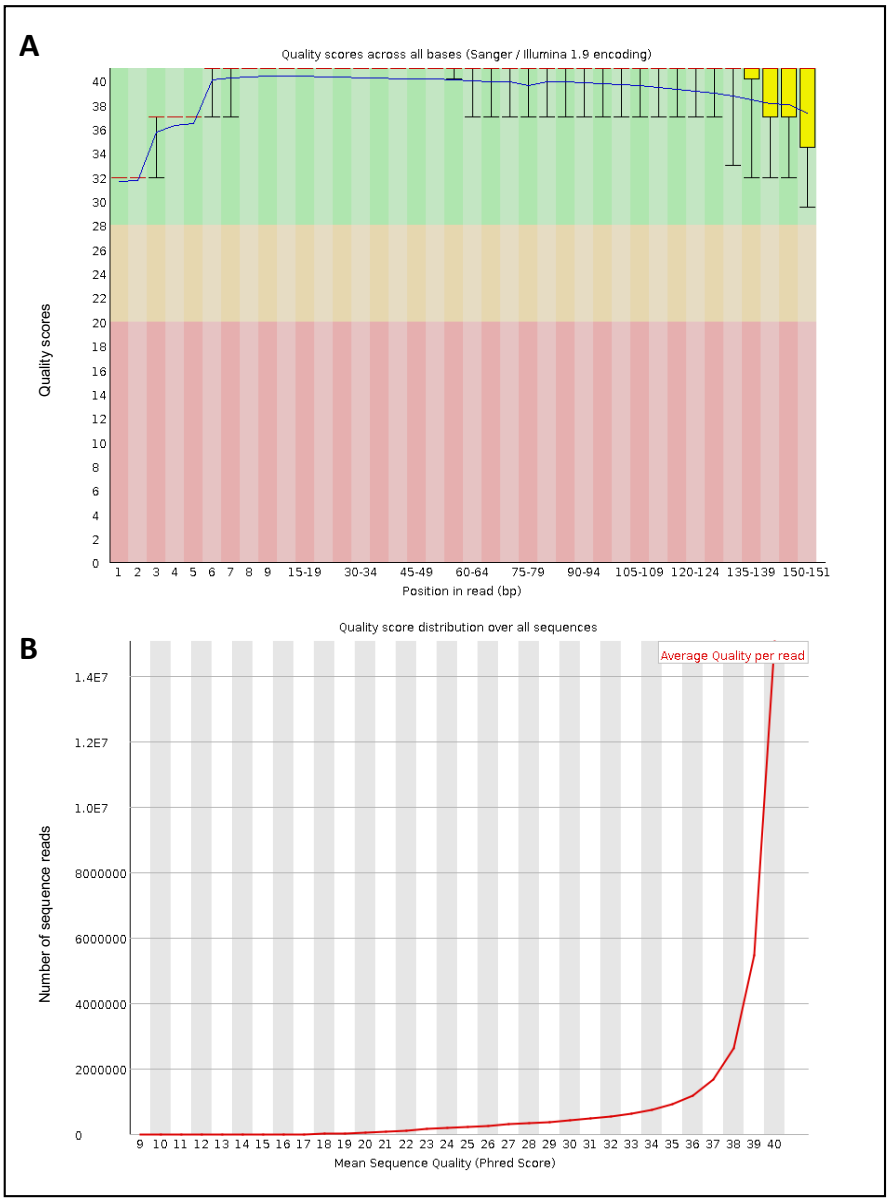


Figure 4.1 QC analysis of WES data for patient ID F1250.
(A) A representative of the QC of the raw fastq file for patient F1250 using a BoxWhisker plot. The plot shows the quality scores across all 150 bases for each read generated by Illumina HiSeq 3000. The x-axis represents the position in read in bp and the y-axis shows the quality Phred scores. Scores below 20 represents poor quality calls, scores between 20-28 represent calls of acceptable quality and the scores above 30 are very high-quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality. **(B)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample F1250. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads.

Depth of coverage was also calculated for all patients as described in Section 3.2.3 and the results are summarised in Table4.2. They show the percentages

of bases that had above 5, 10, 20 and 30 reads. All but one sample had over 90% of bases read above 30 times.

Sample ID	% of bases above 5	% of bases above 10	% of bases above 20	% of bases above 30
F1401	99.5	98.3	92.2	80.7
F1394	99.5	99.2	97.4	93.7
F1343	99.8	99.4	97.9	94.7
F1308	99.6	99.3	98.2	96.0
F1268	99.5	99.1	97.2	93.5
F1261	99.5	99.0	97.0	93.0
F1260	99.6	99.3	98.1	95.8
F1250	99.5	98.9	96.5	91.4
F1233	99.7	99.2	96.8	92.2
F1235	99.5	99.0	96.6	91.9
F1225	99.8	99.4	97.9	94.9
F1213	99.7	99.3	97.2	93.3
F1057_1072	99.8	99.5	98.3	96.0
F1051	99.8	99.4	97.8	94.7
F1039	99.8	99.5	98.0	95.0
F1020	99.6	99.4	98.7	97.1

F164	99.8	99.4	98.0	95.2
F1342	99.8	99.5	98.4	96.1
F1251	99.8	99.3	97.4	93.6
F1375	99.6	99.3	98.0	95.4

Table4.2 Depth of coverage for FEVR WES data.
Sample IDs are provided, and the corresponding percentage of bases covered above certain read depths are listed.

The WES bioinformatics pipeline used to analyse the data is presented in Figure 4.2 and the full details are listed in Appendix 7.7. The bioinformatics analysis was followed as previously described in Section 3.2.7.3 and variants with MAF above ≥0.01% in population were removed. All patients were analysed for dominant and recessive inheritance.

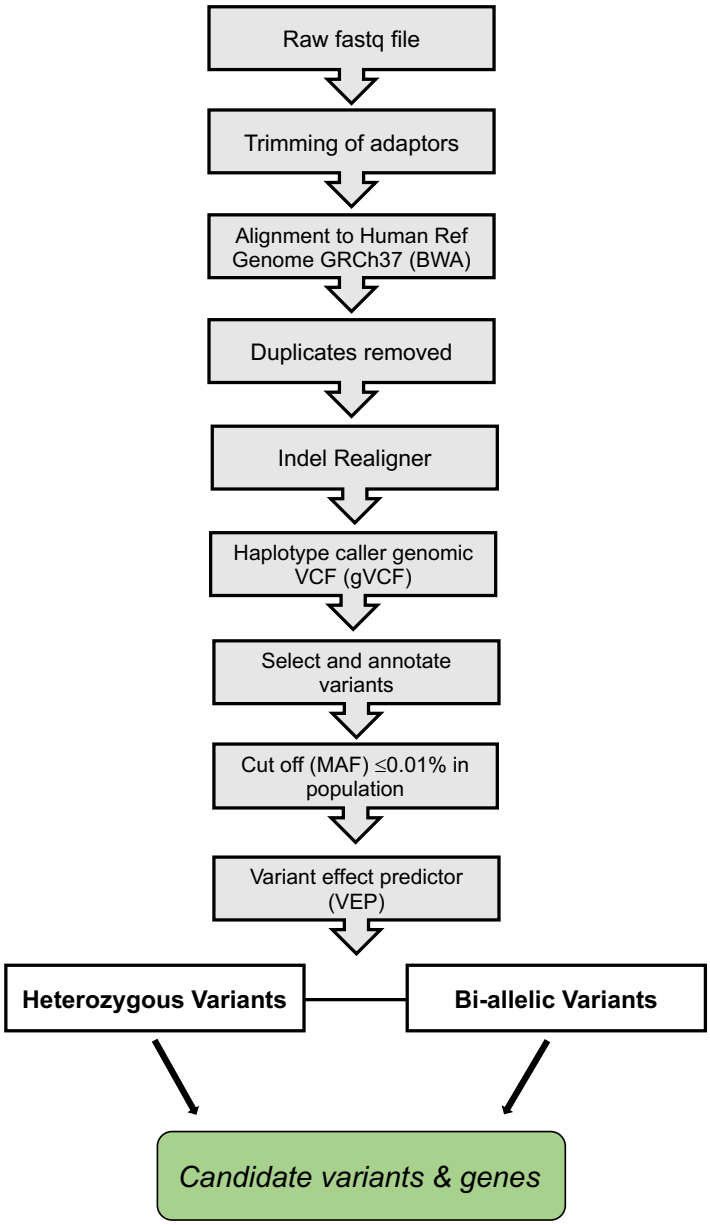


Figure 4.2 Schematic representation of the bioinformatics pipeline.
Flowchart describing the analysis pipeline workflow that was used to analyse WES of all FEVR patients with the aim to uncover new mutations and genes.

The known FEVR genes were searched using ExomeDepth as previously described in Chapter 3, Section 3.2.7.3.

For each patient, the known FEVR genes were investigated first and if no mutations were identified, then the data was analysed to look for dominant and recessive mutations in new candidate FEVR genes. All candidate variants were visualised on IGV (Section 2.19.10) and checked for their presence and frequency in online variant databases such as EVS, ExAC and gnomAD (Section 2.19.5-2.19.7). If appropriate, the conservation of the mutated amino acid residues were visualised using Homologene (Section 2.19.8). All remaining convincing mutations were confirmed by Sanger sequencing (Section 2.11.2).

4.2.2 Mutations in the known FEVR genes

5/20 samples were identified to carry convincing pathogenic variants in the known FEVR genes. The details of each are described below. The samples that were solved were the following: F1250, F1057_1072, F1375, F164 and F1308.

4.2.2.1 *LRP5* c.2254C>T; p.(Arg752Trp) & c.3914G>A; p.(Cys1305Tyr)

Compound heterozygous missense mutations in *LRP5* (NM_002335) were identified in a UK female with FEVR with no reported family history, identifier F1250. The first variant, GRCh37, Chr. 11:68177544, c.2254C>T; p.(Arg752Trp), was located in exon 10. The second variant GRCh37, Chr. 11:68201220, c.3914G>A; p.(Cys1305Tyr) was located in exon 18. Both mutations were visualised on IGV. The allele depth for the variants were wt/mut: C=71/T=58 and wt/mut: G=89/A=71 respectively (Figure 4.3).

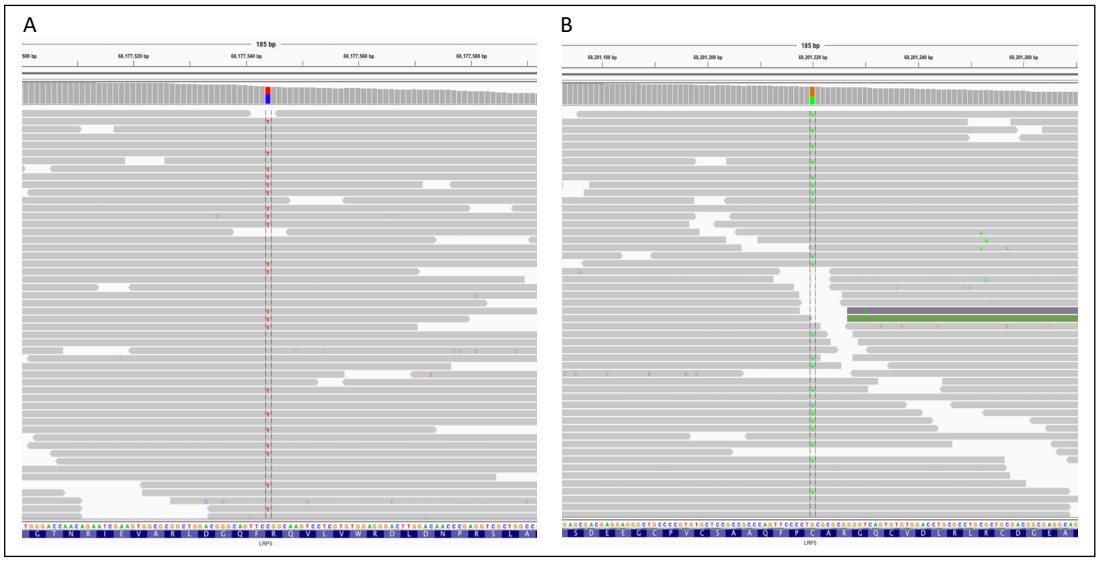


Figure 4.3 IGV snapshot of the compound heterozygous variants in *LRP5* for patient F1250.
(A) IGV screenshot of the *LRP5* c.2254C>T, p.(Arg752Trp) variant. Red colored Ts represent the alternative (mutant) base. Wild type base C is shown in blue in the reference sequence below. **(B)** IGV screenshot of the *LRP5* c.3914G>A, p.(Cys1305Tyr) variant. Green colored As represent the reads with the alternative (mutant) base. Wild type base (G) is shown in orange in the reference sequence below. Grey bars in both images represent reads above the zoomed in region of *LRP5.*

Neither of the two variants were present in the online variant databases, EVS, ExAC and gnomAD (assessed July 2017) (Table 4.3) and both variants were predicted to be pathogenic and damaging by all pathogenicity prediction tools tested (Table 4.4).

EVS	gnomAD	ExAC
0/13,006	0/277,264	0/121,412

Table 4.3 Frequency of c.2254C>T p.(Arg752Trp) variant and c.3914G>A p.(Cys1305Tyr) variant in variant databases.
URLs: EVS database (http://evs.gs.washington.edu/EVS/); gnomAD database (http://gnomad.broadinstitute.org); ExAC database (http://exac.broadinstitute.org).

Variant	CADD	PolyPhen2	SIFT	Blosum62*	PROVEAN	MutPred
c.2254C>T	35	Probably damaging (1.000)	Damaging if ≤ 0.05 (0)	-3	Damaging if ≤ -2.5 (-7.225)	Damaging if ≥ 0.50 (0.891)
c.3914G>A	29.5	Probably damaging (1.000)	Damaging if ≤ 0.05 (0)	-2	Damaging if ≤ -2.5 (-9.174)	Damaging if ≥ 0.50 (0.956)

Table 4.4 Pathogenicity software prediction scores for *LRP5* c.2254C>T p.(Arg752Trp) and c.3914G>A p.(Cys1305Tyr).
URLs: CADD http://cadd.gs.washington.edu (Kircher, 2014); PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/ (Adzhubei et al., 2010); SIFT http://sift.bii.a-star.edu.sg (Ng and Henikoff, 2001); Blosum62 https://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt (Henikoff and Henikoff, 1992); PROVEAN http://provean.jcvi.org/index.php (Choi et al., 2012); MutPred http://mutpred.mutdb.org (Li et al., 2009). *Blosum62 scores range between +3 to -3 and negative scores are more likely to be damaging. CADD scores range from 1 to 99, based on the rank of each variant, with the top 10% of deleterious variants having scores of 10 and above, the top 1% scores of 20 and above and top 0.1% having scores of 30 and above.

The variants were further confirmed by Sanger sequencing (Figure 4.4). No family DNA was available to confirm that the mutations were on different chromosomes.

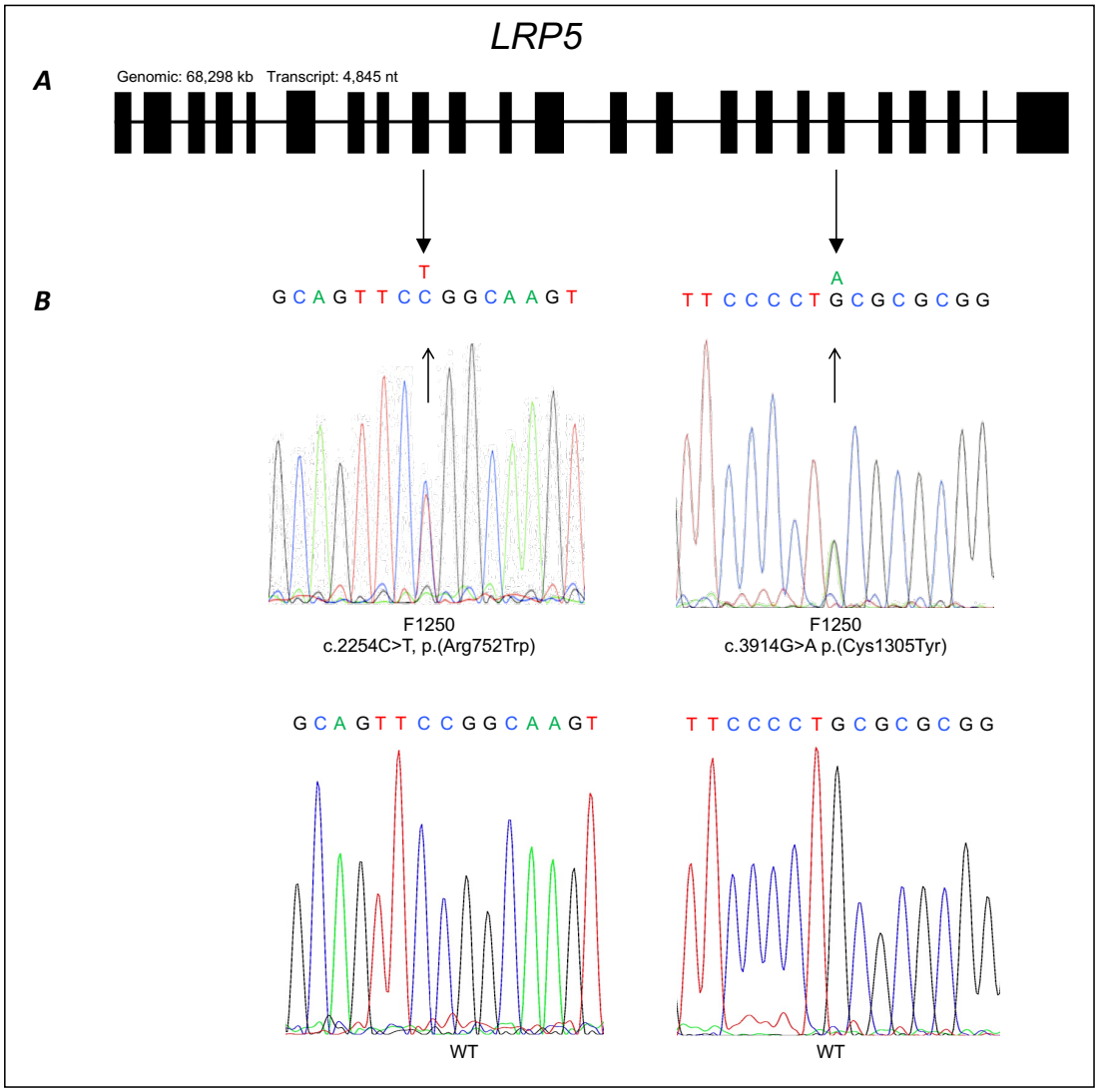


Figure 4.4 *LRP5* c.2254C>T and c.3914G>A variants identified in FEVR patient F1250.
(A) Schematic representation of the *LRP5* transcript. **(B)** Location and sequence traces of the variants on exon 10 and exon 18 with the corresponding wild-type (WT) alleles.

Both variants are fully conserved in all species (Figure 4.5). The variant p.(Arg752Trp) is located in the third YWTD domain and the p.(Cys1305Tyr) is located in the LDL-R-like domain of the LRP5 protein (Figure 4.6).

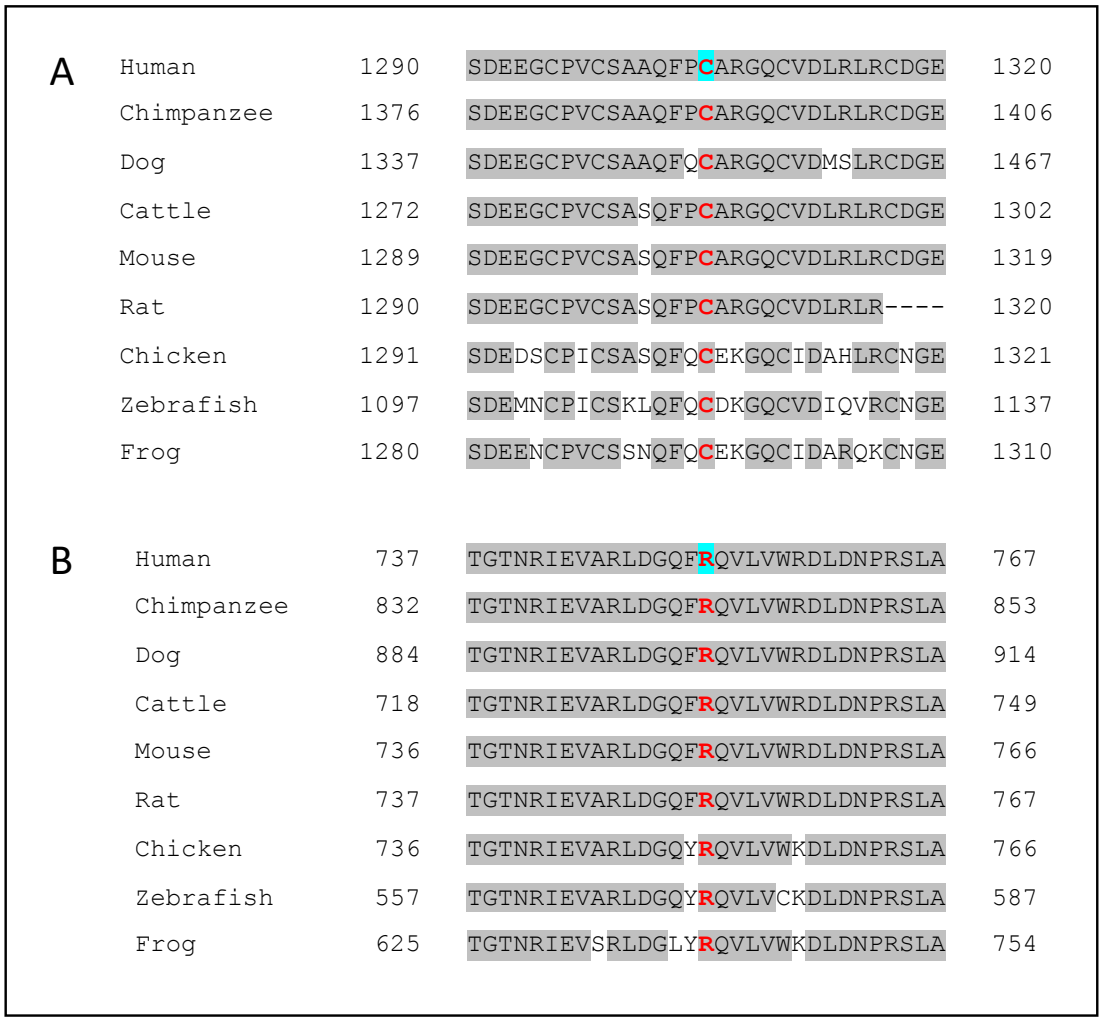


Figure 4.5 Protein sequence alignment of human LRP5 with its orthologues for the p.(Cys1305Tyr) variant (A) and the p.(Arg752Trp) variant (B).
Alignments were calculated using ClustalW**.** Human NP_002326 (Homo sapiens); Chimpanzee XP_508605 (Pan troglodytes); Dog XP_003432463 (Canis lupus familiaris); Cattle XP_002699451 (Bos Taurus); House mouse NP_032539 (Mus musculus); Rat NP_001099791 (Rattus norvegicus); Chicken NP_001012915 (Gallus gallus); Zebrafish NP_001170929 (Danio rerio); Frog XP_002941689 (Xenopus tropicalis). 15 amino acids on either side of the variants are shown. The altered amino acid residues throughout the species are highlighted in red and the affected human amino acid residues are highlighted in red/cyan.

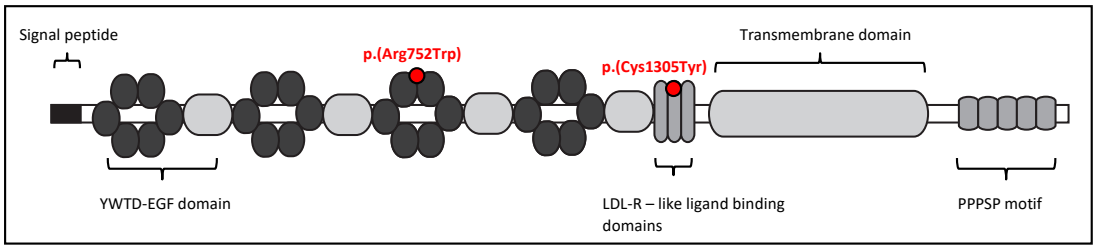


Figure 4.6 Protein structure schematic of LRP5 showing its domains and the p.(Arg752Trp) and p.(Cys1305Tyr) variants.
Individual domains are labelled. Red circles denote the p.(Tyr752Cys) and the p.(Cys1305Tyr) missense variants. The LRP5 structure was adapted from SMART tool (http://smart.embl-heidelberg.de) and He et al. (2004). The YWTD domain – tyrosine (Y), tryptophan (W), threonine (T) and aspartic acid (D) β-propeller domain; EGF domain – epidermal growth factor; LDL-R-like domain – low-density-lipoprotein receptor–like (LDL-R–like) ligand-binding.

4.2.2.2 *LRP5* c.2116 G>A; p.(Gly706Arg) & c.2318+1G>C splice mutation

Compound heterozygous mutations were identified in *LRP5* (NM_002335) in a sporadic UK male diagnosed with FEVR, identifier F1057_1072. The first variant was a missense change located in exon 10, GRCh37, Chr. 11:68177406, c.2116G>A; p.(Gly706Arg). The second variant was a splice site variant at the 3' end of exon 10, GRCh37, Chr. 11:68177609, c.2318+1G>C. Visualisation on IGV confirmed the allele depth of wt/mut: G=73/A=55 for the missense and wt/mut: G=52/C=49 for the splice mutation (Figure 4.7).

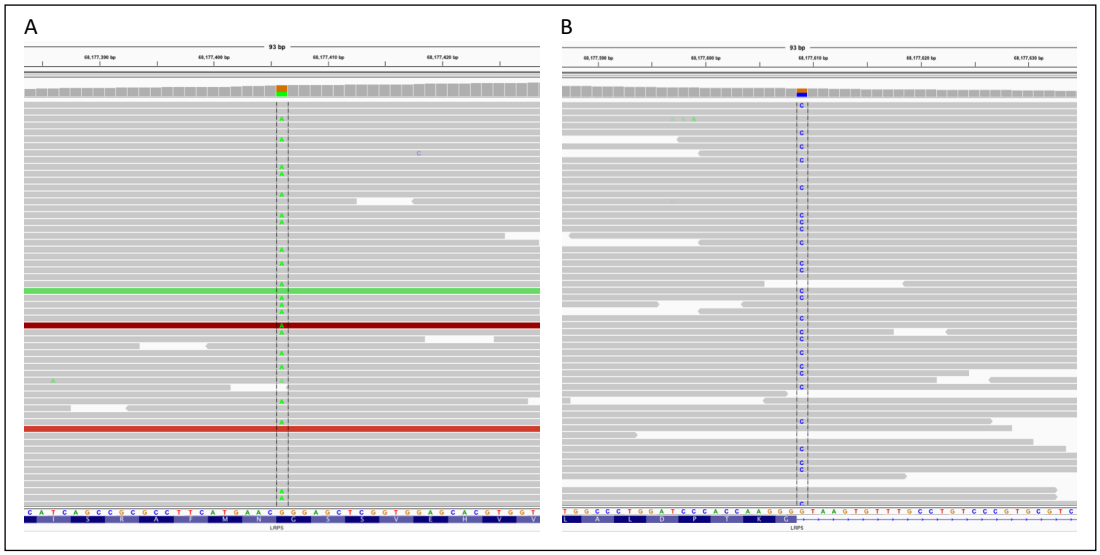


Figure 4.7 IGV snapshot of the compound heterozygous variants in *LRP5* for patient F1057_1072.
(A) IGV screenshot of the *LRP5* c.2116G>A; p.(Gly706Arg) variant. Green colored As represent the alternative (mutant) base. Wild type base (G) is shown in orange in the reference sequence below. **(B)** IGV screenshot of the *LRP5* c.2318+1G>C variant. Blue colored Cs represent the reads with the alternative (mutant) base. Wild type base (G) is shown in orange in the reference sequence below. Grey bars in both images represent reads above the zoomed in region of *LRP5.*

The variant c.2116G>A; p.(Gly706Arg) was found to be present in one allele of a European (non-Finnish) population in gnomAD variant database but was not present in EVS or ExAC (date assessed July 2017). The c.2318+1G>C variant was not reported in any databases (date assessed July 2017) (Table 4.5). The overall frequency of the c.2116G>A; p.(Gly706Arg) variant in the total population is 0.00003234 and the European frequency is 0.00006675 (Table 4.6).

Variant	EVS	gnomAD	ExAC
c.2116G>A	0/13,006	1 heterozygous (1/30,926)	0/121,412
c.2318+1G>C	0/13,006	0/277,264	0/121,412

Table 4.5 Frequency of c.2116G>A p.(Gly706Arg) and c.2318+1G>C variant in variant databases.
EVS (http://evs.gs.washington.edu/EVS/); gnomAD database (http://gnomad.broadinstitute.org); the ExAC database (http://exac.broadinstitute.org).

Population	Allele Count	Allele Number	No. of Homozygotes	Allele Frequency
European (Non-Finnish)	1	14982	0	0.00006675
African	0	8720	0	0.000
Ashkenazi Jewish	0	302	0	0.000
East Asian	0	1622	0	0.000
European (Finnish)	0	3488	0	0.000
Latino	0	838	0	0.000
Other	0	974	0	0.000
South Asian	0	0	0	NA
Total	**1**	**30926**	**0**	**0.00003234**

Table 4.6 Population frequencies generated by gnomAD database for *LRP5* c.2116G>A, p.(Gly706Arg) variant.
The table shows the frequency for this variant by populations. Number of heterozygotes (allele count) and number of homozygotes are shown. The *LRP5* variant is present in 1 allele in European (non-Finnish) population.

The splice variant identified abolishes the invariant GT splice donor site and is therefore likely to be pathogenic. Unfortunately, the precise effect of this variant has not been determined as no RNA was available. However, the most common outcome of splice donor mutations is the deletion of the previous exon and in this case, this would delete exon 10 which would result in a frameshift and a premature stop, p.(Lys697Hisfs*39), presumably resulting in NMD. The c.2116G>A, p.(Gly706Arg) missense variant is predicted to be pathogenic by all six prediction tools (Table 4.7).

CADD	PolyPhen2	SIFT	Blosum62*	PROVEAN	MutPred
32	Probably damaging (1.000)	Damaging if ≤ 0.05 (0.01)	-2	-7.213	Damaging if ≥ 0.50 (0.886)

Table 4.7 Pathogenicity software prediction scores for *LRP5* c.2116G>A p.(Gly706Arg).
URLs: CADD http://cadd.gs.washington.edu (Kircher, 2014); PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/ (Adzhubei et al., 2010); SIFT http://sift.bii.a-star.edu.sg (Ng and Henikoff, 2001); Blosum62 https://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt (Henikoff and Henikoff, 1992); PROVEAN http://provean.jcvi.org/index.php (Choi et al., 2012); MutPred http://mutpred.mutdb.org (Li et al., 2009). *Blosum62 scores range between +3 to -3 and negative scores are more likely to be damaging. CADD scores range from 1 to 99, based on the rank of each variant, with the top 10% of deleterious variants having scores of 10 and above, the top 1% scores of 20 and above and top 0.1% having scores of 30 and above.

Both variants were confirmed in the gDNA of the patient by Sanger sequencing (Figure 4.8) but no family DNA was available to check whether the alleles were on different chromosomes.

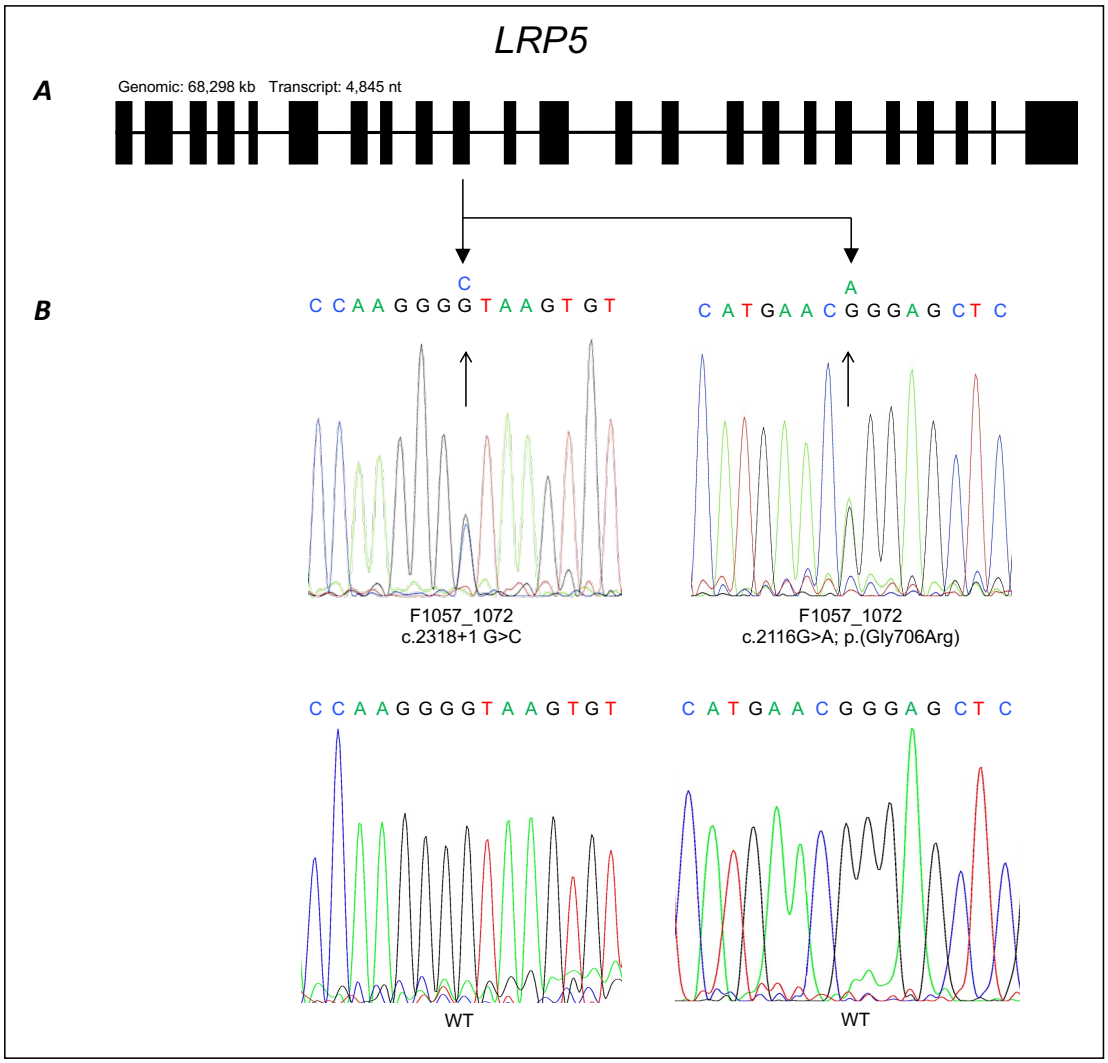


Figure 4.8 *LRP5* c.2116G>A and c.2318+1G>C variants identified in FEVR patient F1057_1072.
(A) Schematic representation of the *LRP5* transcript. **(B)** Location and sequence traces of the variants on exon 10 and exon/intron 10 with the corresponding wild-type (WT) alleles.

The p.Gly706 amino acid residue was found to be fully conserved among all species analysed (Figure 4.9). The p.(Gly706Arg) variant is located in the third YWTD domain (Figure 4.10).

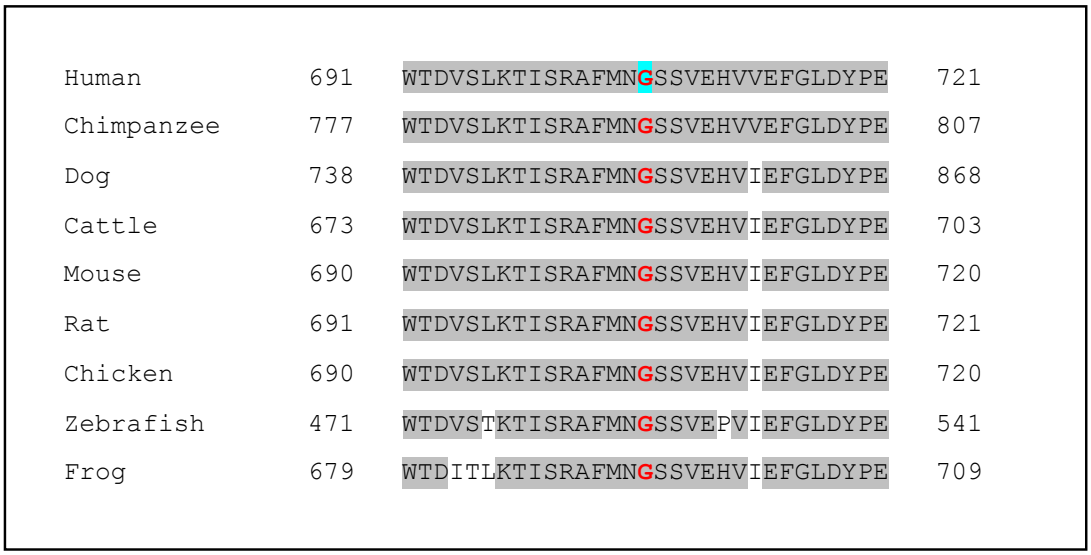


Figure 4.9 Protein sequence alignment of human LRP5 with its orthologues for the p.(Gly706Arg) variant.
Alignments were calculated using ClustalW**.** Human NP_002326 (Homo sapiens); Chimpanzee XP_508605 (Pan troglodytes); Dog XP_003432463 (Canis lupus familiaris); Cattle XP_002699451 (Bos Taurus); House mouse NP_032539 (Mus musculus); Rat NP_001099791 (Rattus norvegicus); Chicken NP_001012915 (Gallus gallus); Zebrafish NP_001170929 (Danio rerio); Frog XP_002941689 (Xenopus tropicalis). 15 amino acids on either side of the variant are shown. The altered amino acid residue throughout the species is highlighted in red and the affected human amino acid residue is highlighted in addition in cyan.

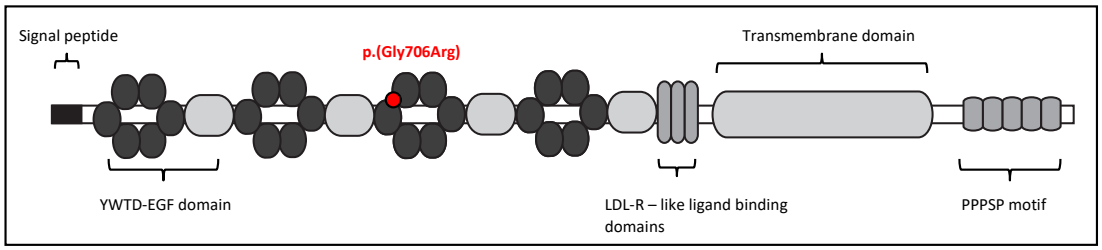


Figure 4.10 Protein structure representation of LRP5 protein, its domains and p.(Gly706Arg).
Individual domains are labelled. Red circle denotes the p.(Gly706Arg) missense variant. The LRP5 structure was adapted from SMART tool (http://smart.embl-heidelberg.de) and (He et al., 2004). The YWTD domain – tyrosine (Y), tryptophan (W), threonine (T) and aspartic acid (D) β-propeller domain; EGF domain – epidermal growth factor; LDL-R-like domain – low-density-lipoprotein receptor–like (LDL-R–like) ligand-binding.

4.2.2.3 *FZD4* c.1513C>T; p.(Gln505*)

A heterozygous nonsense mutation in *FZD4* (NM_012193.3) was identified in a sporadic Belgium female FEVR case, identified F1375, GRCh37, Chr. 11:

86662285, c.1513C>T; p.(Gln505*). The variant was present in exon 2 and had an allele depth of wt/mut: G=39/A=44 (Figure 4.11).

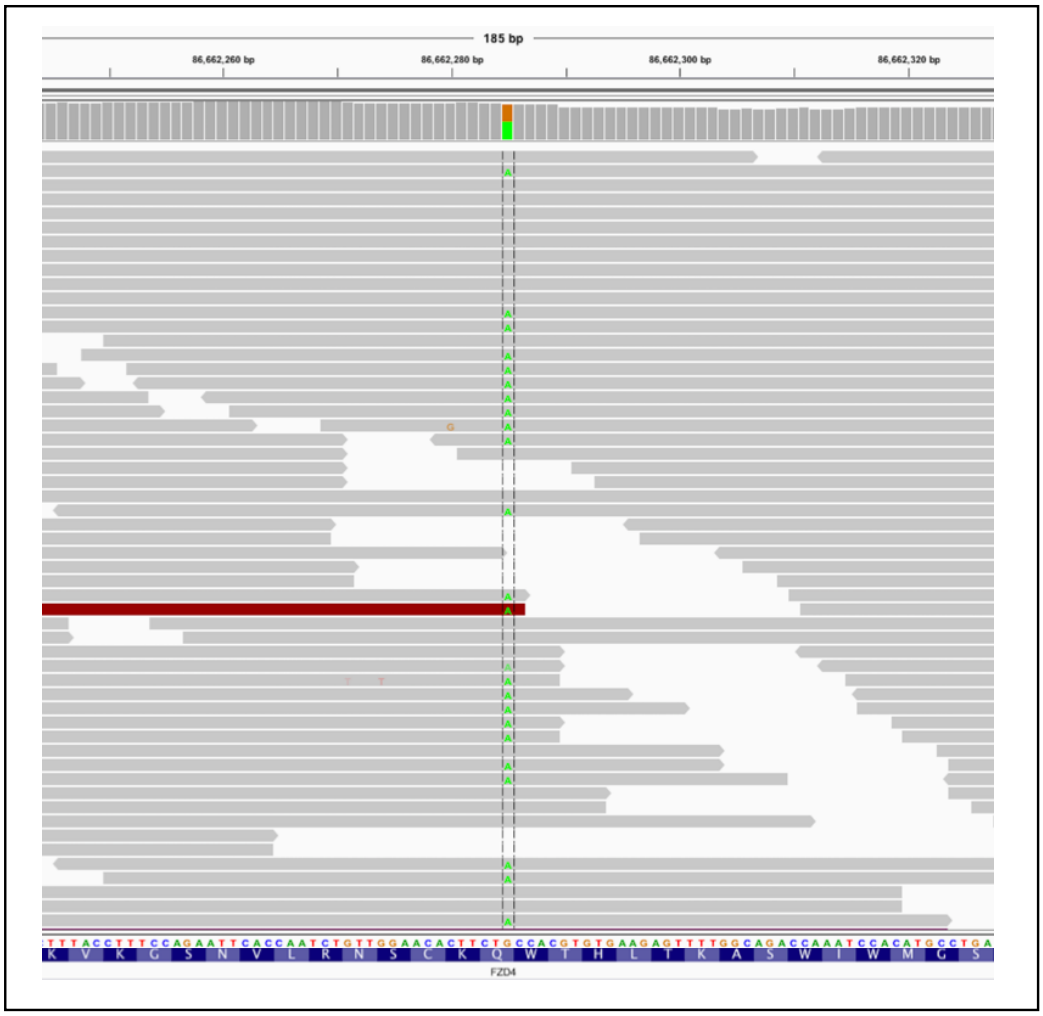


Figure 4.11 IGV snapshot of *FZD4* variant for patient F1375.
IGV screenshot of the *FZD4* c.1513C>T; p.(Gln505*) variant. Green colored As represent the alternative (mutant) base. Wild type base (G) is shown in orange in the reference sequence below. The variant presented in the IGV snapshot is on the opposite strand. Grey bars represent reads above the zoomed in region of *FZD4.*

The c.1513C>T, p.(Gln505*) variant was not present in any of the online variant databases described in this study (Table 4.8). The mutation was confirmed by Sanger sequencing, but no family members were available for segregation (Figure 4.12).

EVS	gnomAD	ExAC
0/13,006	0/277,264	0/121,412

Table 4.8 Frequency of c.1513C>T, p.(Gln505*) variant in variant databases. EVS (http://evs.gs.washington.edu/EVS/): gnomAD database (http://gnomad.broadinstitute.org); ExAC database (http://exac.broadinstitute.org).

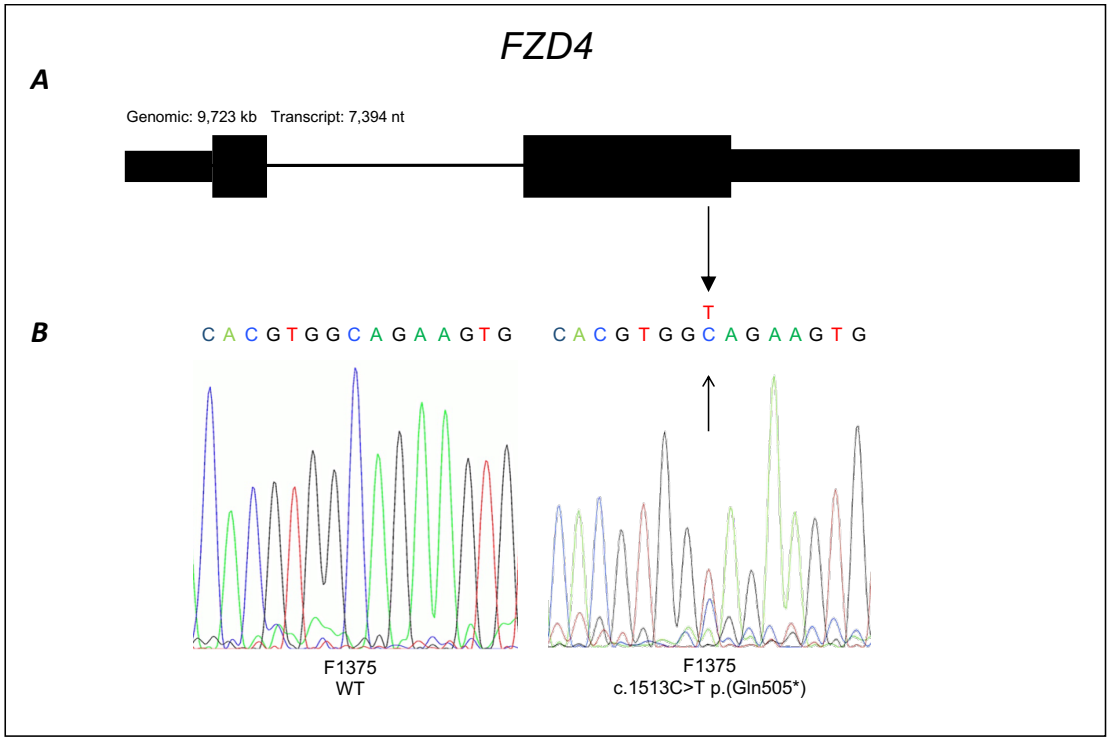


Figure 4.12 *FZD4* c.1513C>T variant identified in FEVR patient F1375.
(A) Schematic representation of the *FZD4* transcript. **(B)** Location and sequence traces on exon 2 with the corresponding wild-type (WT) allele.

The premature stop variant p.(Gln505*) is located in the final exon of *FZD4* so is predicted to avoid NMD and create a truncated protein with an altered DVL association domain and no PDZ domain (Figure 4.13).

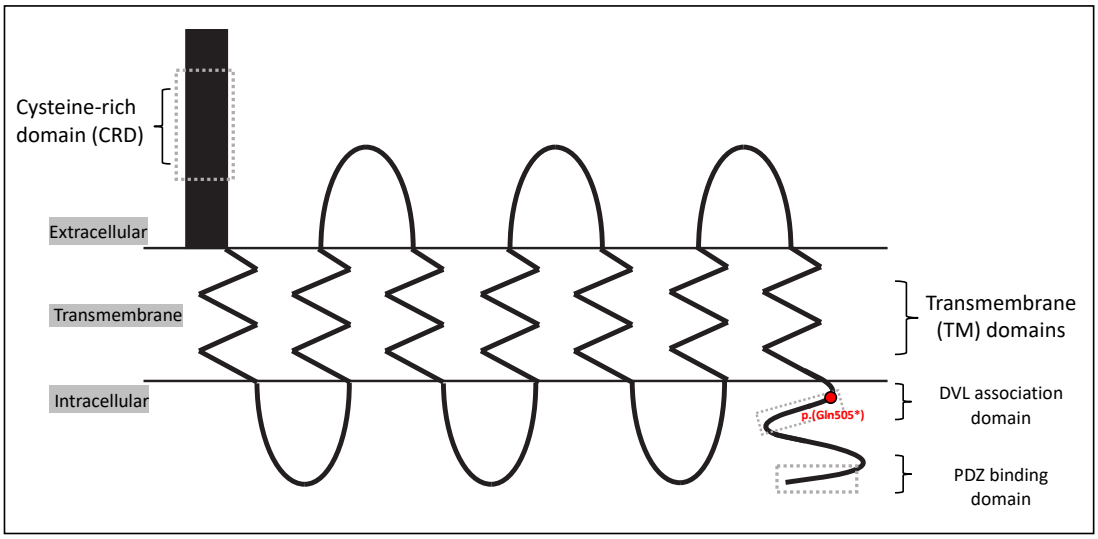


Figure 4.13 Protein structure representation of FZD4 protein, its domains and p.(Gln505*).
Individual domains are labelled. Red circle denotes the p.(Gln505*) variant in the DVL association domain.

4.2.2.4 *KIF11* c.839-840delAG; p.(Arg281Serfs*34)

A heterozygous frameshift variant in *KIF11* (NM_004523) was found in a single sporadic UK male case diagnosed with vitreoretinal dysplasia, identifier F164, GRCh37, Chr. 10:94373182, c.839_840delAG; p.(Arg281Serfs*34). The *KIF11* variant was present in exon 8 with the allele depth of wt/mut: AAG=22/A=28 (Figure 4.14).

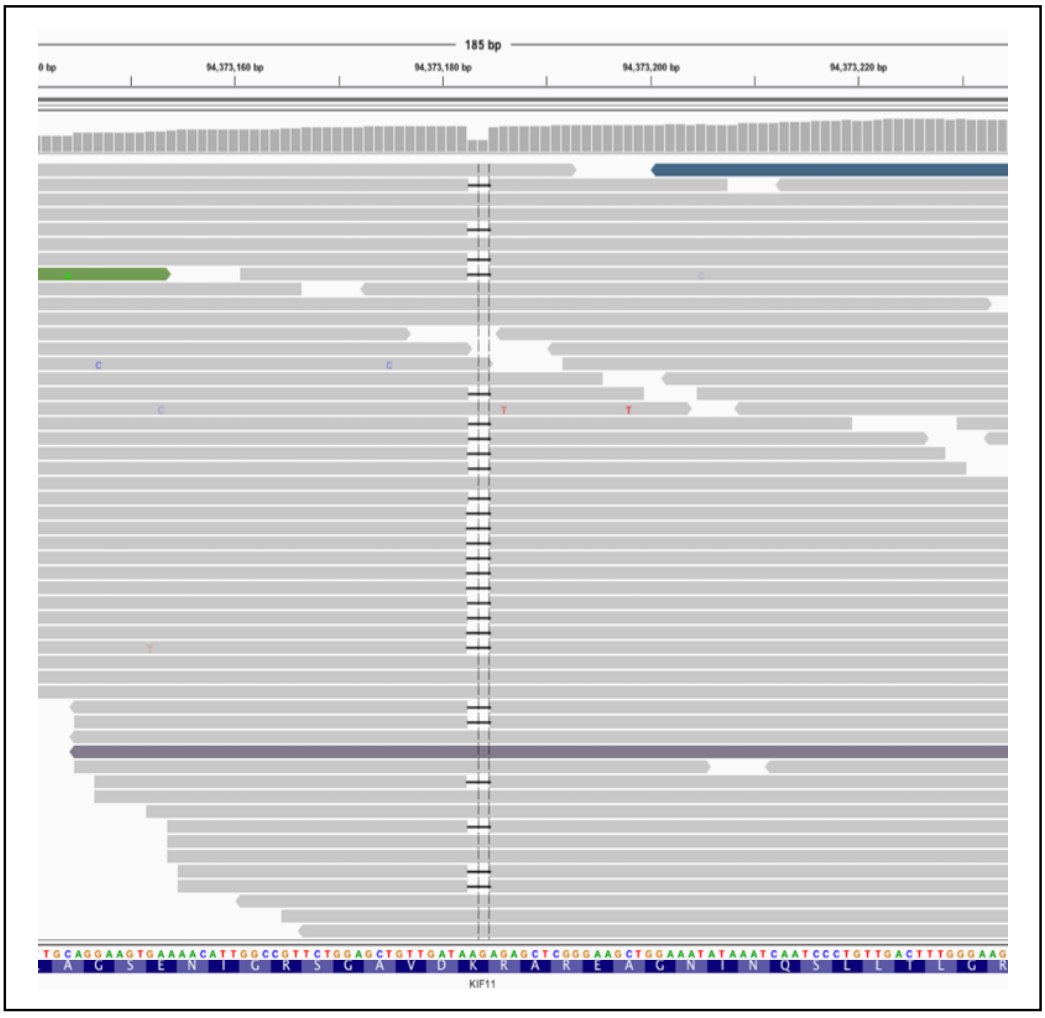


Figure 4.14 IGV snapshot of the *KIF11* variant for patient F164.
IGV screenshot of the *KIF11* c.839-840delAG; p.(Arg281Serfs*34) variant. The black bars across the reads spanning over two bases represent the two-base deletion. Wild type base (AG) is shown in green/orange in the reference sequence below all the reads as only non-wild-type bases are shown. Grey bars represent reads above the zoomed in region of *KIF11.*

The frameshift variant was not present in any online variant databases (Table 4.9) and was confirmed by Sanger sequencing (Figure 4.15). The p.(Arg281Serfs*34) variant disrupts the KISc domain of KIF11 but is predicted to cause NMD (Figure 4.16). No family DNA was available.

EVS	gnomAD	ExAC
0/13,006	0/277,264	0/121,412

Table 4.9 Frequency of c.839-840 p.(Arg281Serfs*34) variant in variant databases.
EVS (http://evs.gs.washington.edu/EVS/); gnomAD database (http://gnomad.broadinstitute.org); ExAC database (http://exac.broadinstitute.org).

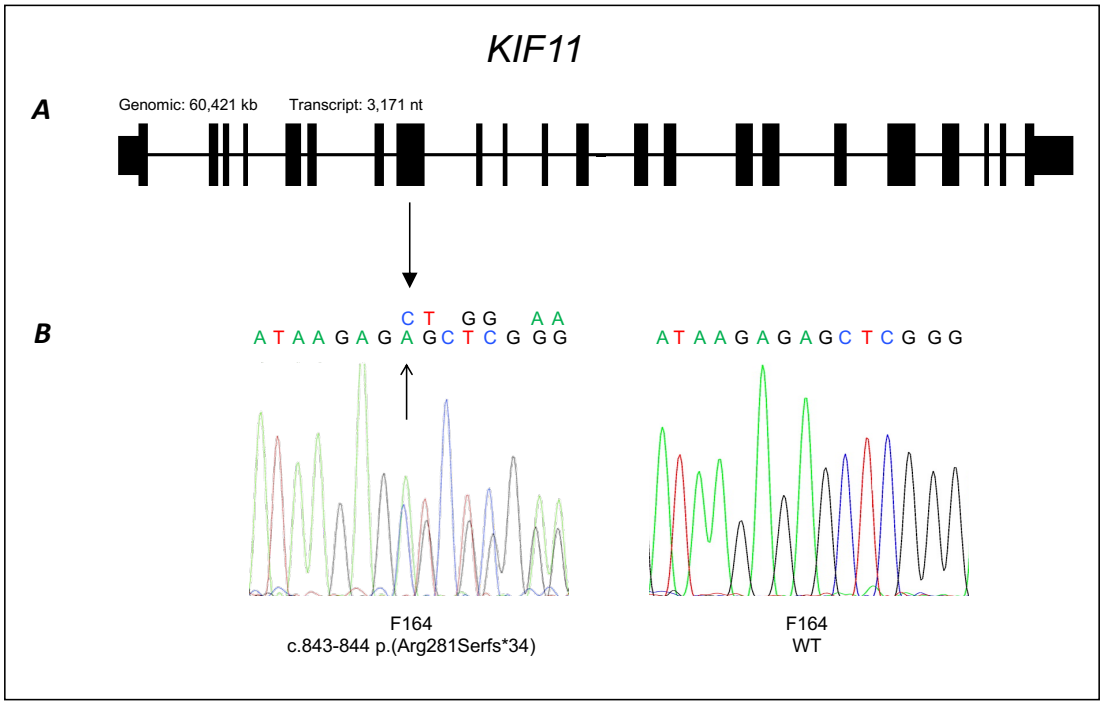


Figure 4.15 *KIF11* c.839-840delAG variant identified in FEVR patient F164.
(A) Schematic representation of the *KIF11* transcript. **(B)** Location and sequence traces on exon 8 with the corresponding wild-type (WT) allele.

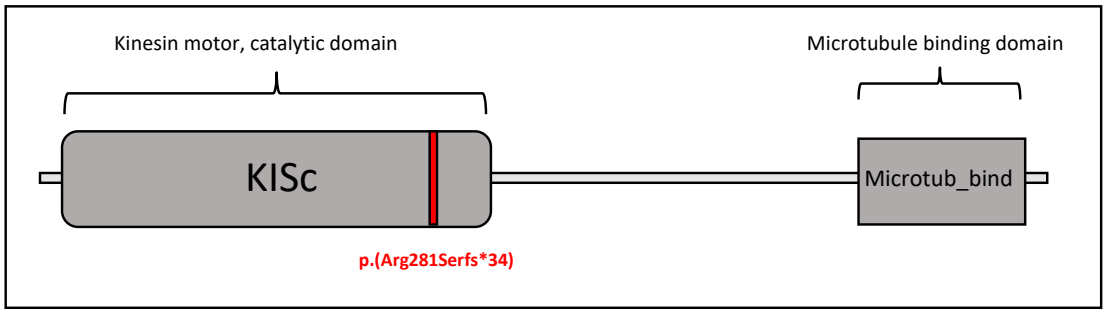


Figure 4.16 Protein structure representation of KIF11 protein, its domains and p.(Arg281Serfs*34).
Individual domains are labelled. Red line denotes the p.(Arg281Serfs*34) variant. Figure adapted from the conserved domain database (CDD) (http://www.ncbi.nlm.nih.gov/Structure/cdd) and Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de) (Schultz et al., 1998; Letunic et al., 2014).

4.2.2.5 *TSPAN12* Heterozygous Deletion of Exons 5, 6 and 7

A heterozygous three-exon deletion in *TSPAN12* (NM_012338) was identified in a sporadic Danish female diagnosed with FEVR, identifier F1308. The large deletion spanning exons 5, 6 and 7 was identified by ExomeDepth (Figure 4.17) (Section 2.20). The breakpoints of the deletion were defined by ExomeDepth as chr7:120446604-120455857 but specific breakpoint were not confirmed. The observed/expected read ratio was 0.497 (82/165), indicating that the deletion was heterozygous. This deletion was not reported in the Database of Genomic Variants (http://dgv.tcag.ca/dgv/app/homeonline) or DECIPHER (http://decipher.sanger.ac.uk) (Firth et al., 2009).

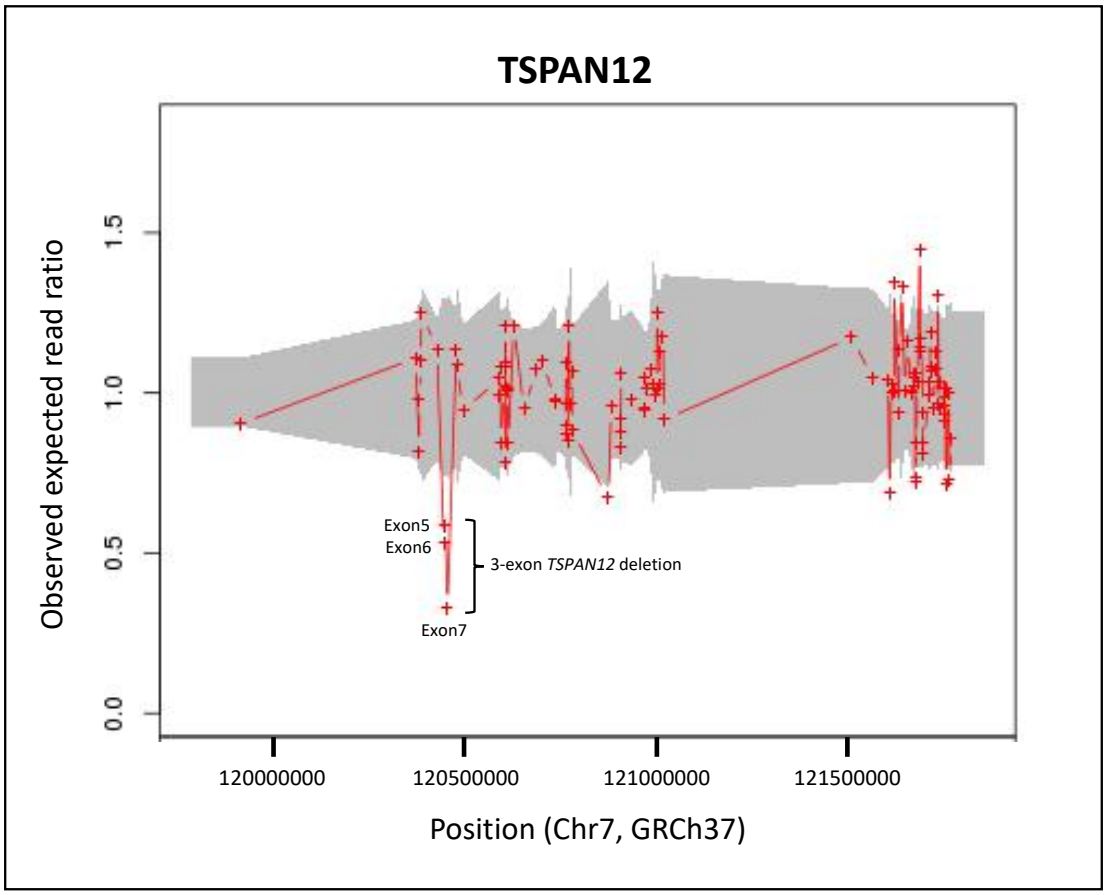


Figure 4.17 Image from ExomeDepth analysis of *TSPAN12* in patient F1308.
The 3-exon deletion is displayed by the lowered expected read ratio to 0.45. Red crosses display the ratio of observed/expected number of reads for each exon and the grey shaded area shows the estimated 99% confidence interval for the particular ratio. The observed/expected number is ~ 1.0 for samples without a CNV and ~ 0.5 for a heterozygous deletion.

TSPAN12 encodes a 305 amino acid receptor containing four transmembrane (TM) domains, a small extracellular loop (SEL) and a large extracellular loop (LEL) (Rubinstein et al., 1996). The removal of exons 5-7 is predicted to produce an in-frame deletion resulting in a 196 amino acids mutant protein missing the third transmembrane domain along with most of the LEL domain (Figure 4.18). This is merely a prediction and RNA from the individual would need to be analysed to fully confirm the outcome of this deletion.

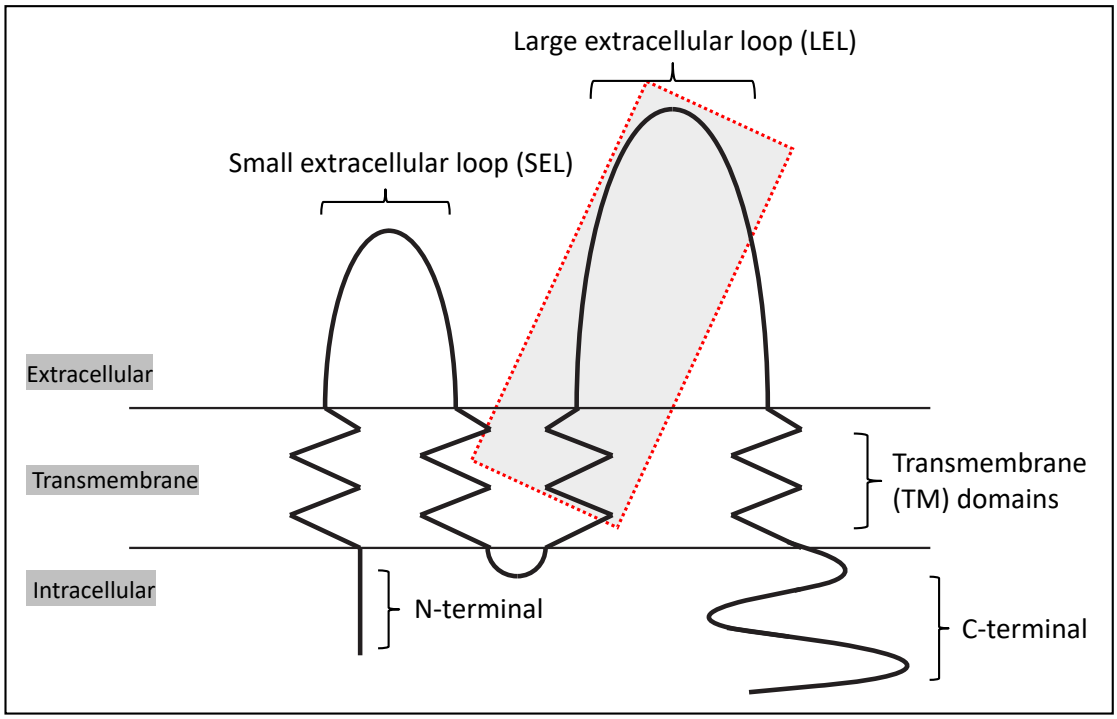


Figure 4.18 Protein structure representation of TSPAN12 protein, its domains and the 3 exon deletion.
Individual domains are labelled. Grey shaded box with dotted red outline denotes the sections of domains that had been deleted with the 3-exon deletion.

4.2.3 Searching for new FEVR genes

No mutations in the known FEVR genes were identified in the remaining 15 exomes. These were therefore reanalysed to try and identify new genes which may underlie this disorder. To aid in this analysis, negative exome data from 10 patients was added to the current dataset. These additional exomes were originally generated and analysed by Panagiotou (2018) and are detailed in Table 4.10.

Patient ID	Ethnicity / Country of Origin	Familial / Sporadic	Phenotype
F170	North European, British	F (affected brother)	FEVR
F1210	North European, British	S	FEVR
F1279	South European, Spanish	F (affected sister)	FEVR, Cleft palate
433	East Asian, Japanese	F (affected mother, uncle, aunt and grandfather)	FEVR
483	North European, German	F (affected son)	FEVR
517	North American, USA	F (affected son)	FEVR
F561	South Asian, Indian	F (affected son and carrier unaffected mother)	PFV, Cataract
1055	Kuwaiti	S	FEVR
1060	South Asian, Malaysian	F (affected brother)	FEVR
1304_1362*	South Asian, Pakistan	F (affected 3 siblings)	FEVR

Table 4.10 Phenotype of additional 10 unsolved FEVR cases.
Ethnicity, origin, phenotype and genes pre-screened are detailed for each individual. FEVR – Familial Exudative Vitreoretinopathy; PFV – Persistent Foetal Vasculature. *1304_1362 was previously found to have a homozygous missense mutation in *NR2E3* identified by Panagiotou (2018).

The raw data for these negative exomes was reanalysed using the current pipeline (Figure 4.2). The depth of coverage for the 10 additional patients is summarised in Table 4.11. Percentages of bases that had above 5, 10, 20 and 30 reads are presented. All samples had on average 73.44% bases covered

above 30 reads. The depth of coverage was not as high as for the original FEVR samples analysed in the study (Table4.2). This is likely due to different exome reagents being used (the samples in Table4.2 used v5 reagent) and the lower amount of data generated on the Illumina Hiseq2500, the platform used to sequence the samples.

The remaining analysis focused on the data from all 25 exomes to try and identify new FEVR genes and mutations.

Sample ID	% of bases above 5	% of bases above 10	% of bases above 20	% of bases above 30
F170	96.4	94.1	88.5	82.7
F1210	96.4	94.3	89.5	84.4
F1279	94.7	90.9	82.8	73.4
433	93.5	88.3	78.2	66.3
483	94.9	91.2	83.6	76.6
517	96.5	94.8	90.9	86.5
F561	96.8	95.4	92.2	88.7
1055	94.7	89.2	74.9	59.2
1060	89.0	79.1	55.7	35.2
F1304_1362	95.4	92.7	86.8	81.4

Table 4.11 Depth of coverage for additional 10 unsolved FEVR WES data.
Sample IDs are provided and the corresponding percentage of bases covered above certain read depths are listed.

4.2.3.1 Identification of a potential mutation in a new *LRP5* transcript

A potential mutation in a new transcript of *LRP5* was identified in a UK girl of Pakistani origin diagnosed with FEVR who has consanguineous parents and affected siblings, identifier F1304_1362. The primary transcript (transcript 1) (NM_002335) of *LRP5* contains 23 exons and produces a 1,615 amino acid protein (Kim et al., 1998). The new transcript, annotated as transcript 2 (NM_001291902.1), is missing the original exon 7 and contains a translation start site in a new exon, denoted as Ex8a, to produce a much shorter protein of 1,034 amino acids (Figure 4.19 & Figure 4.20). The shorter protein lacks the signal peptide, the first two YWTD domains and the first EGF domain (Figure 4.23). The identified variant is a heterozygous missense change which is located in the novel Ex8a, GRCh37, Chr.11:68183886, c.16A>G; p.(Thr6Ala).

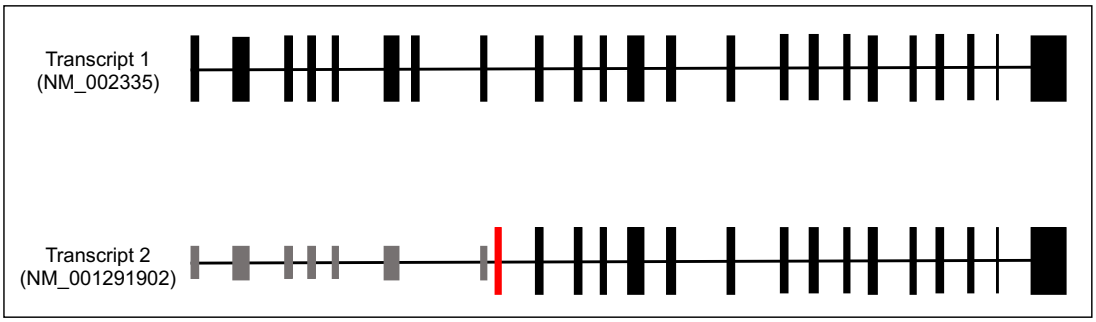


Figure 4.19 The two transcripts of *LRP5*.
There are two transcripts of *LRP5* on RefSeq UCSC Genome Browser (http://genome.ucsc.edu/index.html). Transcript 1 encodes the canonical 1,615 amino acid protein whereas the new transcript 2 encodes a 1,034 amino acid protein by missing exon 7 and introducing a novel exon (Ex8a) with an alternative start site (shaded in red).

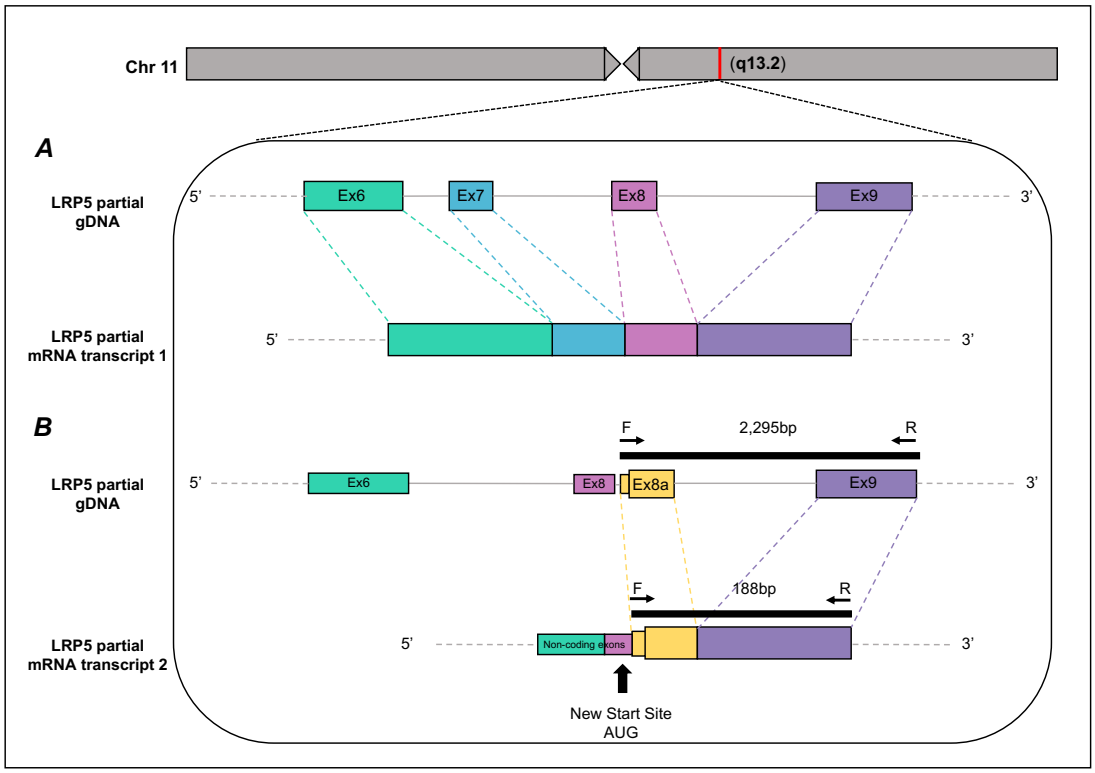


Figure 4.20 Schematic of the splicing of mRNA to generate the two transcripts for *LRP5* gene.
(A) Partial *LRP5* gDNA and corresponding transcript 1 (NM_002335). All exons displayed are translated as indicated in the mRNA. **(B)** Partial gDNA and corresponding partial transcript 2 (NM_001291902.1). The new AUG start site is in the Ex8a (yellow). The corresponding mRNA is displayed below the gDNA. Introns are represented by the continuous grey line connecting the exons. Arrows display the forward (F) and reverse (R) primer with the continuous black line representing the PCR product within the gDNA of 2,295bp and within the cDNA of 188bp. Chromosomal location is shown at the very top of the image with a schematic representation of the chr11 with the p and q arms. Red line across the chromosome symbolises the current position in the *LRP5* gene.

The variant has an allele depth of A=105/G=66 (wt/mut) on IGV (Figure 4.21) and the predicted missense change generated mixed pathogenic prediction scores, with three tools scoring the variant as pathogenic and three as benign (Table 4.12).

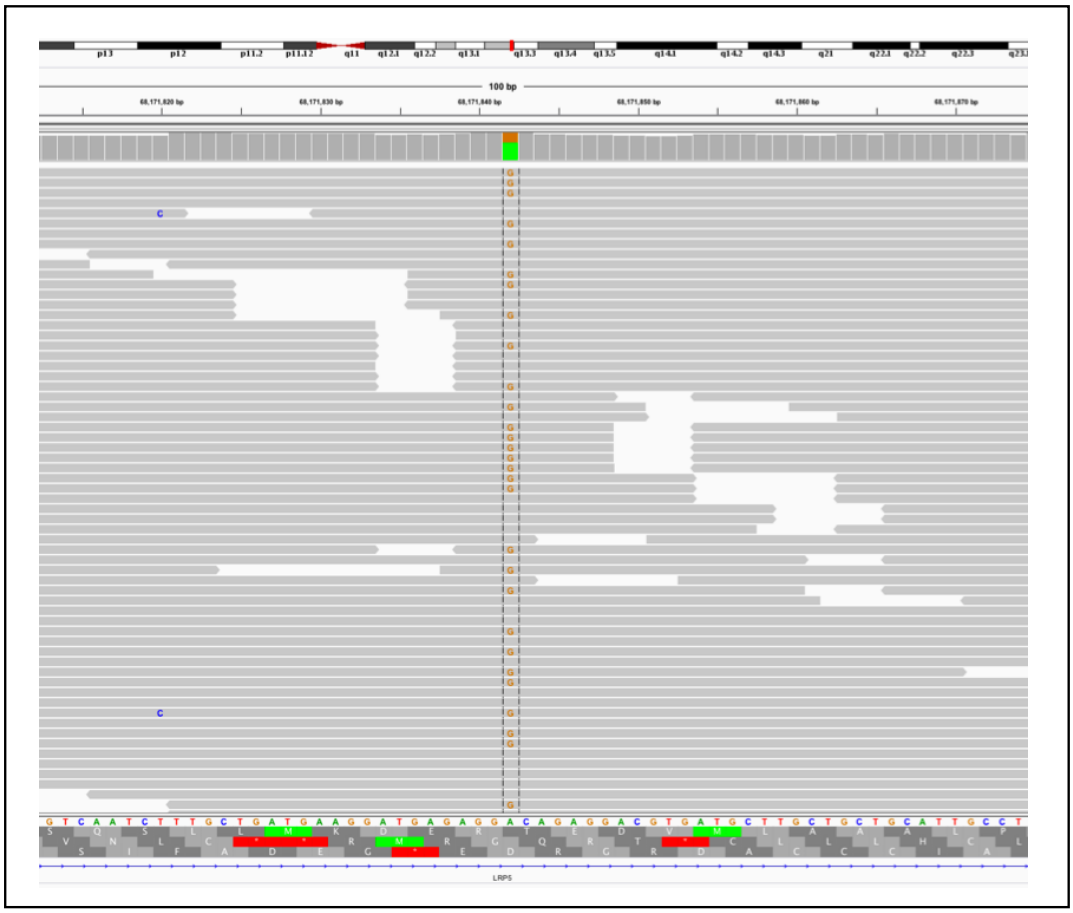


Figure 4.21 IGV snapshot of the *LRP5* variant in Ex8a of transcript 2 (NM_001291902.1) for patient F1304_1362.
IGV screenshot of the *LRP5* c.16A>G p.(Thr6Ala) variant. Orange coloured Gs represent the alternative (mutant) base. Wild type base (A) is shown green in the reference sequence. Grey bars represent reads above the zoomed in region of *LRP5.*

CADD	PolyPhen2	SIFT	Blosum62*	PROVEAN	MutPred
27.1	Benign (0.010)	Damaging if ≤ 0.05 (0)	0	Neutral (0.067)	Damaging if ≥ 0.50 (0.37)

Table 4.12 Pathogenicity software prediction scores for *LRP5* c.16A>G p.(Thr6Ala).
URLs: CADD http://cadd.gs.washington.edu (Kircher, 2014); PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/ (Adzhubei et al., 2010); SIFT http://sift.bii.a-star.edu.sg (Ng and Henikoff, 2001); Blosum62 https://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt (Henikoff and Henikoff, 1992); PROVEAN http://provean.jcvi.org/index.php (Choi et al., 2012); MutPred http://mutpred.mutdb.org (Li et al., 2009). *Blosum62 scores range between +3 to -3 and negative scores are more likely to be damaging. CADD scores range from 1 to 99, based on the rank of each variant, with the top 10% of deleterious variants having scores of 10 and above, the top 1% scores of 20 and above and top 0.1% having scores of 30 and above.

The variant was confirmed in the gDNA of the patient by Sanger sequencing analysis (Figure 4.22). Unfortunately, there was no family DNA currently available to segregate this variant. The changed amino acid is at the very start of the new protein but not in any specific domain (Figure 4.23). No other variants in *LRP5* Ex8a were identified in FEVR or Coats disease patients screened in this thesis.

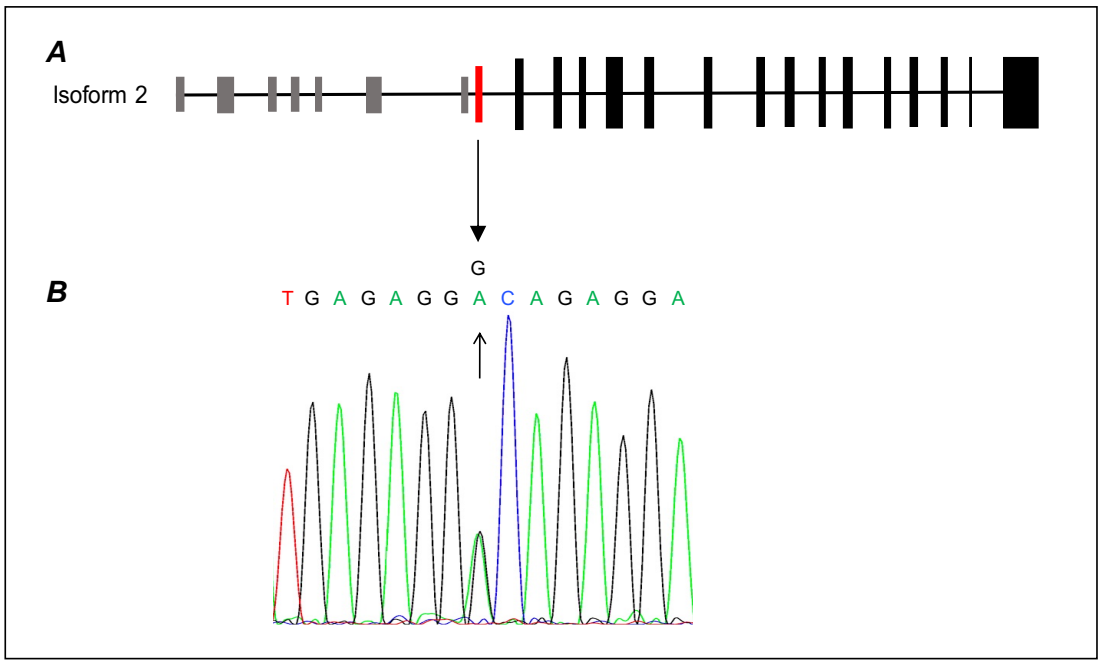


Figure 4.22 *LRP5* variant c.16A>G identified in an FEVR patient F1304_1362 in the Ex8a of transcript 2.
(A) Schematic representation of the *LRP5* transcript. Grey coloured blocks are noncoding exons, red coloured block is the new Ex8a and black coloured blocks are all coding exons.
(B) Location and sequence traces on Ex8a.

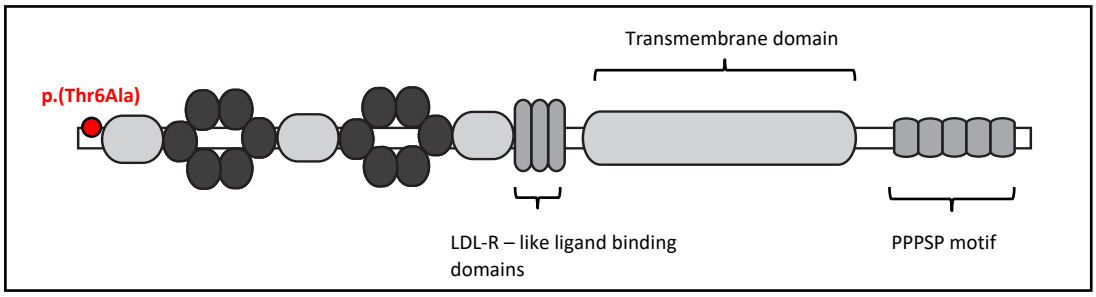


Figure 4.23 Schematic of LRP5 protein structure translated from isoform 2, its domains and p.(Thr6Ala).
Individual domains are labelled. The signal peptide, first two YWTD domains and first EGF domain present on isoform 1 are missing from this new isoform. The LRP5 structure was adapted from SMART tool (http://smart.embl-heidelberg.de) and (He et al., 2004). The YWTD domain – tyrosine, tryptophan, threonine and aspartic acid β-propeller domain; EGF domain – epidermal growth factor; LDL-R-like domain – low-density-lipoprotein receptor–like (LDL-R–like) ligand-binding.

To verify whether the transcript was real, and to compare the mRNA expression of the two documented transcripts, a panel of cDNAs derived from a range of tissues was screened by reverse transcriptase (RT)-PCR (Section 2.22). The cDNA panel included bone marrow, whole brain, fetal liver, heart, kidney, liver, lung, placenta, prostate, skeletal muscle, spleen, testis, thymus, trachea, uterus, colon w/ mucosa, small intestine, spinal cord, stomach and retina. The panel also included cDNA from cell lines RPE1, ARPE19 and HUVECs generated and verified by Soriano (2017). To check whether cDNA synthesis of all tissues was successful, and to control for the amount of cDNA generated, a PCR reaction was performed using *P53* primers (sequences provided in Appendix 7.13) (Figure 4.24C). All tissues were converted to cDNA successfully. gDNA was run alongside as a positive control for the PCR reaction (Figure 4.24).

To verify the presence of transcript 2, primers unique for this transcript were designed within the new exon, Ex8a, and exon 9 (Figure 4.24B). The expected product was 188bp. The primers did not overlap the *LRP5L* pseudogene. For

comparison, primers specific to *LRP5* transcript 1 (located in exon 1 and exon 3) were also used to amplify the cDNA panel (Figure 4.24A). The sequences of the primers are provided in Appendix 7.14. The results showed that the new *LRP5* transcript 2 is a real transcript and is widely expressed as *LRP5* transcript 1.

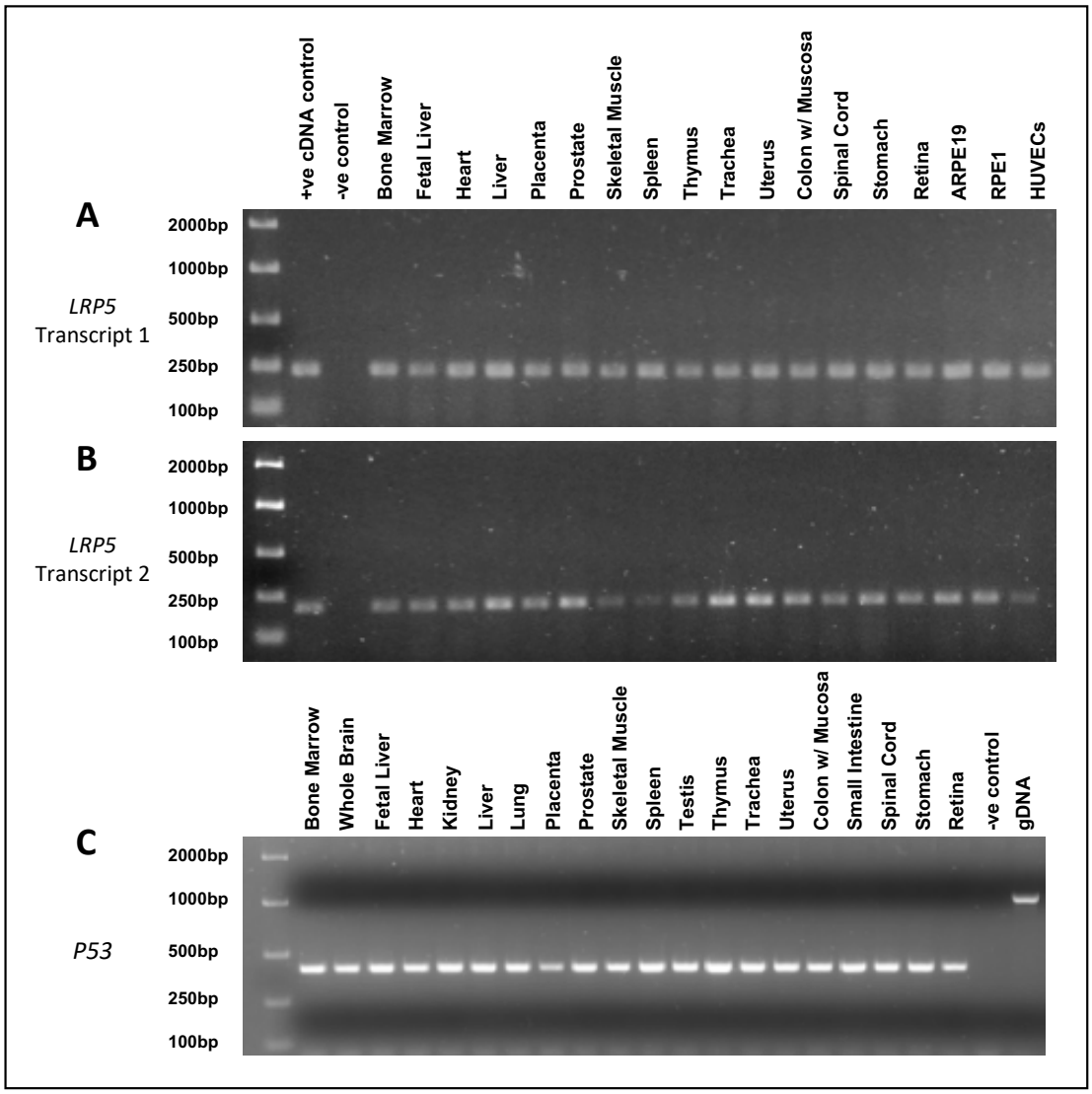


Figure 4.24 Expression of *LRP5* transcript 1 and transcript 2 in a cDNA tissue panel synthesis from total RNA.
(A) cDNA expression of *LRP5* transcript 1. Expression of transcript 1 in tissues and cell lines. Primers were designed without overlapping the pseudogene *LRP5L* to generate a product of 221bp in cDNA. Expression is shown in all tissues. **(B)** cDNA expression of *LRP5* transcript 2**.** Primers were designed to be specific to *LRP5* transcript 2 (by locating one in the new exon, Ex8a) to generate a product of 188bp in cDNA. **(C)** cDNA expression of *P53* to confirm the quality and quantity of cDNA. Product size for cDNA was 410bp and product size expected for gDNA was 1,059bp. For all PCRs, negative control was water, positive cDNA control was stomach and size standard was EasyLadder I (Bioline).

4.2.3.2 Identification of a potential mutation in *DLG1*

A heterozygous potential mutation in a gene called disc large MAGUK scaffold protein 1 (*DLG1*) was identified in an Indian female patient diagnosed with FEVR, identifier F561, GRCh37, Chr. 3:196910495, c.185G>A; p.(Arg62Gln) (NM_001204387 & NM_001204388). This variant was detected in the original analysis performed by Panagiotou (2018),however, the potential significance of this finding didn't become clear until the current study. Professor Jeremy Nathans (Johns Hopkins University) had identified *DLG1* as a candidate FEVR gene through his experiments on retinal angiogenesis mouse models (unpublished confidential data) and he was looking for human mutations. This prompted a search of the FEVR and Coats dataset.

DLG1 has 6 transcripts and the identified variant is only present in transcripts 4 (NM_001204387.1) and 5 (NM_001204388.1) (Figure 4.25). The *DLG1* variant had an allele depth of C=62/T=70 (wt/mut) when visualised on IGV (Figure 4.26).

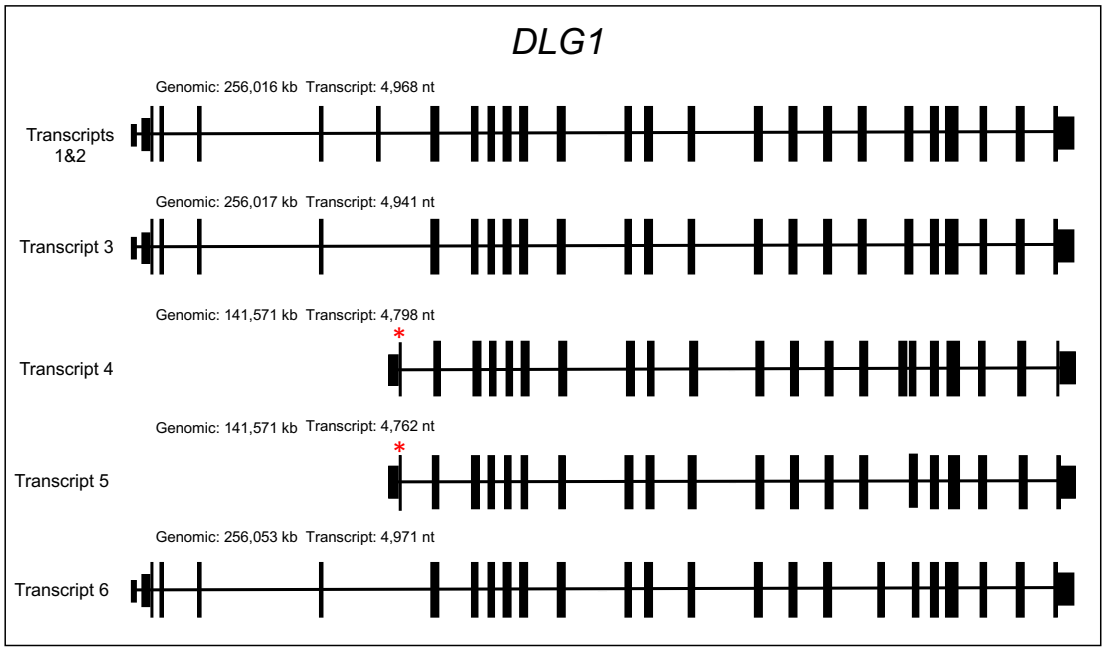


Figure 4.25 Schematic representation of the *DLG1* transcripts.
Transcripts presented as annotated by RefSeq GRCh37/hg19 (https://genome.ucsc.edu). Red star denotes the exon in transcript 4 and 5 where the c.185G>A variant was identified.

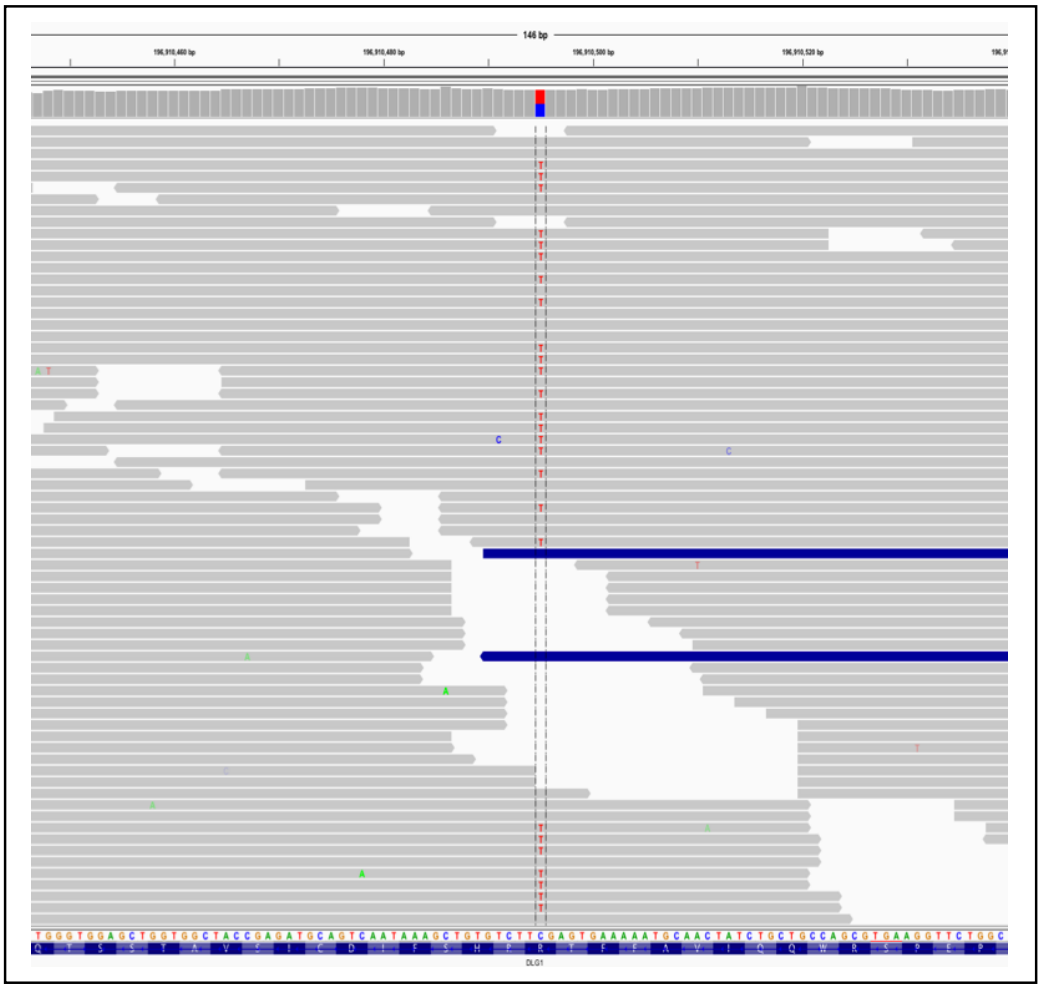


Figure 4.26 IGV snapshot of the *DLG1* variant for patient F561.
IGV screenshot of the *DLG1* c.185G>A; p.(Arg62Gln) variant. The red colored Ts represent the reads with alternative (mutant) base. Wild type base (C) is shown in blue in the reference sequence below. Grey bars represent reads above the zoomed in region of *DLG1.*

The c.185G>A variant was present in two alleles in the gnomAD database but was not in EVS or ExAC (Table 4.13). Both gnomAD alleles were from the Latino population giving an allele frequency of 0.00008386. The patient is Indian but this variant was not reported in 22,594 alleles from the South Asian population on gnomAD. The overall general population frequency of the variant is 0.00001391 (Table 4.14).

EVS	gnomAD	ExAC
Not reported	2 heterozygous (2/143,758)	0/121,412

Table 4.13 Frequency of c.185G>A p.(Arg62Gln) variant in variant databases. EVS (http://evs.gs.washington.edu/EVS/); gnomAD database (http://gnomad.broadinstitute.org); the ExAC database (http://exac.broadinstitute.org).

Population	Allele Count	Allele Number	No. of Homozygotes	Allele Frequency
Latino	2	23850	0	0.00008386
African	0	6520	0	0.000
Ashkenazi Jewish	0	8106	0	0.000
East Asian	0	10122	0	0.000
European (Finnish)	0	16652	0	0.000
European (Non-Finnish)	0	52142	0	0.000
Other	0	3772	0	0.000
South Asian	0	22594	0	0.000
Total	**2**	**143758**	**0**	**0.00001391**

Table 4.14 Population frequencies generated by gnomAD database for *DLG1* c.185G>A, p.(Arg62Gln) variant.
The table shows the frequency for this variant by population. Number of heterozygotes (allele count) and number of homozygotes are shown. The *DLG1* variant is present in 2 alleles in the Latino population.

The deleteriousness of the *DLG1* variant was scored to be pathogenic by two pathogenicity prediction tools but three predicted it to be benign and it only had a low CADD score of 16.42 (Table 4.15).

CADD	PolyPhen2	SIFT	Blosum62*	PROVEAN	MutPred
16.42	Probably damaging (0.997)	Tolerated (0.59)	1	Neutral (0.729)	Damaging if ≥ 0.50 (0.549)

Table 4.15 Pathogenicity software prediction scores for *DLG1* variant c.185G>A; p.(Arg62Gln).
URLs: CADD http://cadd.gs.washington.edu (Kircher, 2014); PolyPhen-2 http://genetics.bwh.harvard.edu/pph2/ (Adzhubei et al., 2010); SIFT http://sift.bii.a-star.edu.sg (Ng and Henikoff, 2001); Blosum62 https://www.ncbi.nlm.nih.gov/Class/FieldGuide/BLOSUM62.txt (Henikoff and Henikoff, 1992); PROVEAN http://provean.jcvi.org/index.php (Choi et al., 2012); MutPred http://mutpred.mutdb.org (Li et al., 2009). *Blosum62 scores range between +3 to -3 and negative scores are more likely to be damaging. CADD scores range from 1 to 99, based on the rank of each variant, with the top 10% of deleterious variants having a score of 10 and above, the top 1% scores of 20 and above and top 0.1% having scores of 30 and above.

The variant was verified in patient F561 and segregated in available family members using Sanger sequencing (Figure 4.28). Five additional family members were screened, and two were found to carry the variant in a heterozygous state F566 (I:2), the patient's mother, and F562 (III:1), the patient's son (Figure 4.27). The son was diagnosed with FEVR, but the mother was reported to be unaffected although no detailed clinical analysis had been undertaken and she may represent a case of non-penetrance or be asymptomatic.

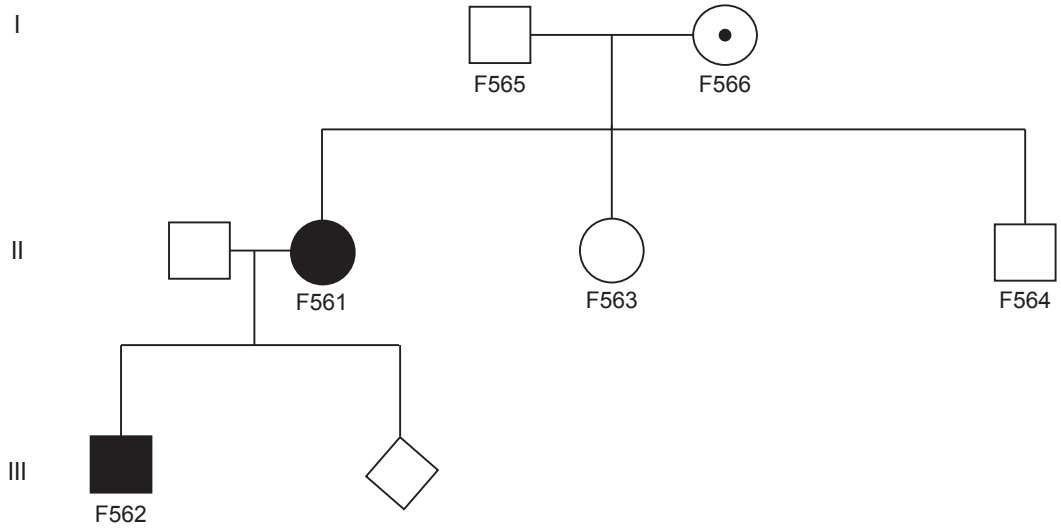


Figure 4.27 Family tree for patient F561.
DNA was available from family members that have ID numbers. Upon the screening of all patients by Sanger sequencing, patient F566 and F562 were also found to carry the same heterozygous variant as patient F561. Squares denote males, circles females and rotated square unknown gender. Shaded symbols denote affected patients. The symbol with a dot indicates that the family member carried the variant but was not diagnosed with FEVR.

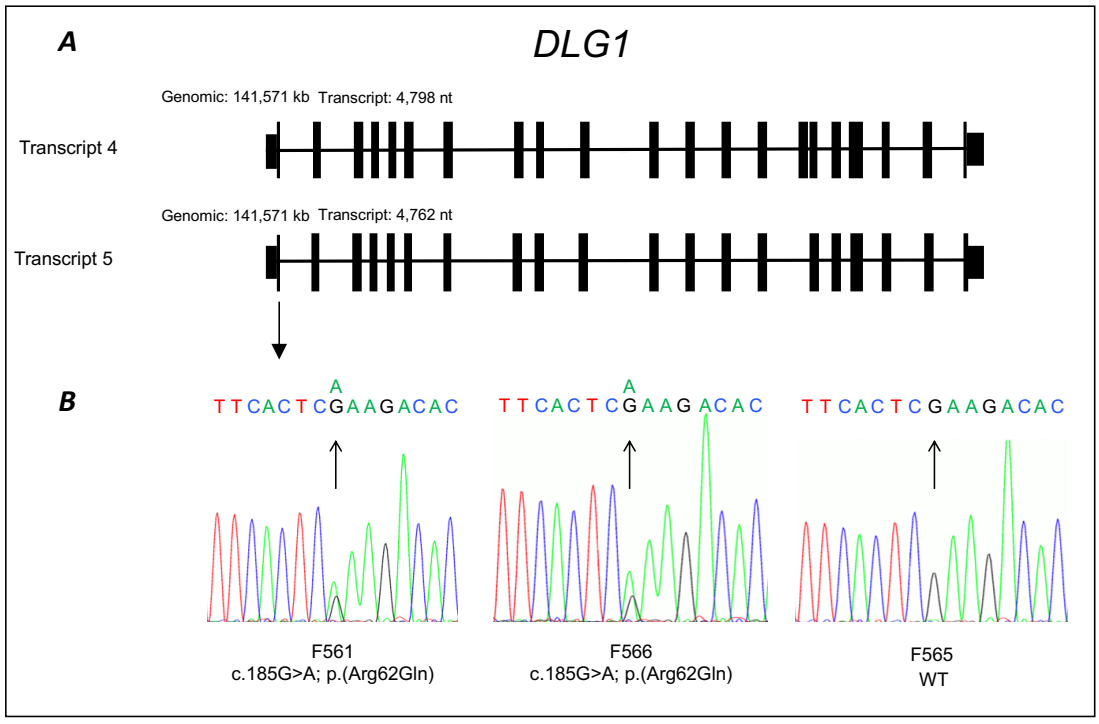


Figure 4.28 *DLG1* c.185G>A variant identified in FEVR patient F561.
(A) Schematic representation of the *DLG1* transcript 4 and 5. (**B)** Location and sequence traces of the variant on exon 1 with the corresponding wild-type (WT) allele.

The protein schematic and the location of the p.(Arg62Gln) variant is shown in Figure 4.29. The variant is present just outside of the PDZ domain and the peptide produced by transcripts 4 and 5 lacks the LS27 domain.

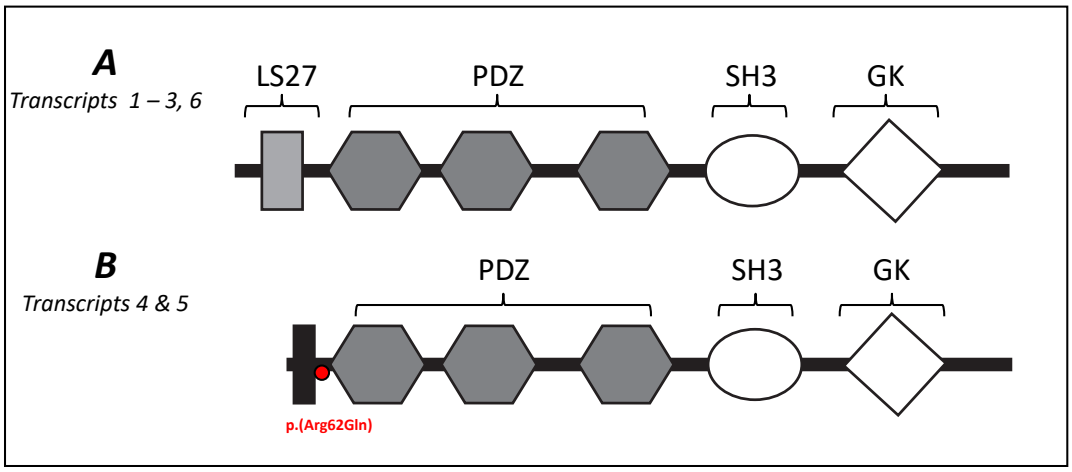


Figure 4.29 Protein structure representation of DLG1 protein, its different transcripts and p.(Arg62Gln).
Individual domains are labelled. Red circle denotes the p.(Arg62Gln) missense variant found in patient F561. Figure was adapted from Uezato et al. (2015).

The expression of *DLG1* transcripts 1 (NM_001098424.1), 2 (NM_004087.2), 3 (NM_001204386.1) and 6 (NM_001363865.1) was investigated in a range of tissues, including tissues of the retina and cDNA from other retinal and vascular cells (RPE1, ARPE19 and HUVECs). The same cDNA panel was generated as described in Section 4.2.3.1 and the cDNA was evaluated by the use of *P53* primers (Figure 4.30B).

Only transcripts 1,2, 3 and 6 were tested as primers for transcripts 4 and 5 were not able to be optimised. Initially, stomach cDNA was used for optimisation, but the primers failed to amplify a product. However, the failed optimisation with stomach cDNA could indicate that the expression of *DLG1* transcript 4 and 5 was simply not expressed in this tissue cDNA, not that the primers were not optimised. For this reason, another set of primers was tested

for transcripts 4 and 5 with stomach cDNA. However, optimisation with further set of primers was still not successful. In past studies, *DLG1* expression was documented to be present in digestive tract and brain tissues (Mori et al., 1998). The expression of transcripts 4 and 5 was also consulted in expressed sequence tags (EST) and mRNA databases (UCSC Genome Browser). For humans, there were two mRNA clones present, AK294855 and AK294772. Both of these mRNAs had source tissue type of brain. Therefore, brain cDNA was used to optimise the primer sets but, again, this was not successful. Ideally, primers used would also be able to amplify a product in gDNA, however, in this case this was not possible as the gDNA product would have been over 33,000bp.

For these reasons, tissue expression of transcripts 1,2,3 and 6 was determined instead and primers were designed and successfully optimised. All transcripts appear to be widely expressed (Figure 4.30). Transcripts 1,2 and 6 can't be differentiated as they produce a product of the same size, but they all show to have decreased expression in whole brain and spinal cord in comparison to transcript 3. Importantly, all of these transcripts were found to have expression in the retina, in RPE-derived cell lines, RPE1 and ARPE1 and in vascular cells HUVECs (Figure 4.30A).

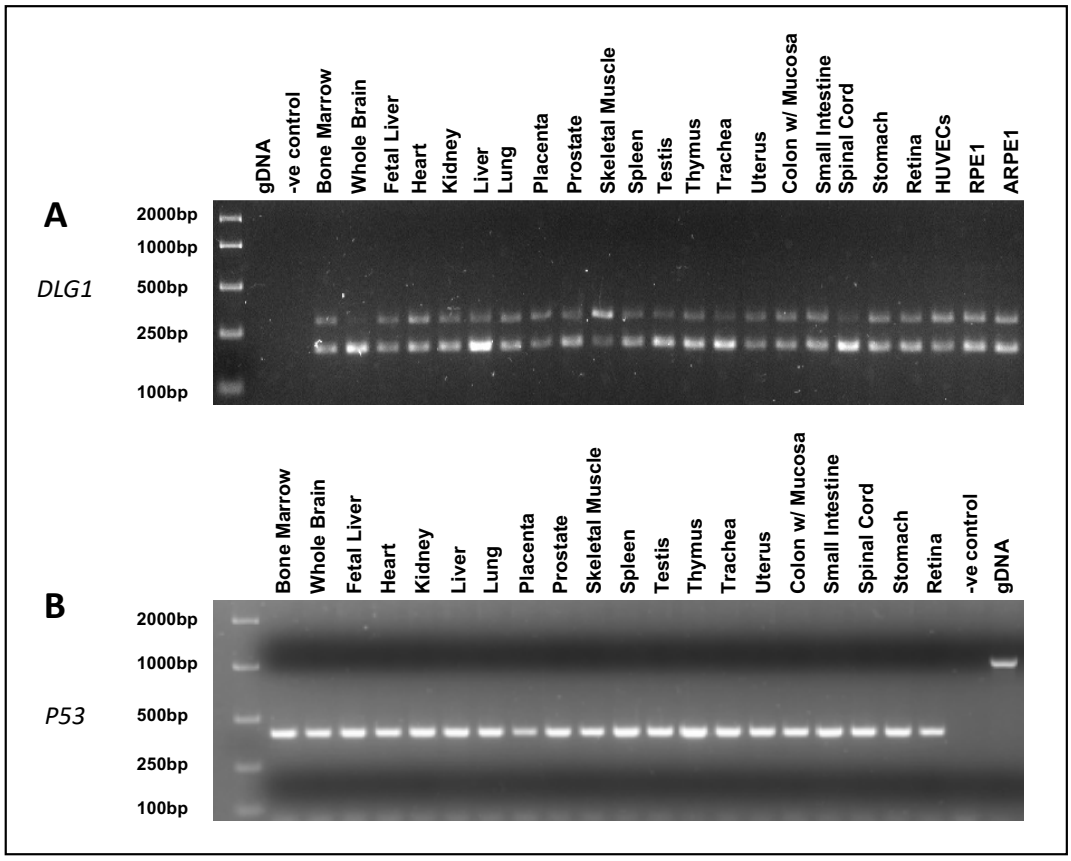


Figure 4.30 cDNA tissue expression panel for *DLG1*.
(A) Expression of *DLG1* in a range of tissues. Transcript 1, transcript 2 and transcript 6 predicted product is 289bp. These products are represented by the top band on the gel but it is not possible to distinguish which transcripts represent the band. The bottom band at size 190bp is a product of transcript 3 of *DLG1*. *DLG1* expression ranges across many tissues. Ladder used was EasyLadder I (Bioline). **(B)** 1μg/μl of total RNA from Total RNA Master Panel II (Clontech) was used for synthesis of cDNA. Retinal cDNA was purchased from Clontech (Catalog No. 637216). A panel of 20 human tissues was used and *P53* primers used to confirm the synthesis from RNA to cDNA. Water served as -ve control. Product size for cDNA was 410bp and product size expected for gDNA was 1059bp. Ladder used was EasyLadder I (Bioline). Primers are provided in Appendix 7.17.

4.2.4 Identification of common genes/variants between FEVR and Coats Disease cohorts

Due to the overlap in phenotype and the underlying genetic causes, the variant output lists between FEVR and Coats disease were compared in order to identify variants in both datasets that were present in the same gene. This revealed three genes *MUC4*, *ATN1* and *RPL14* that contained variants in a number of FEVR and Coats disease patients. The analysis only included

variants with a MAF <0.0001 (<0.01%) in ExAC, were predicted to have an effect on the coding amino acid sequence and a high CADD score. Variants were also filtered unless they passed the quality threshold control filter that was used in the bioinformatic analysis and had allele depth of 50/50 ratio for heterozygous variants. Genes were only selected if variants were found in at least 2 patients in the merged FEVR and Coats disease cohort. *MUC4*, *ATN1* and *RPL14* were the only common genes shared between the two data sets, however all variants were microsatellites that were present in ≥4/7 Coats disease patients and in ≥10/25 unsolved FEVR patients. The variants in each gene were therefore deemed to be likely non-pathogenic. This analysis was only preliminary and will need to be concluded in a greater detail in the near future.

4.3 Discussion

In this study, a panel of 20 FEVR patients were screened by WES to globally look at patients' coding regions to uncover mutations in either the known FEVR genes or new genes. The majority of these cases had previously undergone some level of pre-screening although 2 patients had undergone no previous screening. Mutations in known FEVR genes were identified in 5/20 cases. The remaining 15 FEVR -ve cases were combined with an additional 10 FEVR -ve exomes from a previous study to aid in the identification of new FEVR genes. A putative mutation was found in a new exon within a novel transcript of *LRP5*, and this transcript was verified at the mRNA level. Furthermore, a potential variant was identified in two transcripts of *DLG1*, a new potential FEVR gene highlighted as a candidate through collaborators.

4.3.1 Mutations in known FEVR genes

The analysis identified 7 potential different mutations in the known FEVR genes in 5 cases. In this study, none of the cases presented with missed mutations in the already pre-screened genes, except for patient F1308 who had been pre-screened for *TSPAN12* by Sanger sequencing, which is not able to cover large exonic deletions or insertions unlike the WES technology used in this study. All of these patients were sporadic. Unsolved exomes were also checked for variants in all the known retinal disease genes, as some of these are known to be phenocopies for FEVR such as *VCAN*, *BEST1*, and *KCNJ13* (Hejtmancik et al., 2008; Brézin et al., 2011; Pasquay et al., 2015; Panagiotou, 2018) but no putative variants were identified. Therefore, the remaining 15 cases remain unsolved.

Patient F1250, diagnosed with FEVR, was found to carry compound heterozygous variants within the *LRP5* gene. Recessive mutations in *LRP5* are commonly associated with OPPG (Gong et al., 2001) and dominant mutations with FEVR (Toomes, et al., 2004b). However, recessive *LRP5* mutations have been reported in a few FEVR families. In the original study describing recessive *LRP5* mutations in FEVR, no bone phenotypes were investigated (Jiao et al. 2004). However, it is likely that the bone phenotype was missed in these patients as it has been shown many times that FEVR patients with dominant or recessive mutations in *LRP5* have reduced bone mineral density (Gong et al., 2001; Toomes et al., 2004b; Downey et al., 2006). Patient F1250 was recruited when only one year of age and may have been too young to have displayed a bone phenotype but following this diagnosis

they can be referred for a DEXA scan or orthopaedic assessment and undergo bisphosphonate treatment if required. It was previously shown that bone phenotypes are easily missed in FEVR patients as broken bones are common in blind children (Toomes et al., 2004b) and even in recessive cases, which tend to be more severe, the bone phenotype may be subtle and only evident upon DEXA scan (Downey et al. 2006). The parents of this child are likely to be at risk of having FEVR and reduced bone mineral density as they are potentially carriers of one of the mutations identified in F1250. Following this genetic result, they will also be followed up in clinic to look for a subtle eye or bone phenotype and can be offered counselling for future pregnancies. If applicable, cascade molecular testing and counselling can be implemented to the wider family.

The first mutation p.(Arg752Trp) identified in patient F1250 has been previously reported in a UK patient with OPPG and congenital blindness (Alonso et al., 2015). This mutation maps to the third YWTD β-propeller which was documented to be responsible for binding of antagonists such as Dkk1 (Mao et al., 2001; Semënov et al., 2001). The effects of this mutation are likely to reduce Norrin-β-catenin signalling but the mechanism underlying this defect is currently unknown. The second variant identified in F1250 p.(Cys1305Tyr) was present in the LDL-R like domain. This mutation was previously documented in the thesis of Dr Sarah Hull (2016) who identified it in an infant diagnosed with poor vision, bilateral retinal folds and nystagmus. This patient was also shown to have lowered bone density indicating OPPG. Two other mutations located in the LDL-R like domain had been previously published. In

a study by Toomes et al. (2004a), a mutation p.Cys1361Gly was identified in a 6 year-old boy from Australia diagnosed with FEVR. Another reported mutation in the LDL-R like domain was a homozygous missense change, p.Glu1367Lys, identified in an affected brother and sister from a consanguineous family of European descent who were diagnosed with FEVR and retinal detachments (Jiao et al., 2004). The role of the LDL-R like domain within the Wnt or Norrin-β-catenin signalling pathway remains unknown (Macdonald and He, 2012). Given this data, both mutations look very convincing in this patient.

In FEVR patient F1057_1072, two potential variants were identified in *LRP5*, a missense variant and a splice variant. The splice site mutation, c.2318+1G>C is very convincing as mutations affecting the first two intronic nucleotides of the splice donor site are almost always pathogenic (Krawczak et al., 1992). In this case, the donor site of GT was changed to CT. Unfortunately, no RNA was available for this patient so the precise effects on splicing have not been confirmed. There are a number of scenarios which can occur if the mutation resides in the splice site. Aberrant splicing can cause intron retention (which then becomes part of the final mRNA), skipping of an exon (resulting in smaller mRNA transcript) and aberrant product or the activation of a cryptic splice site instead of the mutated splice site (Maquat, 1996). The majority of these are likely to generate premature stop codons due to the shift in the reading frame (Maquat, 2004). Although not required for this very convincing splice mutation, for those which don't affect the conserved GT-AG splice sites, it is still very challenging to establish the relationship between

a splice mutation and the disease. To overcome this problem, splicing assays using mini- and midi-gene constructs have been developed to assess splicing mutations' effects (Steffensen et al., 2014; Sangermano et al., 2016). This splicing variant was not previously reported, and it further highlights the need for screening not only the exonic regions but also the exon-intron boundaries in establishing the cause of the disease. The second mutation identified in F1057_1072 was a missense mutation p.(Gly706Arg) located in exon 10 and mapping to the third YWTD β-propeller. This mutation also looked very convincing as it was rare, fully conserved and predicted to be pathogenic by all six prediction tools. However, this variant was recently found to be annotated as a polymorphism after it was found in a heterozygous state in a FEVR patient, but segregation analysis found it to be homozygous in the patient's unaffected mother (Hull, 2016). Therefore, it is likely that F1057_1072 has a dominant form of FEVR caused by the *LRP5* splice mutation. This molecular diagnosis is consistent with the milder eye phenotype observed in this patient and the lack of severe bone pathology, as this patient was recruited as an adult. This data shows how challenging it can be to assess whether a mutation is pathogenic, especially when it is backed up with convincing data as presented in this study.

In FEVR patient F1375, a nonsense variant in *FZD4* p.(Gln505*) was identified. This mutation has previously been reported by Toomes et al. (2004b), where it was identified in an Australian FEVR family with European ancestry. The mutation occurs in the second exon and is located just outside of the DVL association domain. It results in the loss of the PDZ domain and perhaps

alteration of the DVL association domain, both of which were found to be crucial for proper signal activation (Punchihewa et al., 2009 & Bertalovitz et al., 2016). As the nonsense mutation is located in the last exon, it is not predicted to undergo NMD and could produce truncated protein. This could be functionally confirmed by techniques such as protein truncation test (PTT) which was developed to determine whether a mutation results in a truncated protein or not (Roest et al., 1993). It works by the isolation of nucleic acids, amplification of specific region of interest followed by *in vitro* transcription and translation of the product that is biotin-labelled, or N-terminus tagged and can be visualized through immunoblots (Vossen & den Dunnen, 2004).

The heterozygous *KIF11* variant p.(Arg281Serfs*34) was identified in this study in a F164 patient diagnosed with sporadic vitreoretinal dysplasia. The patient's phenotype resembled Norrie disease but no mutation in *NDP* was identified so they were enrolled in this study. Dominant *KIF11* mutations are known to cause MLCRD and CDMMR (Section 1.8.2.8). In 2014, dominant mutations in *KIF11* were also reported to cause FEVR showing phenotypic overlap between FEVR and MLCRD (Robitaille et al., 2014). F164 was not reported to have microcephaly, however, this phenotype is easily missed by ophthalmologists unless they specifically look for it (Li et al., 2016; Robitaille et al., 2014). Like this mutation, it has been shown that most mutations in *KIF11* that cause an ocular phenotype are nonsense, frameshift and splice site mutations (Ostergaard et al., 2012; Robitaille et al., 2014). This mutation has not previously been reported in FEVR, MLCRD or CDMMR.

To date, there have been seventeen *KIF11* mutations described in FEVR patients. In the initial study where Robitaille et al. (2014) have identified the phenotypic overlap between FEVR and MLCRD/CDMMR caused by *KIF11* mutations, five heterozygous mutations in *KIF11* were described in patients diagnosed with FEVR. One of these mutations (p.Arg47*) was previously published in a study by Hazan et al. (2012) where the patient was diagnosed with MLCRD but not FEVR. Further seven mutations were reported in a study by Li et al. (2016) and all patients were reported to have FEVR. Three of these patients also had microcephaly and one chorioretinopathy, confirming the phenotypic overlap described by Robitaille et al. (2014). Another four heterozygous mutations were identified in a study by Hu et al. (2016) in FEVR patients. The patients were reported to have variable FEVR phenotypes with or without MLCRD. The last study has identified one frameshift mutation in a patient diagnosed with FEVR and no other phenotypes, such as MLCRD or CDMMR were described (Rao et al., 2017). In this study, the patient carrying the *KIF11* mutation was also described to only have vitreoretinal dysplasia but no microcephaly, or other phenotypic features of MLCRD/CDMMR were documented. Similarly, the study by Hu et al. (2016) have also identified *KIF11* variants in patients who were only diagnosed with abnormal ocular phenotype. These data show that *KIF11* can only be mutated in patients with FEVR or FEVR-related retinopathies and should be part of the diagnostic testing even for patients who do not present with classical features of MLCRD or CDMMR.

4.3.2 CNVs

A large 3 exon deletion in *TSPAN12*, encompassing exons 5-7, was identified using ExomeDepth in patient F1308. *TSPAN12* had previously been screened

in this patient by Sanger sequencing but no mutation was identified, only a common synonymous homozygous SNP in exon 8 (c.765G>C, rs41623). Large intragenic deletions and/or insertions are easily missed by Sanger sequencing, whereas WES allows for fast and effective CNV analysis that is crucial in diagnostics (Ellingford et al., 2017; Ellingford et al., 2018a). It was only in recent studies that heterozygous deletions in *TSPAN12* were documented, one encompassing the entire gene and another deleting exon 4 (Okamoto et al., 2011; Miller et al., 2015; Seo et al., 2016).

ExomeDepth was successfully used for the CNV analysis in this study, however, the software generates a large number of false positives which makes it hard to use to detect CNVs in non-candidate genes. A comparison study of different CNV detection tools came to a similar conclusion by showing that ExomeDepth had the highest sensitivity of all tools tested but also contained a higher rate of false positives (Tan et al., 2014). The best option to identify CNVs is high coverage WGS data (Tan et al., 2014).

Apart from *TSPAN12*, rare large exonic deletions in *LRP5* and *FZD4* have previously been reported (Narumi et al., 2010; Seo et al., 2015). Exonic deletions in *NDP* are common, presumably due to their easy detection due to *NDP* being located on the X chromosome (Arai et al., 2014; Chen et al., 1993b; Halpin & Sims, 2008; Rivera-Vega et al., 2005; Schuback et al., 1995). The number of deletions in autosomal dominant genes will have been vastly underreported due to the limitations of Sanger sequencing.

The deletion of exons 5-7 causes an in-frame 109 amino acid deletion and the domains which would be deleted are the third transmembrane domain and most of the LEL domain. TSPAN12 is part of the tetraspanin protein family and it is an auxiliary protein that contains four TM domains, SEL and LEL which is known to be important for its function (Rubinstein et al., 1996). TSPAN12 is described as an auxiliary protein which helps form large signalling complexes on the cell surface (Junge et al., 2009). It has an important role in Norrin-β-catenin signalling by facilitating ligand selectivity (Lai et al., 2017).

The *TSPAN12* deletion is likely to have a drastic effect on the stability, trafficking and folding of the protein. The LEL domain mediates interactions with proteins and is involved in homodimerisation. The TM domains are involved in inter- and intramolecular interactions (Stipp et al., 2003) and are crucial for the stability, folding and transport of TSPAN12 (Kovalenko et al., 2005). The importance of these regions of the protein are also highlighted by the previous identification of two missense mutations in these domains which cause FEVR, one in the third TM domain (p.Leu101His) and another in the LEL domain (p.Met210Arg) (Poulter et al., 2010).

The identification of these mutations within the known genes of the Norrin-β-catenin signalling pathway helps deepen the understanding of FEVR, the mutated proteins and phenotype-genotype correlations. As previously documented, only 50% of the diagnosed FEVR cases carry mutations in the currently known genes (Salvo et al., 2015), therefore, it is essential to identify

new genes implicated in this disorder and to establish their involvement in the retinal vasculature.

4.3.3 *LRP5* transcript 2

In FEVR patient F1304_1362, a variant p.(Thr6Ala) in the new exon (Ex8a) of transcript 2 (NM_001291902.1) of *LRP5* was identified. This child was previously found to have a convincing potential homozygous missense mutation in *NR2E3* (NM_014249) GRCh37, Chr.15:72105736, c.755T>C, p.(Leu252Pro) (Panagiotou, 2018). This variant was rare and predicted to be pathogenic and the patient had a large region of homozygosity surrounding *NR2E3* and came from a consanguineous union. However, the FEVR phenotype has not previously been associated with mutations in *NR2E3* and follow-up clinical analysis of this patient and sampling of family DNA is ongoing. Therefore, as it was still unsolved, this case was included in the present study.

The new transcript, *LRP5* transcript 2 (NM_001291902.1), was only recently added to RefSeq after the addition of RNAseq data (Pruitt et al., 2014). This isoform is predicted to produce a shorter protein of 1,034 amino acids. The signal peptide, the first two YWTD domains and the first EGF domain are not present in the predicted protein. The protein would not contain binding sites for the Wnt ligands, antagonists such as WISE and the location of putative binding sites for Norrin on LRP5's YTWD domains 1 and 2 would also be affected by the shorter protein (Ke et al., 2013).

An expression panel was generated to determine whether the newly annotated *LRP5* transcript 2 was real, whether it was expressed in the retinal tissues and cell lines and if it showed differential expression to transcript 1 (NM_002335.3). The results showed that both transcripts appear to be ubiquitously expressed (Figure 4.24). This data for the canonical transcript 1 is consistent with numerous studies which have shown that it is widely expressed in liver, pancreas, prostate, placenta, small intestine, heart, lung, stomach, skeletal muscle, kidney, spleen, thymus, testis, colon, and brain (Kim et al., 1998; Hey, 1998; Kato et al., 2002). To verify whether the new transcript is encoding, a WB would need to be performed. In the WB performed in Chapter 3 (Section 3.2.6.2), isoform 2 was not detected, however, this is most likely due to the image exposure as endogenous full length LRP5 was also not detected. WB analysis would therefore need to be repeated in cell lysates generated without LRP5 expression constructs. The antibody used in the initial experiment targets the C-terminal intracellular part of LRP5 and could, therefore, be used to detect if the shorter LRP5 protein is generated from isoform 2. Immunohistochemistry could also be used to detect the shorter polypeptide in the retinal cells (Kato et al., 2002). However, a specific antibody would need to be used that would only target the new protein.

Since LRP6 has a very similar role to LRP5, being a co-receptor in Wnt-β-catenin signalling pathway, a search for a similar shorter transcript was carried out. However, *LRP6* only produces one transcript NM_002336.2 comprising of 23 exons and generates 1,613 amino acid protein as recorded by RefSeq on UCSC Genome browser (https://genome.ucsc.edu, accessed July 2018).

There is no other transcript for *LRP6* that would be similar to the transcript 2 of *LRP5*.

The new transcript could possibly be a long non-coding RNA (lncRNA). The importance of non-coding RNAs has grown considerably over recent years and lncRNAs have been shown to be key regulators of gene expression (Bhat et al., 2016; Chen & Carmichael, 2010). There are many reported lncRNAs in the eye and these are thought to play a role in development and disease (Jiang et al., 2016; Karali & Banfi, 2018; Zhou et al., 2015).

Since the protein produced from transcript 2 would be missing the signal peptide, it is unlikely to act as a receptor and would likely have a novel unknown function. With the missing signal peptide and first two YWTD domains and its retained ability to interact with the destruction complex, the TCF activity might only be reduced and not completely halted. This would replicate in the milder phenotype that is seen in FEVR patients. This is merely a hypothesis and the role of the new transcript and protein would still need to be confirmed and evaluated.

The novel exons of transcript 2 have not been screened in FEVR cohorts and they are also not present on targeted arrays, such as the Manchester 175 eye gene service, therefore, they may be more mutations underlying FEVR which could have been missed.

4.3.4 Mutation in candidate gene *DLG1*

A heterozygous missense variant in *DLG1* p.(Arg62Gln) was identified in patient F561 which partly segregated with the phenotype in the family (Figure 4.27). This family was not found to harbour any mutations in any of the FEVR genes assessed by WES and ExomeDepth analysis. A potential heterozygous frameshift variant in *KLHL7* was previously found in this family but it didn't segregate with the affected family members so was discarded (Panagiotou, 2018). Mutations in *KLHL7* are associated with adRP (Friedman et al., 2009).

The proband F561 presented with unilateral disease which contributed to her diagnosis of PHPV. The affected son F562 presented with bilateral FEVR, which would raise the question of the mother having a somatic disease as seen in the study by Black et al. (1999) and Coats disease (Chapter 3). However, if the variant in *DLG1* is real, this would argue against this hypothesis and would confirm the marked asymmetry that is seen in some FEVR cases (Robitaille et al., 2011). The mother of the proband F566 was not reported to be affected but was found to be heterozygous for the mutation suggesting that she may be an asymptomatic mutation carrier or a case of reduced penetrance. With a very high rate of asymptomatic carriers and reduced penetrance in FEVR, it complicates the segregation within the disease (Toomes et al., 2004b; Toomes et al., 2004a; Poulter et al., 2010; Panagiotou et al., 2017). Therefore, only the segregation in affected members can be trusted in FEVR, although there are rare reports of phenocopies in FEVR families which can complicate this (Poulter et al., 2012).

DLG1 was highlighted as a potential gene in *in vivo* experiments by a collaborator (unpublished confidential data from Professor Jeremy Nathans). Upon this discovery, the FEVR cohort dataset was searched for any variants in *DLG1*. DLG1 belongs to a family of proteins called membrane-associated guanylate kinase homologs (MAGUKs) which are localised at cell-cell junctions and are involved in signalling (Azim et al., 1995). Dlg1 was previously shown to be involved in inhibition of axonal stimulation of myelination when interacting with PTEN and causing demyelination that is distinctive of neuropathies (Cotter et al., 2010). DLG1 encodes the synaptic PDZ proteins which interact with the N-methyl-D-aspartate type glutamate receptor (NMDAR) receptor. This receptor is thought to be underactive in schizophrenia patients (Uezato et al., 2015). In a study by Uezato et al. (2017), it was found that genetic variants in *DLG1* have an effect on its interaction and regulation with the glutamate receptors. This disrupted interaction was found to lead to early-onset schizophrenia at a specific stage of brain development.

APC, a component of the Wnt and Norrin signalling pathways, was shown to promote localisation of DLG1 (Anastas and Moon, 2013) through DLG1 binding to APC via its PDZ domains (Matsumine et al., 1996). Dlg1 was found to be a tumour suppressor with a role in controlling proliferation (Humbert et al., 2003). It was shown to be directly involved in PI3K-Akt signalling and also in L1CAM signalling pathways. PI3K-Akt signalling is involved in angiogenesis, proliferation and growth which makes it a good candidate for playing a role in the vascularisation of the retina (Cho et al., 2009).

DLG1 contains 6 different transcripts and the variant was present in transcripts 4 and 5 which contain an alternative start site to transcripts 1-3 and 6 (Figure 4.25). These shorter isoforms produce a smaller protein that is missing the LS27 domain, which is responsible for the formation of a DLG1 oligomer. *DLG1* has been previously documented to be deleted in 3q29 microdeletion syndrome in heterozygous state. These patients presented with mental retardation, microcephaly and speech delay (Carroll et al., 2011). No FEVR has been reported in cases with 3q29 deletions but this doesn't mean it isn't present. As previously stated, many FEVR mutation carriers are mild or asymptomatic and often in children with multiple anomalies sight problems can be missed. A similar situation of undiagnosed FEVR was seen in many patients with *CTNNB1* deletion syndrome (Panagiotou et al., 2017).

As this gene was identified as a potential candidate gene in a mouse study of defective vasculature from a collaborator, it could potentially play a role in FEVR or FEVR related retinopathies. Dlg1 KO mouse was found to be embryonic lethal (http://www.informatics.jax.org/allele/allgenoviews/MGI:3710242) and this would explain why a germline Mendelian mutation would need to be in a tissue specific transcript. Mutation hotspots in exons only expressed in certain tissues have been reported before, example is the *ORF15* in RP (Vervoort et al., 2000).

Mouse studies have already been performed but *in vitro* studies that show the involvement of the particular transcript and the mutation identified in this study in retinal vasculature are also necessary. The determination of DLG1 as a

potential component of Norrin-β-catenin pathway or its localisation to the retinal layers would require the use of functional assays such as binding assays, co-localisation assays, proximity ligation assay, immunofluorescence (IF), co-immunoprecipitation, luciferase assay or mass spectrometry (Fredriksson et al., 2002; Xu et al., 2004; Smallwood et al., 2007; Qin et al., 2008; Junge et al., 2009; Ke et al., 2013). IF could be carried out to see the specific localisation of DLG1 within the retinal layers, as its expression was found in retinal cDNA (Figure 4.30). Co-localisation experiments would indicate whether DLG1 co-localises with any of the proteins of the Norrin-β-catenin signalling pathway. To further establish whether there is an interaction between molecules, binding assays such as co-immunoprecipitation could be used to purify the target, DLG1, together with other molecules, such as FZD4, LRP5, TSPAN12 or NDP, and evaluate whether any interaction occurs. Mass spectrometry is one of the most common methods used in understanding the biological functions of proteins and how they interact. It allows for identification of protein-protein interactions which could be used in identifying the targets which DLG1 interacts with. It has been developed to identify a single or multiple protein of interest (Smits and Vermeulen, 2016). Other methods for identifying protein-protein interactions when in close proximity is the proximity ligation assay (Fredriksson et al., 2002) (Section 3.3.8). These assays would be carried out to establish whether DLG1 does interact with any of the known components of the Norrin-β-catenin signalling pathway or whether there is a new pathway that could be involved in the vascularisation of the retina. DLG1 could be interacting with ATOH7, ZNF408 or RCBTB1 proteins which have all been described to be involved in the pathogenicity of FEVR but are not known

components of Norrin-β-catenin signalling pathway (Collin et al., 2013; Khan et al., 2012; Wu et al., 2016).

Additional patients with *DLG1* mutations would now need to be identified to confirm that *DLG1* is a new FEVR gene. Ideally, collaborators would be contacted to check through their cohorts for any variants, but since this data is confidential, other routes would need to be taken. These could include searching the variant platforms of the European Retinal Disease Consortium (ERDC) (https://www.erdc.info/index.php/candidate-ird-genes), the UK-Inherited Retinal Disease Consortium (IRDC) or the Matchmaker (https://www.matchmakerexchange.org) (Philippakis et al., 2015). These sources provide a platform where variants in various genetic eye diseases with unknown pathogenic significance can be shared to help identify multiple cases with variants in the same gene and overlapping phenotypes. A further resource is the 100,000 Genomes project (http://www.genomicsengland.co.uk) which includes a number of FEVR patients and can be used to consult variants in candidate genes.

The discovery of *DLG1* variant in FEVR patient, which segregated within the family strengthens the *in vivo* mouse studies (unpublished data from collaborator) that this gene could be involved in the proper vascularisation of the retina. However, further studies need to be performed to fully confirm DLG1's role in the Norrin-β-catenin signalling pathway.

4.3.5 Shared genes between FEVR and Coats disease cohorts

The analysis of shared genes between FEVR and Coats disease cohorts identified 3 genes, *MUC4*, *RPL14* and *ANT1. RPL14* and *MUC4* have not been previously linked to an eye phenotype. However, mouse Atn1 was shown to be expressed in developing retinas and was also shown to act as a repressor which interacts with Tlx, a receptor in ocular development (Zhang et al., 2006). *Tlx*$^{-/-}$ mice were shown to have loss of visual acuity, reduced cell number and thinning of INL, IPL, ONL, OPL and photoreceptor layers. Tlx was shown to be required for retina-specific differentiation, proliferation but, interestingly, not cell type specification. Interaction of Atn1 to Tlx allows for the precise differentiation and proliferation of retinal progenitor cells (Zhang et al., 2006).

As this analysis was preliminary, the data sets generated in this study require much deeper analysis in order to aid the identification of genes that may be causative of these rare disorders. To aid in this analysis, there are some upcoming bioinformatic tools that aim to help in the identification of causative mutations. One such new tool called DOMINO has only recently been developed and it aids in identifying pathogenic genes within dominant disorders (Quinodoz et al., 2017). The tool works on assessing the likelihood that any given gene will harbour dominant mutations. Identifying mutations in the same gene, such as *LRP5* in this study, in multiple patients would also help in confirming that the particular gene is involved in the pathogenesis of the disease and it would also strengthen the pathogenicity of the already identified mutations. This was previously shown in a case of identifying mutations in the

CTNNB1 gene in patients with FEVR where two familial and one sporadic case have all segregated within the family members showing the same phenotype (Panagiotou et al., 2017). It is a lot more challenging to prove a gene pathogenic when only one family is shown to be carrying the variant. This was shown to be challenging in this study as the unsolved exomes of FEVR require further analysis to identify causative mutations as no family members were available for segregation. A very efficient way of assessing if the variant could be pathogenic is to look at the previously mentioned online platforms such as ERDC, IRDC or the Matchmaker where variants can be exchanged within the research community (Section 4.3.4). Another possible way of searching for new candidate genes is in published data sets, such as the transcriptome data set recently published for *ZNF408* (Karjosukarso et al., 2018).

4.3.6 WES methodology

The bioinformatics pipeline used in this study to identify variants and novel genes was previously used in a number of gene identification studies in their team (El-Asrag et al., 2015; Panagiotou et al., 2017). As seen from the successful identification of mutations in known FEVR genes, this pipeline is very effective. This was also shown by the identification of 3 exon deletion in *TSPAN12* in patient F1308 (Section 4.2.2.5), which was not identified when screening *TSPAN12* by Sanger sequencing. However, even though WES covers all coding regions, it still lacks the identification of structural variants such as inversions or deep intronic variants. These non-coding intronic variants have been shown to be frequently linked to many diseases. One such example is the deep intronic mutation in *CEP290* identified in a study by den Hollander et al. (2006) in patients with LCA. They had found that a deep

intronic mutation created a splice site and inserted a cryptic exon into the mRNA of *CEP290*. This mutation was found to be present in 21% of patients in their cohort. Another study by Cipriani et al. (2017) had similar findings where duplication events within the non-coding regions were seen to cause North Carolina macular dystrophy. WGS has been shown to be more advanced in identifying such structural changes or intronic variants as it covers all regions within the patients' genomes (Ellingford et al., 2016). In a study by Belkadi et al. (2015), data from WES and WGS were compared between 6 unrelated cases. They had found that the WGS has successfully identified 3% more coding SNVs which were missed by WES. Large duplication and inversion causative of the disease phenotypes were also identified by WGS in various eye diseases (Small et al., 2016; Lord et al., 2017). This confirms that WGS is more superior to WES and should be used more routinely in patients' diagnostics in near future.

As WGS is becoming increasingly popular in genetic research, there are some other points that need to be considered. In a study by Royer-Bertrand and Rivolta (2015), it was highlighted that the diagnosis of the disease would have to be very accurate to select for the appropriate pipeline and filtering. To account for sequencing errors and false positives, numerous avenues could be explored, such as providing more optimised reference genome or updating of the common variant databases that are used within the pipeline. It is also very important to note that undergoing WGS carries the probability of incidental findings that are not specific to the patient's diagnosis and, therefore, creating ethical issues (Section 5.7).

4.4 Summary

In this chapter, a cohort of 20 unsolved FEVR sporadic and familial cases were put forward for WES to identify causative variants in the known FEVR genes and to identify potential genes that may be involved in this disorder. Seven variants in five individuals (all sporadic) presented with variants in known FEVR genes. In an additional cohort of 10 FEVR -ve patients, one variant in a familial case was identified in a potential new candidate gene for FEVR, *DLG1*. Additionally, a variant was also identified in a new transcript of *LRP5* in a familial case of FEVR.

To further allow for the search of novel genes, the WES data from Coats disease patients (Chapter 3) were compared with the FEVR data but no candidates were identified. However, the comparison between these two data sets was only preliminary and requires much deeper analysis which will be carried out in the near future.

5 Chapter: General Discussion

5.1 Key findings

The aim of this study was to identify novel genes and mutations underlying Coats disease and FEVR. Chapter 3 describes WES of FFPE-extracted DNA from Coats disease eye globes leading to the identification of *LRP5* as a new Coats disease gene. Chapter 4 describes WES in a cohort of 20 FEVR patients resulting in the identification of mutations in known FEVR genes in five patients. Chapter 4 also describes the search for novel genes by reanalysing exome data from 10 unsolved FEVR patients [generated in a separate study (Panagiotou, 2018)] with the current dataset, leading to the identification of a potential mutation in a new candidate FEVR gene, *DLG1*, and a potential FEVR mutation in a newly defined *LRP5* transcript.

The aim of this research is to increase the understanding of Coats disease and FEVR and provide insights into the cellular mechanisms that control normal retinal vasculature development and/or lead to abnormal vasculature and blindness. Although both Coats disease and FEVR are rare disorders, abnormal retinal vasculature is a feature of many common causes of blindness including diabetic retinopathy (DR) (Eshaq et al., 2017), age-related macular degeneration (AMD) (Al-Zamil and Yassin, 2017) and ROP (Drenser, 2016), and so this research may also increase the understanding of these disorders. Furthermore, the majority of known Coats and FEVR genes encode components of the Norrin-β-catenin signalling pathway (Ye et al., 2010), and

this pathway has been shown to play a role in the development of the vasculature of the brain (Engelhardt and Liebner, 2014), ear (Rehm et al., 2002) and ovaries (Hsieh et al., 2005), as well as the retina (Xu et al., 2004). Proteins within this pathway have also been linked to many different forms of cancer (Bassett et al., 2016; Planutis et al., 2014; Reinartz et al., 2016). Therefore, the identification of new mutations and genes which affect Norrin-β-catenin signalling will provide further insight into this complex pathway. Consequently, this research has the potential to have wide-reaching impact and may aid in the development of treatments and therapies for a host of disorders.

5.2 Identifying mutated genes underlying Coats disease

Only one study has previously investigated somatic mutations in Coats disease patients, and these were identified using Sanger sequencing (Black et al., 1999). PCR-based Sanger sequencing limits the ability to identify large structural variants or to globally look across all genes, plus PCR amplification can be problematic in FFPE-derived DNA. In this study, NGS technology, in particular WES, was used to advance on this earlier study to identify additional genes underlying Coats disease. This analysis identified a heterozygous mutation in *LRP5* c.2951A>G p.(Tyr984Cys) in one Coats disease eye. Further bioinformatic and functional experiments confirmed that the variant is highly likely to be pathogenic, therefore identifying *LRP5* as a new gene involved in the pathogenesis of Coats disease. For the remaining eyes, additional candidate variants were identified but these require further experiments to prove their role in the disease.

One limitation of this study was the lack of matched gDNA from the Coats patients who had their eyes enucleated. The availability of this resource would have allowed for the germline and somatic mutations to be distinguished in each patient. This information would allow all the somatic mutations in the eye to be identified, and this shorter list could be interrogated as candidate Coats genes. In tumour-derived samples, the list of somatic mutations is usually substantial and often clonal in nature due to mutations in oncogenes and tumour suppressor genes (Harris et al., 2017). However, in Coats disease, the variant list is expected to be much shorter and the variants are expected to be present in the majority of the cells making this a powerful dataset for gene discovery.

Distinguishing germline and somatic mutations would also answer the very important question of whether Coats disease is only a somatic disease, or whether it is caused by germline mutations or a combination of germline and somatic mutations. The Black study identified a somatic mutation in *NDP* (Black et al., 1999) but the Wu study identified a germline mutation in *RCBTB1* in Coats disease (Wu et al., 2016). This suggests that both scenarios exist.

FEVR is known for its extremely variable clinical presentation and a significant number of patients are asymptomatic (Section 1.8). In the FEVR cases who do display symptoms, there is often marked asymmetry between the severity of the eyes and one eye can appear unaffected (Toomes & Downey, 2005) (updated: September 22, 2011). Clearly, there is the potential for some FEVR patients to be diagnosed with Coats disease in this situation. Furthermore,

although over 95% of Coats cases are reported to have unilateral disease (Shields et al., 2001b), the latest imaging technologies, such as ultra-wide field angiography, are now showing that significant numbers have avascularity in the fellow eye. For example, Blair and colleagues found abnormalities in the unaffected eye in 22/32 (68%) Coats cases (Blair et al., 2013) and Rabiolo and colleagues reported avascularity in 7/9 (77%) unaffected eyes (Rabiolo et al., 2017). These findings suggest that germline mutations play a role in Coats disease or that the clinical diagnosis of FEVR and Coats disease overlap significantly. Interestingly, there is also a hypothesis suggesting that FEVR may not be a simple Mendelian disorder but that there may be genetic modifiers which are responsible for the variable phenotype seen in patients. There is preliminary evidence suggesting that those patients with severe disease harbour two pathogenic mutations, which can either be in the same gene or in two separate genes (Poulter et al., 2012; Panagiotou, 2018). In this situation, a Coats patient may have one germline mutation and a second somatic mutation in the retina.

A further issue that needs to be considered is the nature of the FFPE data and how it affects the bioinformatics analysis. One of the first studies to carry out NGS sequencing from FFPE tissues was Schweiger et al. (2009). In their study, comparison between the genome-wide parallel sequencing from matched frozen and FFPE breast tissue from the same patient, with ranging age ischemic times upon preparation was made. The study showed that the number of unknown SNPs were higher in FFPE samples. The same phenomenon had been observed by other studies showing that the FFPE

embedded samples contain more sequence variants in comparison to frozen tissues (Quach et al., 2004 & Hedegaard et al., 2014). However, in another study, targeted sequencing was performed on carcinoma fresh frozen (FF) and FFPE paired samples and it was shown that the same variants were detected in both tissue formats (Spencer et al., 2013). Therefore, there is a need for developing bioinformatics pipelines that would concentrate on identifying and removing as many FFPE induced artefacts as possible to minimise the strain that is put on the analysis.

In the original design of this experiment, blood-derived DNA was expected to be available for WES to investigate these possibilities. Unfortunately, it was more difficult than expected to track down the donors of the eyes to obtain a blood sample. The private health care system in the USA meant that there was no continuity of patient care within one organisation, as seen in the NHS system utilised in the UK. Therefore, once the patient had undergone enucleation, they were no longer seen by the ophthalmologist and their contact details were not kept up to date. This made it extremely difficult to track down the patients and no genomic samples were available within the time-course of this study. As an alternative, attempts were made to extract DNA from parts of the eyes derived from different embryonic lineages, but these were unsuccessful. Thankfully, the USA team have persisted, and four patients have now been located and DNA collected from two of these. Sadly, the patient with the *LRP5* mutation and the putative *ISPD* mutation has not been located yet.
Another limitation of this study was the fact that only FFPE tissue was available for genetic analysis. FFPE remains the best way of preserving histological

tissues and samples, but as this study shows, extracting good quality DNA is problematic (Section 3.3.2). Although cryo-preservation can preserve tissue structure without damaging DNA (Mazur, 1970), the storage difficulties mean most archives are still prepared using FFPE (Hedegaard et al., 2014). In the past, Coats diseased eyes were frequently enucleated as they were misdiagnosed as retinoblastoma (Huang et al., 2010). New imaging techniques have now made it possible to distinguish between these two disorders (Soliman et al., 2016) so enucleations are now only performed as a last resort in cases where the eye is painful, so that circadian rhythms are preserved where possible (Flynn-Evans et al., 2014). This has meant that the availability of new Coats disease eyes is extremely limited so historical FFPE archives, such as the one utilised in this study, are often the only source of Coats disease tissue. Fortunately, at the end of this study, a source of cryo-preserved Coats tissue (vitreoretinal membranes) became available through Professor Birgit Lorenz (Justus-Liebig-Universitat Giessen, Germany). This sample also comes with matched gDNA and parental DNA. A further Coats eye enucleated by Alex Levin, the main collaborator on this study, has also been cryo-preserved and stored and will hopefully be available for future analysis once consent has been obtained.

The availability of a third round of WES data from the FFPE samples, WES data from matched gDNA from a proportion of the enucleated patients, and matched genomic/somatic NGS data from the two cryo-preserved Coats eyes will greatly enhance the dataset generated in this study and will hopefully lead to the identification of further genes underlying Coats disease.

5.3 The future of FFPE

As FFPE remains the best way of preserving histological tissues and samples, it is crucial that new protocols are derived that will help improve upon the damage introduced to the DNA by its fixation process. The need for fully optimised protocols for DNA extraction, such as the one described in this study, are essential for samples of all ages. The bioinformatics analysis also needs to concentrate on creating new pipelines that would focus on removing or minimising the most common sequencing artefacts in FFPE samples. The need for these procedures is high as working with FFPE samples is not only very challenging but it is also extremely time consuming to achieve a reliable result. Cryo-preservation is another technique that has partly replaced the formalin fixation of tissues, in order to avoid the difficulties of extracting good quality, amplifiable DNA (Mazur, 1970) and it also preserves the tissues for an unlimited period of time. However, most archives still include FFPE fixation as samples are easier to collect and store for a prolonged period of time (Hedegaard et al., 2014). It is therefore of value to create these protocols as FFPE fixation will most likely still be used in the near future.

5.4 Identifying mutations and new genes in FEVR

Chapter 4 described the identification of mutations in known FEVR genes and also the identification of a potential mutation in a new candidate FEVR gene, *DLG1*, and a potential FEVR mutation in a newly defined *LRP5* transcript.

5.4.1 Mutations in known FEVR genes

New mutations in known FEVR genes were identified in 5/20 patients (25%) using WES technology. This technology was shown to be extremely efficient in finding the underlying cause of patients' phenotypes. Identified mutations were found in genes that make up the Norrin-β-catenin signalling pathway (except for *KIF11*), which have already been well established to play a role in FEVR (Chen et al., 1993a; Junge et al., 2009; Robitaille et al., 2002; Robitaille et al., 2014; Toomes et al., 2004b).

It is important to point out that one of the pathogenic mutations was a CNV in *TSPAN12* which was previously missed by standard Sanger sequencing methods, further confirming that NGS is essential for identification of large structural variants (Ellingford et al., 2017; Lord et al., 2017; Ellingford et al., 2018a). Therefore, CNV analysis should be incorporated in the diagnostic testing of FEVR patients to prevent prolonged investigation into the discovery of pathogenic mutations and possibly a delayed treatment. The identification of more mutations in the known FEVR genes allows for further understanding into the mechanisms of this pathway and its role in retinal vasculature.

5.4.2 *LRP5* transcript 2

In another patient, a variant c.16A>G, p.(Thr6Ala) in transcript 2 of *LRP5* was identified and this transcript was found to have expression in the retina as well as retinal and vascular cell lines. The precise role of this shorter polypeptide in retinal vasculature and in Norrin-β-catenin signalling has not been previously studied and therefore remains to be elucidated. However, as the patient was diagnosed with FEVR and no other putative mutations have been identified,

either by WES or CNV analysis, it is fair to assume that this shorter transcript also plays an important role in the retinal vasculature. However, it is not clear whether this transcript still acts through the same Norrin-β-catenin signalling pathway or whether it is involved in other processes which play a role in retinal vasculature but are still unknown.

5.4.3 *DLG1* as a candidate gene for FEVR

A potential candidate gene, *DLG1,* was discovered to play a role in retinal vasculature by a collaborator (unpublished data), and a subsequent analysis of a cohort of FEVR -ve patients has identified a variant in *DLG1,* c.185G>A; p.(Arg62Gln). Two additional family members were also found to carry the same mutation where the mother of the affected patient was found to be an asymptomatic carrier. It is known that there is a high rate of asymptomatic carriers and reduced penetrance in FEVR, which can make it difficult to carry out segregation within the disease and, therefore, more variants in more families need to be identified to strengthen the hypothesis that *DLG1* is involved in pathogenesis of FEVR (Toomes et al., 2004a; Toomes et al., 2004b; Poulter et al., 2010; Panagiotou et al., 2017). Prior to this study, *DLG1* has never been documented to play a role in retinal vasculature but given the current *in vivo* experiments provided by a collaborator and the fact it was previously shown to be directly involved in PI3K-Akt signalling responsible for angiogenesis, proliferation and growth (Cho et al., 2009) it seems to be a very promising candidate.

5.5 Identifying the remaining genes and mutations underlying FEVR and Coats Disease

If one assumes that the mutations in *DLG1* and the new *LRP5* transcript are real, then 23 FEVR patients remain unsolved in this study. Similarly, six Coats eyes remain genetically unsolved, although this number will hopefully drop when further WES analysis is undertaken. There are a number of different scenarios which could explain these findings. The simplest one is that there are new genes involved in the pathogenesis of FEVR/Coats which remain to be identified. If this is true, the data in this study suggests that the new genes are rare and only explain a handful of cases as no common genes were identified in the datasets (Section 4.3.5). This can lead to problems in proving the pathogenic nature of a variant/gene as additional mutations and families are required to satisfy the burden of proof that a gene is causative for a disorder. Potential ways of overcoming this issue involve collaborating with other FEVR/Coats researchers by either sharing variant lists or candidate genes. The FEVR research community collaborate extensively and have shared and screened candidates in the past (Collin et al., 2013; Dixon et al., 2016; Musada et al., 2016; Panagiotou et al., 2017; Rao et al., 2017). In the future, it is hoped that raw NGS data can be shared, and ethical approval has just been obtained to share genetic data with collaborating researchers. Alternative strategies include sharing variants within the wider inherited retinal disease community. The Leeds team participate in two consortia which share variants for all inherited retinal diseases, the UK IRDC and the ERDC (www.erdc.info). There are also online platforms, such as Matchmaker exchange, where variants can be exchanged with the whole research

community (Arachchi et al., 2018). Presenting data at conferences is another opportunity.

Another possible explanation for the unsolved exomes is that the mutations have been missed using this technology. Although WES is far superior to single gene-based screens or panel screening strategies (LaDuca et al., 2017; Clark et al., 2018) it does miss certain mutations. Some regions of the genome are not covered in WES including GC-rich regions, repetitive genes and intronic regions, which could harbour mutations affecting splicing or transcription (Carss et al., 2017). Although CNVs can be detected using WES data, such as the *TSPAN12* deletion identified in this study (Section 4.3.2), structural variants such as inversions can be missed (Lord et al., 2017). There are plenty of examples in the eye field where similar scenarios have been found to account for a significant number of patients. For example, there is a deep intronic splicing mutation in *CEP290* which accounts for 20% of LCA cases in Northern Europe (den Hollander et al., 2006). Similarly, North Carolina macular dystrophy has been shown to be caused by non-coding mutations which are believed to disrupt the transcription of retinal genes (Small et al., 2016; Cipriani et al., 2017). Mutations could also be present in microRNAs or long non-coding RNAs as these weren't screened in the present study but have been shown to cause retinal disease (Conte et al., 2015; Karali and Banfi, 2018). Performing WGS in this negative cohort may therefore identify "missing" mutations (Ellingford et al., 2016; Ellingford et al., 2018b), although given the huge amount of data generated in WGS, additional

strategies may need to be implemented to help identify the pathogenic variants, such as RNAseq and linkage analysis.

There is also the possibility that mutations are present in the dataset but have not been selected due to incorrect assumptions about the nature of the mutations being applied to the dataset and/or limitations in the bioinformatic pipeline. Obvious deficits include not analysing mutations in the UTRs of genes and removing synonymous variants, even though both are known to cause disease (Braun et al., 2013; Coppieters et al., 2015). The filtering of variants in this study was highly reliant on *in silico* prediction tools and in particular CADD scores. Even though variants were conservatively filtered to only remove those with a CADD score below 10, there are examples of pathogenic missense mutations with scores below this threshold such as the p.(Glu257Lys) *NMNAT1* mutation in LCA which has a CADD score of 9 (Koenekoop et al., 2012).

The research presented in this thesis is based on the premise that FEVR and Coats disease are solely caused by genetic mutations. Clearly, genetic mutations have been identified in a significant proportion of FEVR cases and in a few instances of Coats disease so there is a clear justification for this hypothesis. However, it is possible that a proportion of these diseases is not genetic, and this could explain why identifying mutations in the remaining samples is problematic. Non-genetic causes of the phenotype could be due to infection, oxidative stress, inflammation, ischemia or some other unidentified environmental cause. Infection with the roundworm parasite *Toxocara canis*

can lead to an eye phenotype resembling FEVR and toxocariasis is a differential diagnosis of FEVR (Toomes and Downey, 2005; Ahn et al., 2014). Similarly, ROP is another differential diagnosis for FEVR and this disorder is traditionally caused by high-oxygen levels in premature babies (Nichols and Lambertsen, 1969) but has also recently been linked to *in utero* light exposure (Rao et al., 2013). Oxidative stress and inflammation have been shown to play a role in the development of ROP, DR and AMD (Rivera et al., 2017) so it's possible that FEVR and Coats are also affected by similar processes.

5.6 Benefits of molecular diagnosis

The mutations identified in the current study will lead to immediate impact for the patients and their families; identifying mutations confirms the clinical diagnosis, facilitates molecular testing of family members and allows accurate genetic counselling. A genetic diagnosis of an asymptomatic patient allows them to be monitored so that interventions can be taken at the first signs of retinal ischemia, potentially saving sight. Furthermore, as mutations in *LRP5* are known to cause osteoporosis (Downey et al., 2006; Toomes et al., 2004b), the patients identified with *LRP5* mutations can be referred for a DEXA scan and treated with bisphosphonates, or similar medication, if applicable (Streeten et al., 2008). All of these benefits will also be relevant to Coats patients if the mutations are found to be germline.

A future benefit of this research is that molecularly diagnosed patients can often enrol in clinical trials, which aids the development of new therapies. A central database called Leiden Open Variation Database (LOVD) of molecularly characterised FEVR patients from multiple international centres is

available at Leeds and this resource helps recruit patients for clinical trials, a major bottle neck for rare diseases (http://www.lovd.nl/3.0/home) (Fokkema et al., 2011). Once treatments are approved for licensing by the Medicines and Healthcare products Regulatory Agency (MHRA), or similar international establishments, a molecular diagnosis is often a pre-requisite required to access many emerging therapies (Moore, 2017).

5.7 Incidental & secondary findings

One of the main ethical issues that accompanies NGS technologies is how to handle additional mutations identified outside the objectives of the recruiting study, so called incidental findings. After consulting widely with the genomics community, the American College of Medical Genetics and Genomics (ACMG) recommended that only actionable germline mutations should be reported back to the patients (Green et al. 2013). Actionable findings are those where medical intervention lowers risks of mortality and morbidity and includes things like *BRCA1* mutations. This list is reviewed frequently and currently has 59 genes registered (Kalia et al., 2017). Patients decide whether they want to be informed of such incidental findings and have the right to opt out on the informed consent form (Ayuso et al., 2015). These guidelines have been widely adopted and in schemes such as the 100,000 Genomes project, the actionable genes are automatically analysed and fed back to recruiting clinicians as secondary findings (http://www.genomicsengland.co.uk). The remaining mutations unrelated to the patient's primary phenotype are called incidental findings and these are not reported. However, the handling of secondary mutations has not been straight forward in clinical settings. Secondary findings need to be confirmed in a diagnostic laboratory to avoid the reporting of false-

positive results back to patients (Directors, 2012) and there is a significant cost associated with this. Furthermore, the problem of distinguishing between rare benign variants and pathogenic mutations and clinically managing the frequently encountered “variant of unknown significance” still remains (Bertier et al., 2016; Ormondroyd et al., 2018).

In this study, no specific steps were made to investigate the ACMG gene list to look for actionable mutations, but no obvious secondary findings were identified in the generation of the variant lists.

5.8 Future treatments and therapies

There is currently no cure for Coats disease or FEVR. Current treatments for Coats disease include cryotherapy, laser photocoagulation, subretinal fluid drainage with vitrectomy and anti-VEGF therapy (Cackett et al., 2010; Goel & Kumar, 2016; He et al., 2010; Imaizumi et al., 2016; Lin et al., 2010; Park et al., 2016; Shields et al., 2001a; Shields et al., 2001b; Sigler et al., 2014; Villegas et al., 2014). These treatments target secondary complications to prevent further progression of the disease (Section 1.5.3). Current treatments for FEVR are similar (Shaikh and Trese, 2002; Shukla et al., 2003; Margolis et al., 2004; Dickinson et al., 2006; Quiram et al., 2008; Sisk, 2012; Yamane et al., 2014; Henry et al., 2015; Dixon et al., 2016), although the use of anti-VEGF agents have not been as successful as those seen for Coats disease (Quiram et al., 2008; Tagami et al., 2008). New therapies are therefore needed for these disorders.

Gene replacement therapy has been successful for inherited retinal degeneration, with RPE65 treatment for Leber congenital amaurosis (LCA)

leading the way but treatments for retinoschisis, choroideremia and others are in the pipeline (Hardcastle et al., 2018). Similar strategies could be utilised for FEVR and although only small genes were amenable to this approach in the past, new dual AAV vectors are becoming available which fit large genes like *LRP5* (McClements and Maclaren, 2017).

The CRISPR-based gene editing technology is another potential therapeutic option for FEVR/Coats disease (Horvath et al., 2010; Hung et al., 2016) and studies in mice have successfully utilised this technology to delete mutations (Ruan et al., 2017), knock-down mutant alleles (Hung, et al., 2016) and edit genes in the retina (Arno et al., 2016). As the retinal vasculature develops before birth in humans, gene editing technologies would ideally need to be applied during embryogenesis and steps to use CRISPR/Cas9 in human embryos are currently being investigated (Liang et al., 2015) making this therapeutic avenue a future possibility.

Another potential strategy to treat FEVR/Coats disease is the use of iPSCs to generate vascular progenitor cells or endothelial cells from iPSCs to repair retinal telangiectasia. This was done recently in a study by Park et al. (2014) who injected iPSC-vascular progenitors into the vitreous of a retinal ischemia-reperfusion mouse model. The results showed the vascular progenitors were seen to integrate into the injured retinal vessels for up to 45 days. This shows a very promising outcome and opens a very exciting avenue in treating diseases with abnormal retinal vasculature.

The Norrin-β-catenin pathway is an attractive therapeutic target for FEVR, Coats and other vitreoretinopathies. Research is already ongoing focusing on targeting components of this pathway including therapeutic antibodies targeting TSPAN12 (Bucher et al., 2017) and treating ROP and Diabetic retinopathy (DR) with Norrin (Dailey et al., 2017). Research defining and characterising the components of this pathway, such as this study, underpins the development of therapies like these.

Another emerging therapy for FEVR, and possibly Coats disease, is early treatment with topical and intravitreal steroids at the first signs of capillary drop out (Thanos et al., 2016). Crucial to this treatment is the early identification of patients at risk and the molecular diagnosis of patients is key to this.

5.9 Limitations

One of the main limitations of this project was the lack of matched gDNA for all patients in the Coats cohort in Chapter 3. One of the aims of this study was to establish whether Coats disease is a somatic or germline disease, however, without the matched gDNA this was not possible to conclude. By obtaining patients' gDNA, it would allow for all the somatic and germline mutations to be identified and serve as candidate gene lists for Coats disease. Having these lists would further allow for effective process of identifying top candidate genes and variants, with the possibility of characterising Coats disease at the genomic level in the very near future.

Another significant limitation of the study presented in this thesis was the availability of only FFPE DNA. Due to the fixation and the age of the samples, this has caused major issues in DNA extraction and WES preparation (detailed in Sections 3.3.2 - 3.3.4) that could have potentially been overcome if another source of DNA was available. However, this was a limitation that was out of control and could not have been improved upon throughout the study.

The bioinformatics pipeline could also be considered as a further limitation. The pipeline used in this study, although very successful in identification of genes in other eye diseases from gDNA (El-Asrag et al., 2015; Panagiotou et al., 2017), was not designed to analyse FFPE samples and to control for the induced damage caused by the fixation process (Section 3.3.1.1). This limited the analysis in a number of ways and would need to be revisited and optimised for future analyses. Other filtering strategies mentioned in section 3.3.6 could be implemented into the new pipeline to help reduce the number of artefacts associated with FFPE DNA.

5.10 Future Work

To be able to progress with the research presented in this thesis, a number of experiments would need to be carried out. In Chapter 3, four out of seven Coats disease patients have now been located and their gDNA has been collected. Therefore, in the near future, the NGS analysis of their DNAs will hopefully aid in answering the question whether Coats disease is germline or somatic. Additionally, a further source of cryo-preserved Coats tissue (vitreoretinal membranes) was also collected along with matched gDNA and

parental DNA. WES or WGS (depending on the amount of DNA extracted) will be performed on these samples in the near future with the aim of comparing the WES lists generated from somatic tissue and identifying genes implicated in this disorder.

In Chapter 4, a new gene *DLG1* was identified in a patient with FEVR and was shown to play a role in retinal vasculature in *in vivo* mouse model by a collaborator. However, more mutations in other families need to be identified prior to the establishment of *DLG1* as a new FEVR gene. Therefore, in the near future, more FEVR cohorts will be screened for variants in this gene to strengthen the hypothesis that it is mutated in patients with abnormal vasculature and plays a role in FEVR.

Additionally, in Chapter 4, a novel mutation in the recently described transcript 2 of *LRP5* was identified in FEVR patient. The novel parts of this *LRP5* transcript will be screened in new and existing FEVR cohorts in the very near future to identify more variants and establish its role in FEVR.

5.11 Conclusion

The work presented in this thesis is a step towards better understanding of Coats disease and FEVR with the identification of new candidate genes for both diseases. The work also highlights the use of WES technology to effectively identify causative mutations or genes which could have been missed by other methods. With the identification of *LRP5* as a Coats disease gene, this opens up an extremely exciting time in understanding the pathology of this disease which has never been able to be investigated before. It is documented that only around 50% of known FEVR patients carry mutations in

published FEVR genes (Salvo et al., 2015), therefore, the identification of *DLG1* expands on this number and allows for a deeper understanding of this complex disorder. Altogether, the work presented in this thesis leads to improvements in diagnostics, management, treatments and future therapies for patients affected by Coats disease, FEVR or other diseases of abnormal retinal vasculature.

6 Chapter: References

Adzhubei, I.., Schmidt, S., Peshkin, L., Ramensky, V.., Gerasimova, A., Bork, P., Kondrashov, A.. and Sunyaev, S.. 2010. A method and server for predicting damaging missense mutations. *Nat Methods*. **7**(4), pp.248–249.

Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D., Brister, J.R., Bryant, S.H., Canese, K., Charowhas, C., Clark, K., Dicuccio, M., Dondoshansky, I., Federhen, S., Feolo, M., Funk, K., Geer, L.Y., Gorelenkov, V., Hoeppner, M., Holmes, B., Johnson, M., Khotomlianski, V., Kimchi, A., Kimelman, M., Kitts, P., Klimke, W., Krasnov, S., Kuznetsov, A., Landrum, M.J., Landsman, D., Lee, J.M., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Marchler-Bauer, A., Karsch-Mizrachi, I., Murphy, T., Orris, R., Ostell, J., O'sullivan, C., Panchenko, A., Phan, L., Preuss, D., Pruitt, K.D., Rodarmer, K., Rubinstein, W., Sayers, E., Schneider, V., Schuler, G.D., Sherry, S.T., Sirotkin, K., Siyan, K., Slotta, D., Soboleva, A., Soussov, V., Starchenko, G., Tatusova, T.A., Todorov, K., Trawick, B.W., Vakatov, D., Wang, Y., Ward, M., Wilbur, W.J., Yaschenko, E. and Zbicz, K. 2016. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*. **44**(D1), pp.D7–D19.

Ahn, S.J., Ryoo, N. and Woo, S.J. 2014. Ocular toxocariasis: clinical features, diagnosis, treatment, and prevention. *Asia Pacific allergy*. **4**(1), pp.134–141.

Ahn, V.E., Chu, M.L.H., Choi, H.J., Tran, D., Abo, A. and Weis, W.I. 2011. Structural basis of Wnt signaling inhibition by Dickkopf binding to LRP5/6. *Developmental Cell*. **21**(5), pp.862–873.

Aiello, L.., Avery, R.., Arrigg, P.., Keyt, B.A., Jampel, H.., Shah, S.T., Pasquale, L.R., Thieme, H., Iwamoto, M.., Park, J.E., Nguyen, H. V, Ferrara, N. and King, G.L. 1994. Vascular Endothelial Growth Factor in Ocular Fluid of Patients with Diabetic Retinopathy and Other Retinal Disorders. *The New England Journal of Medicine*. **331**(22), pp.1480–7.

Al-Zamil, W. and Yassin, S. 2017. Recent developments in age-related macular degeneration: a review. *Clinical Interventions in Aging*. **Volume 12**, pp.1313–1330.

Alexandridou, A. and Stavrou, P. 2002. Bilateral Coats' disease: long-term follow up. *Acta ophthalmologica Scandinavica*. **80**(1), pp.98–100.

van Allen, E.M., Wagle, N., Stojanov, P., Perrin, D.L., Cibulskis, K., Marlow, S., Jane-Valbuena, J., Friedrich, D.C., Kryukov, G., Carter, S.L., McKenna, A., Sivachenko, A., Rosenberg, M., Kiezun, A., Voet, D., Lawrence, M., Lichtenstein, L.T., Gentry, J.G., Huang, F.W., Fostel, J., Farlow, D., Barbie, D., Gandhi, L., Lander, E.S., Gray, S.W., Joffe, S., Janne, P., Garber, J., Macconaill, L., Lindeman, N., Rollins, B., Kantoff, P., Fisher, S.A., Gabriel, S., Getz, G. and Garraway, L.A. 2014. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nature Medicine*. **20**(6), pp.682–688.

Alonso, N., Soares, D.C., V McCloskey, E., Summers, G.D., Ralston, S.H. and

Gregson, C.L. 2015. Atypical Femoral Fracture in Osteoporosis Pseudoglioma Syndrome Associated with Two Novel Compound Heterozygous Mutations in *LRP5*. *Journal of Bone and Mineral Research*. **30**(4), pp.615–620.

Alpern, M., Bastian, B. and Moeller, J. 1982. In search of the elusive long-wave fundamental. *Vision Res.* **22**, pp.627–634.

Anastas, J.N. and Moon, R.T. 2013. WNT signalling pathways as therapeutic targets in cancer. *Nature Reviews Cancer*. **13**(1), pp.11–26.

Anderson, B. 1968. Ocular effects of changes in oxygen and carbon dioxide tension. *Transactions of the American Ophthalmological Society*. **66**, pp.423–74.

Arachchi, H., Wojcik, M.H., Weisburd, B., Jacobsen, J.O.B., Valkanas, E., Baxter, S., Byrne, A.B., O'Donnell-Luria, A.H., Haendel, M., Smedley, D., MacArthur, D.G., Philippakis, A.A. and Rehm, H.L. 2018. *matchbox* : An open-source tool for patient matching via the Matchmaker Exchange. *Human Mutation*., pp.1–19.

Arai, E., Fujimaki, T., Yanagawa, A., Fujiki, K., Yokoyama, T., Okumura, A., Shimizu, T. and Murakami, A. 2014. Familial cases of Norrie disease detected by copy number analysis. *Japanese Journal of Ophthalmology*. **58**(5), pp.448–454.

Arno, G., Agrawal, S.A., Eblimit, A., Bellingham, J., Xu, M., Wang, F., Chakarova, C., Parfitt, D.A., Lane, A., Burgoyne, T., Hull, S., Carss, K.J., Fiorentino, A., Hayes, M.J., Munro, P.M., Nicols, R., Pontikos, N., Holder, G.E., Black, G., Hall, G., Ingram, S., Gillespie, R., Manson, F., Sergouniotis, P., Inglehearn, C., Toomes, C., Ali, M., McKibbin, M., Poulter, J., Khan, K., Lord, E., Nemeth, A., Downes, S., Yu, J., Lise, S., Ponitkos, N., Plagnol, V., Michaelides, M., Hardcastle, A.J., Cheetham, M.E., Webster, A.R., van Heyningen, V., Asomugha, C., Raymond, F.L., Moore, A.T., Li, Y., Cukras, C. and Chen, R. 2016. Mutations in REEP6 Cause Autosomal-Recessive Retinitis Pigmentosa. *American Journal of Human Genetics*. **99**(6), pp.1305–1315.

Ayuso, C., Millan, J.M. and Dal-Re, R. 2015. Management and return of incidental genomic findings in clinical trials. *Pharmacogenomics Journal*. **15**(1), pp.1–5.

Azim, A.C., Knoll, J.H.M., Marfatia, S.M., Peel, D.J., Bryant, P.J. and Chishti, A.H. 1995. DLG1: Chromosome Location of the Closest Human Homologue of the Drosophila Discs Large Tumor Suppressor Gene. , pp.613–616.

Bahassi, E.M. and Stambrook, P.J. 2014. Next-generation sequencing technologies: Breaking the sound barrier of human genetics. *Mutagenesis*. **29**(5), pp.303–310.

Bass, S.J., Sherman, J. and Giovinazzo, V. 2011. Bilateral Coats' response in a female patient leads to diagnosis of facioscapulohumeral muscular dystrophy. *Optometry*. **82**(2), pp.72–76.

Bassett, E.A., Tokarew, N., Allemano, E.A., Mazerolle, C., Morin, K., Mears, A.J., McNeill, B., Ringuette, R., Campbell, C., Smiley, S., Pokrajac, N.T., Dubuc, A.M., Ramaswamy, V., Northcott, P.A., Remke, M., Monnier, P.P., Potter, D., Paes, K., Kirkpatrick, L.L., Coker, K.J., Rice, D.S., Perez-Iratxeta, C., Taylor, M.D. and Wallace, V.A. 2016. Norrin/frizzled4 signalling in the preneoplastic niche blocks medulloblastoma initiation.

eLife. **5**(NOVEMBER2016), pp.1–27.

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q.B., Antipenko, A., Shang, L., Boisson, B., Casanova, J.-L. and Abel, L. 2015. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proceedings of the National Academy of Sciences*. **112**(17), pp.5473–5478.

Berger, W., Van De Pol, D., Bächner, D., Oerlemans, F., Winkens, H., Hameister, H., Wieringa, B., Hendriks, W. and Ropers, H.H. 1996. An animal model for Norrie disease (ND): Gene targeting of the mouse ND gene. *Human Molecular Genetics*. **5**(1), pp.51–59.

Bertalovitz, A.C., Pau, M.S., Gao, S., Malbon, C.C. and Wang, H. 2016. Frizzled-4 C-terminus Distal to KTXXXW Motif is Essential for Normal Dishevelled Recruitment and Norrin-stimulated Activation of Lef/Tcf-dependent Transcriptional Activation. *Journal of Molecular Signaling*. **11**, pp.1–7.

Bertier, G., Hétu, M. and Joly, Y. 2016. Unsolved challenges of clinical whole-exome sequencing: A systematic literature review of end-users' views Donna Dickenson, Sandra Soo-Jin Lee, and Michael Morrison. *BMC Medical Genomics*. **9**(1), pp.1–12.

Betts, M.J. and Russell, R.B. 2003. Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists*., pp.289–316.

Bhat, S.A., Ahmad, S.M., Mumtaz, P.T., Malik, A.A., Dar, M.A., Urwat, U., Shah, R.A. and Ganai, N.A. 2016. Long non-coding RNAs: Mechanism of action and functional utility. *Non-coding RNA Research*. **1**(1), pp.43–50.

Black, G.C.M., Perveen, R., Bonshek, R., Cahill, M., Clayton-Smith, J., Lloyd, I.C. and McLeod, D. 1999. Coats' disease of the retina (unilateral retinal telangiectasis) caused by somatic mutation in the NDP gene: A role for norrin in retinal angiogenesis. *Human Molecular Genetics*. **8**(11), pp.2031–2035.

Blair, M.P., Ulrich, N., Hartnett, E. and Shapiro, M.J. 2013. Peripheral retinal nonperfusion in fellow eyes in coats disease. *Retina*. **33**(8), pp.1694–1699.

Bourhis, E., Tam, C., Franke, Y., Bazan, J.F., Ernst, J., Hwang, J., Costa, M., Cochran, A.G. and Hannoush, R.N. 2010. Reconstitution of a Frizzled8·Wnt3a·LRP6 signaling complex reveals multiple Wnt and Dkk1 binding sites on LRP6. *Journal of Biological Chemistry*. **285**(12), pp.9172–9179.

Boyden, L.., Mao, J., Belsky, J., Mitzner, L., Farhi, A., Mitnick, M., Wu, D., Insogna, K. and Lifton, R.P. 2002. High bone density due to a mutation in LDL-receptor-related protein 5. *Medicine*. **346**(20), pp.1462–1471.

Braun, T.A., Mullins, R.F., Wagner, A.H., Andorf, J.L., Johnston, R.M., Bakall, B.B., Deluca, A.P., Fishman, G.A., Lam, B.L., Weleber, R.G., Cideciyan, A. V., Jacobson, S.G., Sheffield, V.C., Tucker, B.A. and Stone, E.M. 2013. Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease. *Human Molecular Genetics*. **22**(25), pp.5136–5145.

Brézin, A.P., Nedelec, B., Barjol, A., Rothschild, P.-R., Delpech, M. and Valleix, S. 2011. A new VCAN/versican splice acceptor site mutation in a French Wagner family associated with vascular and inflammatory ocular features. *Molecular vision*. **17**(June), pp.1669–78.

Brown, N.L., Patel, S., Brzezinski, J. and Glaser, T. 2001. Math5 is required

for retinal ganglion cell and optic nerve formation. *Development*. **128**(13), pp.2497–2508.

Brown, N.L., Dagenais, S.L., Chen, C. and Glaser, T. 2002. Molecular characterization and mapping of ATOH7, a human atonal homolog with a predicted role in retinal ganglion cell development. . **13**(2), pp.95–101.

Brzezinski, J.A., Schulz, S.M., Crawford, S., Wroblewski, E., Brown, N.L. and Glaser, T. 2003. Math5 null mice have abnormal retinal and persistent hyaloid vasculatures. *In*: *Developmental Biology*. ACADEMIC PRESS INC ELSEVIER SCIENCE 525 B ST, STE 1900, SAN DIEGO, CA 92101-4495 USA, p.553.

Bucher, F., Zhang, D., Aguilar, E., Sakimoto, S., Diaz-Aguilar, S., Rosenfeld, M., Zha, Z., Zhang, H., Friedlander, M. and Yea, K. 2017. Antibody-Mediated Inhibition of Tspan12 Ameliorates Vasoproliferative Retinopathy Through Suppression of β-Catenin Signaling. *Circulation*. **136**(2), pp.180–195.

Cackett, P., Wong, D. and Cheung, C.M.G. 2010. Combined intravitreal bevacizumab and argon laser treatment for Coats' disease. *Acta Ophthalmologica*. **88**(2), pp.48–49.

Canny, C.L.B. and Oliver, G.L. 1976. Fluorescein Angiographic Findings in Familial Exudative Vitreoretinopathy. *Archives of Ophthalmology*. **94**(7), pp.1114–1120.

Carrigan, M., Duignan, E., Malone, C.P.G., Stephenson, K., Saad, T., McDermott, C., Green, A., Keegan, D., Humphries, P., Kenna, P.F. and Farrar, G.J. 2016. Panel-Based population next-generation sequencing for inherited retinal degenerations. *Scientific Reports*. **6**(September), pp.1–9.

Carroll, L.S., Williams, H.J., Walters, J., Kirov, G., O'Donovan, M.C. and Owen, M.J. 2011. Mutation screening of the 3q29 microdeletion syndrome candidate genes DLG1 and PAK2 in schizophrenia. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*. **156**(7), pp.844–849.

Chaitankar, V., Karakulah, G., Ratnapriya, R., Giuste, F.O., Brooks, M.J. and Swaroop, A. 2016. *Next Generation Sequencing Technology and Genomewide Data Analysis: Perspectives for Retinal Research*.

Chang, T.H., Hsieh, F.L., Zebisch, M., Harlos, K., Elegheert, J. and Jones, E.Y. 2015. Structure and functional properties of norrin mimic wnt for signalling with Frizzled4, Lrp5/6, and proteoglycan. *eLife*. **4**, pp.1–27.

Chen, J., Yan, H., Ren, D. ni, Yin, Y., Li, Z., He, Q., Wo, D., Ho, M.S. chun, Chen, Y., Liu, Z., Yang, J., Liu, S. and Zhu, W. 2014. LRP6 dimerization through its LDLR domain is required for robust canonical Wnt pathway activation. *Cellular Signalling*. **26**(5), pp.1068–1074.

Chen, L.L. and Carmichael, G.G. 2010. Decoding the function of nuclear long non-coding RNAs. *Current Opinion in Cell Biology*. **22**(3), pp.357–364.

Chen, Z. Y., Battinelli, E.M., Fielder, A., Bundey, S., Sims, K., Breakefield, X.O., Craig, I.W. 1993a. A mutation in the Norrie disease gene (NDP) associated with X-linked familial exudative vitreoretinopathy. *Nature Genetics*. **3**, pp.73–96.

Chen, Z.Y., Battinelli, E.M., Hendriks, R.W., Powell, J.F., Middleton-Price, H., Sims, K.B., Breakefield, X.O. and Craig, I.W. 1993b. Norrie Disease Gene: Characterization of deletions and possible function. *Genomics*. **16**,

pp.533–535.
Cho, D., Mier, J.W. and Atkins, M.B. 2009. PI3K/Akt/mTOR pathway: a growth and proliferation pathway *In*: *Renal Cell Carcinoma*., pp.267–285.
Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. and Chan, A.P. 2012. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS ONE*. **7**(10).
Cipriani, V., Silva, R.S., Arno, G., Pontikos, N., Kalhoro, A., Valeina, S., Inashkina, I., Audere, M., Rutka, K., Puech, B., Michaelides, M., Van Heyningen, V., Lace, B., Webster, A.R. and Moore, A.T. 2017. Duplication events downstream of IRX1 cause North Carolina macular dystrophy at the MCDR3 locus. *Scientific Reports*. **7**(1), pp.1–9.
Clark, M.M., Stark, Z., Farnaes, L., Tan, T.Y., White, S.M., Dimmock, D. and Kingsmore, S.F. 2018. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *npj Genomic Medicine*. **3**(1), pp.1–10.
Cnossen, W.R., te Morsche, R.H.M., Hoischen, A., Gilissen, C., Chrispijn, M., Venselaar, H., Mehdi, S., Bergmann, C., Veltman, J.A. and Drenth, J.P.H. 2014. Whole-exome sequencing reveals LRP5 mutations and canonical Wnt signaling associated with hepatic cystogenesis. *Proceedings of the National Academy of Sciences*. **111**(14), pp.5343–5348.
Coats, G. 1908. Forms of retinal diseases with massive exudation. *Royal London Ophthalmic Hospital Reports*. **17**, p.440–525.
Collin, R.W.J., Nikopoulos, K., Dona, M., Gilissen, C., Hoischen, A., Boonstra, F.N., Poulter, J.A., Kondo, H., Berger, W., Toomes, C., Tahira, T., Mohn, L.R., Blokland, E.A., Hetterschijt, L., Ali, M., Groothuismink, J.M., Duijkers, L., Inglehearn, C.F., Sollfrank, L., Strom, T.M., Uchio, E., van Nouhuys, C.E., Kremer, H., Veltman, J.A., van Wijk, E. and Cremers, F.P.M. 2013. ZNF408 is mutated in familial exudative vitreoretinopathy and is crucial for the development of zebrafish retinal vasculature. *Proceedings of the National Academy of Sciences*. **110**(24), pp.9856–9861.
Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P., Fitzgerald, T., Hu, M., Ihm, C.H., Kristiansson, K., MacArthur, D.G., MacDonald, J.R., Onyiah, I., Pang, A.W.C., Robson, S., Stirrups, K., Valsesia, A., Walter, K., Wei, J., Tyler-Smith, C., Carter, N.P., Lee, C., Scherer, S.W. and Hurles, M.E. 2010. Origins and functional impact of copy number variation in the human genome. *Nature*. **464**(7289), pp.704–712.
Conte, I., Hadfield, K.D., Barbato, S., Carrella, S., Pizzo, M., Bhat, R.S., Carissimo, A., Karali, M., Porter, L.F., Urquhart, J., Hateley, S., O'Sullivan, J., Manson, F.D.C., Neuhauss, S.C.F., Banfi, S. and Black, G.C.M. 2015. MiR-204 is responsible for inherited retinal dystrophy associated with ocular coloboma. *Proceedings of the National Academy of Sciences*. **112**(25), pp.E3236–E3245.
Coppieters, F., Todeschini, A.L., Fujimaki, T., Baert, A., De Bruyne, M., Van Cauwenbergh, C., Verdin, H., Bauwens, M., Ongenaert, M., Kondo, M., Meire, F., Murakami, A., Veitia, R.A., Leroy, B.P. and De Baere, E. 2015. Hidden Genetic Variation in LCA9-Associated Congenital Blindness Explained by 5′UTR Mutations and Copy-Number Variations of NMNAT1.

Human Mutation. **36**(12), pp.1188–1196.

Coppieters, F., Ascari, G., Dannhausen, K., Nikopoulos, K., Peelman, F., Karlstetter, M., Xu, M., Brachet, C., Meunier, I., Tsilimbaris, M.K., Tsika, C., Blazaki, S. V., Vergult, S., Farinelli, P., Van Laethem, T., Bauwens, M., De Bruyne, M., Chen, R., Langmann, T., Sui, R., Meire, F., Rivolta, C., Hamel, C.P., Leroy, B.P. and De Baere, E. 2016. Isolated and Syndromic Retinal Dystrophy Caused by Biallelic Mutations in RCBTB1, a Gene Implicated in Ubiquitination. *American Journal of Human Genetics*. **99**(2), pp.470–480.

Costa, R.M.A., Chiganças, V., Galhardo, R.D.S., Carvalho, H. and Menck, C.F.M. 2003. The eukaryotic nucleotide excision repair pathway. *Biochimie*. **85**(11), pp.1083–1099.

Cotter, L., Ozçelik, M., Jacob, C., Pereira, J.A., Locher, V., Baumann, R., Relvas, J.B., Suter, U. and Tricaud, N. 2010. Dlg1-PTEN interaction regulates myelin thickness to prevent damaging peripheral nerve overmyelination. *Science*. **328**(5984), pp.1415–1418.

Crisan, D. and Mattson, J.C. 1993. Retrospective DNA analysis using fixed tissue specimens. *DNA and Cell Biology*. **12**(5), pp.455–464.

Criswick, V.G. and Schepens, C.L. 1969. Familial Exudative Vitreoretinopathy. *American Journal of Ophthalmology*. **68**(4), pp.578–594.

Dailey, W.A., Drenser, K.A., Wong, S.C., Cheng, M., Vercellone, J., Roumayah, K.K., Feeney, E. V., Deshpande, M., Guzman, A.E., Trese, M. and Mitton, K.P. 2017. Norrin treatment improves ganglion cell survival in an oxygen-induced retinopathy model of retinal ischemia. *Experimental Eye Research*. **164**, pp.129–138.

Daneman, R., Agalliu, D., Zhou, L., Kuhnert, F., Kuo, C.J. and Barres, B.A. 2009. Wnt/beta-catenin signaling is required for CNS, but not non-CNS, angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. **106**, pp.641–646.

Davis, C.G., Goldstein, J.L., Südhof, T.C., Anderson, R.G.W., Russell, D.W. and Brown, M.S. 1987. Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. *Nature*. **326**(6115), pp.760–765.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., McKenna, A., Fennell, T.J., Kernytsky, A.M., Sivachenko, A.Y., Cibulskis, K., Gabriel, S.B., Altshuler, D. and Daly, M.J. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. **43**(5), pp.491–498.

Dickinson, J.L., Sale, M.M., Passmore, A., Fitzgerald, L.M., Wheatley, C.M., Burdon, K.P., Craig, J.E., Tengtrisorn, S., Carden, S.M., Maclean, H. and Mackey, D.A. 2006. Mutations in the NDP gene: Contribution to Norrie disease, familial exudative vitreoretinopathy and retinopathy of prematurity. *Clinical and Experimental Ophthalmology*. **34**(7), pp.682–688.

Dietrich, D., Uhl, B., Sailer, V., Holmes, E.E., Jung, M. and Meller, S. 2013. Improved PCR Performance Using Template DNA from Formalin-Fixed and Paraffin-Embedded Tissues by Overcoming PCR Inhibition. . **8**(10), pp.1–10.

Directors, A.B. of 2012. Points to consider in the clinical application of genomic

sequencing. *Genet Med*. **14**(8), pp.759–761.

Dixon, M.W., Stem, M.S., Schuette, J.L., Keegan, C.E. and Besirli, C.G. 2016. CTNNB1 mutation associated with familial exudative vitreoretinopathy (FEVR) phenotype. *Ophthalmic Genetics*. **37**(4), pp.468–470.

Do, H. and Dobrovic, A. 2012. Dramatic reduction of sequence artefacts from DNA isolated from formalin-fixed cancer biopsies by treatment with uracil-DNA glycosylase. *Oncotarget*. **3**(5), pp.546–58.

Do, H. and Dobrovic, A. 2015. Sequence artifacts in DNA from formalin-fixed tissues: Causes and strategies for minimization. *Clinical Chemistry*. **61**(1), pp.64–71.

Dorrell, M.I., Friedlander, M. and Smith, L.E.H. 2007. Retinal Vascular Development. In: Joussen A.M., Gardner T.W., Kirchhof B., Ryan S.J. (eds). *Retinal Vascular Disease. Springer, Berlin, Heidelberg*.

Downey, L.M., Bottomley, H.M., Sheridan, E., Ahmed, M., Gilmour, D.F., Inglehearn, C.F., Reddy, a, Agrawal, a, Bradbury, J. and Toomes, C. 2006. Reduced bone mineral density and hyaloid vasculature remnants in a consanguineous recessive FEVR family with a mutation in LRP5. *The British journal of ophthalmology*. **90**(9), pp.1163–7.

Drenser, K.A. 2016. Wnt signaling pathway in retinal vascularization. *Eye and Brain*. **Volume 8**, pp.141–146.

Dubruc, E., Putoux, A., Labalme, A., Rougeot, C., Sanlaville, D. and Edery, P. 2014. A new intellectual disability syndrome caused by CTNNB1 haploinsufficiency. *American Journal of Medical Genetics, Part A*. **164**(6), pp.1571–1575.

Edward, D.P., Mafee, M.F., Garcia-Valenzuela, E. and Weiss, R.A. 1998. Coats' disease and persistent hyperplastic primary vitreous. Role of MR imaging and CT. *Radiologic clinics of North America*. **36**(6), p.1119–31, x.

Egbert, P.R., Chan, C.C. and Winter, F.C. 1976. Flat preparations of the retinal vessels in Coats' disease. *Journal of pediatric ophthalmology*. **13**(6), pp.336–339.

Eisenberger, T., Neuhaus, C., Khan, A.O., Decker, C., Preising, M.N., Friedburg, C., Bieg, A., Gliem, M., Issa, P.C., Holz, F.G., Baig, S.M., Hellenbroich, Y., Galvez, A., Platzer, K., Wollnik, B., Laddach, N., Ghaffari, S.R., Rafati, M., Botzenhart, E., Tinschert, S., Börger, D., Bohring, A., Schreml, J., Körtge-Jung, S., Schell-Apacik, C., Bakur, K., Al-Aama, J.Y., Neuhann, T., Herkenrath, P., Nürnberg, G., Nürnberg, P., Davis, J.S., Gal, A., Bergmann, C., Lorenz, B. and Bolz, H.J. 2013. Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: The example of retinal dystrophies. *PLoS ONE*. **8**(11).

El-Asrag, M.E., Sergouniotis, P.I., McKibbin, M., Plagnol, V., Sheridan, E., Waseem, N., Abdelhamed, Z., McKeefry, D., Van Schil, K., Poulter, J.A., Black, G., Hall, G., Ingram, S., Gillespie, R., Ramsden, S., Manson, F., Hardcastle, A., Michaelides, M., Cheetham, M., Arno, G., Thomas, N., Bhattacharya, S., Moore, T., Nemeth, A., Downes, S., Lise, S., Lord, E., Johnson, C.A., Carr, I.M., Leroy, B.P., De Baere, E., Inglehearn, C.F., Webster, A.R., Toomes, C., Ali, M., Watson, C.M., El-Asrag, M.E., Parry, D.A., Morgan, J.E., Logan, C. V., Carr, I.M., Sheridan, E., Charlton, R., Johnson, C.A., Taylor, G., Toomes, C., McKibbin, M., Inglehearn, C.F.

and Ali, M. 2015. Biallelic Mutations in the Autophagy Regulator DRAM2 Cause Retinal Dystrophy with Early Macular Involvement. *American Journal of Human Genetics*. **96**(6), pp.948–954.

Ellingford, J.M., Barton, S., Bhaskar, S., Williams, S.G., Sergouniotis, P.I., O'Sullivan, J., Lamb, J.A., Perveen, R., Hall, G., Newman, W.G., Bishop, P.N., Roberts, S.A., Leach, R., Tearle, R., Bayliss, S., Ramsden, S.C., Nemeth, A.H. and Black, G.C.M. 2016. Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology*. **123**(5), pp.1143–1150.

Ellingford, J.M., Campbell, C., Barton, S., Bhaskar, S., Gupta, S., Taylor, R.L., Sergouniotis, P.I., Horn, B., Lamb, J.A., Michaelides, M., Webster, A.R., Newman, W.G., Panda, B., Ramsden, S.C. and Black, G.C.M. 2017. Validation of copy number variation analysis for next-generation sequencing diagnostics. *European Journal of Human Genetics*. **25**(6), pp.719–724.

Ellingford, J.M., Horn, B., Campbell, C., Arno, G., Barton, S., Tate, C., Bhaskar, S., Sergouniotis, P.I., Taylor, R.L., Carss, K.J., Raymond, L.F.L., Michaelides, M., Ramsden, S.C., Webster, A.R. and Black, G.C.M. 2018a. Assessment of the incorporation of CNV surveillance into gene panel next-generation sequencing testing for inherited retinal diseases. *Journal of Medical Genetics*. **55**(2), pp.114–121.

Ellingford, J.M., Beaman, G., Webb, K., O'Callaghan, C., Hirst, R.A., The 100, 000 Genomes Project, Black, G.C. and Newman, W.G. 2018b. Whole genome sequencing enables definitive diagnosis of Cystic Fibrosis and Primary Ciliary Dyskinesia. *bioRxiv*., p.438–838.

Engelhardt, B. and Liebner, S. 2014. Novel insights into the development and maintenance of the blood-brain barrier. *Cell and Tissue Research*. **355**(3), pp.687–699.

Ergun, S.G., Akay, G.G., Ergun, M.A. and Perçin, E.F. 2017. LRP5-linked osteoporosis-pseudoglioma syndrome mimicking isolated microphthalmia. *European Journal of Medical Genetics*. **60**(3), pp.200–204.

Eshaq, R.S., Aldalati, A.M.Z., Alexander, J.S. and Harris, N.R. 2017. Diabetic retinopathy: Breaking the barrier. *Pathophysiology*. **24**(4), pp.229–241.

Fei, P., Zhang, Q., Huang, L., Xu, Y., Zhu, X., Tai, Z., Gong, B., Ma, S., Yao, Q., Li, J., Zhao, P. and Yang, Z. 2014. Identification of two novel LRP5 mutations in families with familial exudative vitreoretinopathy. *Molecular vision*. **20**(July 2013), pp.395–409.

Fernandes, B.F., Odashiro, A.N., Maloney, S., Zajdenweber, M.E., Lopes, A.G. and Burnier, M.N. 2006. Clinical-histopathological correlation in a case of Coats' disease. *Diagnostic pathology*. **1**, p.24.

Firth, H. V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Vooren, S. Van, Moreau, Y., Pettett, R.M. and Carter, N.P. 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American Journal of Human Genetics*. **84**(4), pp.524–533.

Flynn-Evans, E.E., Tabandeh, H., Skene, D.J. and Lockley, S.W. 2014. Circadian rhythm disorders and melatonin production in 127 blind women with and without light perception. *Journal of Biological Rhythms*. **29**(3),

pp.215–224.
Fokkema, I.F.A.C., Taschner, P.E.M., Schaafsma, G.C.P., Celli, J., Laros, J.F.J. and den Dunnen, J.T. 2011. LOVD v.2.0: The next generation in gene variant databases. *Human Mutation*. **32**(5), pp.557–563.
Fox, C.H., Johnson, F.B., Whiting, J. and Roller, P.P. 1985. Formaldehyde Foxation. . **33**(8), pp.845–853.
Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gústafsdóttir, S.M., Östman, A. and Landegren, U. 2002. Protein detection using proximity-dependent DNA ligation assays. *Nature Biotechnology*. **20**(5), pp.473–477.
French, D. and Edsall, J.T. 1945. The reactions of formaldehyde with amino acids and proteins. *Advances in Protein Chemistry*. **2**, pp.1–335.
Friedman, J.S., Ray, J.W., Waseem, N., Johnson, K., Brooks, M.J., Hugosson, T., Breuer, D., Branham, K.E., Krauth, D.S., Bowne, S.J., Sullivan, L.S., Ponjavic, V., Gränse, L., Khanna, R., Trager, E.H., Gieser, L.M., Hughbanks-Wheaton, D., Cojocaru, R.I., Ghiasvand, N.M., Chakarova, C.F., Abrahamson, M., Göring, H.H.H., Webster, A.R., Birch, D.G., Abecasis, G.R., Fann, Y., Bhattacharya, S.S., Daiger, S.P., Heckenlively, J.R., Andréasson, S. and Swaroop, A. 2009. Mutations in a BTB-Kelch Protein, KLHL7, Cause Autosomal-Dominant Retinitis Pigmentosa. *American Journal of Human Genetics*. **84**(6), pp.792–800.
Fruttiger, M. 2007. Development of the retinal vasculature. *Angiogenesis*. **10**(2), pp.77–88.
Fruttiger, M., Calver, A.R. and Richardson, W.D. 2000. Platelet-derived growth factor is constitutively secreted from neuronal cell bodies but not from axons. *Current Biology*. **10**(20), pp.1283–1286.
Fu, X., Sun, H., Klein, W.H. and Mua, X. 2006. β-catenin is essential for lamination but not neurogenesis in mouse retinal development. *Developmental Biology*. **299**(2), pp.424–437.
Fuhrmann, S. 2010. Eye Morphogenesis and Patterning of the Optic Vesicle. *October*. (93), pp.61–84.
Fukatsu, T. 1999. Acetone preservation: A practical technique for molecular analysis. *Molecular Ecology*. **8**(11), pp.1935–1945.
Gass, J.D.M. 1968a. A Fluorescein Angiographic Study of Macula. *Archives of Ophthalmology*. **80**(5), pp.569–582.
Gass, J.D.M. 1968b. A Fluorescein Angiographic Study of Macular Dysfunction Secondary to Retinal Vascular Disease. *Archives of Ophthalmology*. **80**(5), pp.606–617.
Gass, J.D.M. and Blodi, B.A. 1993. Idiopathic Juxtafoveolar Retinal Telangiectasis: Update of Classification and Follow-up Study. *Ophthalmology*. **100**(10), pp.1536–1546.
Gillio-Tos, A., De Marco, L., Fiano, V., Garcia-Bragado, F., Dikshit, R., Boffetta, P. and Merletti, F. 2007. Efficient DNA extraction from 25-year-old paraffin-embedded tissues: Study of 365 samples. *Pathology*. **39**(3), pp.345–348.
Goel, N. and Kumar, V. 2016. Intravitreal Bevacizumab as an adjunct to laser in the management of adult onset Coats' disease. *Graefe's Archive for Clinical and Experimental Ophthalmology*. **254**(9), pp.1861–1862.
Gong, Y., Slee, R.B., Fukai, N., Rawadi, G., Roman-Roman, S., Reginato, A.M., Wang, H., Cundy, T., Glorieux, F.H., Lev, D., Zacharin, M., Oexle,

K., Marcelino, J., Suwairi, W., Heeger, S., Sabatakos, G., Apte, S., Adkins, W.N., Allgrove, J., Arslan-Kirchner, M., Batch, J.A., Beighton, P., Black, G.C.M., Boles, R.G., Boon, L.M., Borrone, C., Brunner, H.G., Carle, G.F., Dallapiccola, B., De Paepe, A., Floege, B., Halfhide, M.L., Hall, B., Hennekam, R.C., Hirose, T., Jans, A., Jüppner, H., Kim, C.A., Keppler-Noreuil, K., Kohlschuetter, A., LaCombe, D., Lambert, M., Lemyre, E., Letteboer, T., Peltonen, L., Ramesar, R.S., Romanengo, M., Somer, H., Steichen-Gersdorf, E., Steinmann, B., Sullivan, B., Superti-Furga, A., Swoboda, W., Van den Boogaard, M.J., Van Hul, W., Vikkula, M., Votruba, M., Zabel, B., Garcia, T., Baron, R., Olsen, B.R. and Warman, M.L. 2001. LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell*. **107**(4), pp.513–523.

Graw, J. 2003. The genetic and molecular basis of congenital eye defects. *Nature Reviews Genetics*. **4**(11), p.876.

Graw, J. 2010. Eye development. *Current Topics in Developmental Biology*. **90**, pp.343–386.

Green, R.C., Berg, J.S., Grody, W.W., Kalia, S., Korf, B.R., Martin, C.L., Mcguire, A., Nussbaum, R.L., Daniel, J.M.O., Ormond, K.E., Rehm, H.L., Watson, M.S., Marc, S. and Biesecker, L.G. 2013. American College of Medical Genetics and Genomics ACMG Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing. *American College of Medical Genetics*. **4472**(7), pp.1–29.

Gursoy, H., Erol, N., Bilgec, M.D., Basmak, H., Kutlay, O. and Aslan, H. 2015. Bilateral Coats' Disease Combined with Retinopathy of Prematurity. *Case Reports in Ophthalmological Medicine*. **2015**, pp.1–4.

Haddad, R., Font, R.L. and Reeser, F. 1978. Persistent hyperplastic primary vitreous. A clinicopathologic study of 62 cases and review of the literature. *Review Literature And Arts Of The Americas*. **23**(2), pp.123–134.

Haik, B.G. 1991. Advanced Coats' disease. *Transactions of the American Ophthalmological Society*. **89**, pp.371–476.

Halpin, C. and Sims, K. 2008. Twenty years of audiology in a patient with Norrie disease. *International Journal of Pediatric Otorhinolaryngology*. **72**(11), pp.1705–1710.

Hardcastle, A.J., Sieving, P.A., Sahel, J.-A., Jacobson, S.G., Cideciyan, A. V., Flannery, J.G., Beltran, W.A. and Aguirre, G.D. 2018. Translational Retinal Research and Therapies. *Translational Vision Science & Technology*. **7**(5), p.8.

Harris, G., O'Toole, S., George, P., Browett, P. and Print, C. 2017. Massive parallel sequencing of solid tumours – challenges and opportunities for pathologists. *Histopathology*. **70**(1), pp.123–133.

Hartikka, H., Mäkitie, O., Männikkö, M., Doria, A.S., Daneman, A., Cole, W.G., Ala-Kokko, L. and Sochett, E.B. 2005. Heterozygous mutations in the LDL receptor-related protein 5 (LRP5) gene are associated with primary osteoporosis in children. *Journal of Bone and Mineral Research*. **20**(5), pp.783–789.

Hazan, F., Ostergaard, P., Ozturk, T., Kantekin, E., Atlihan, F., Jeffery, S. and Ozkinay, F. 2012. A novel KIF11 mutation in a Turkish patient with microcephaly, lymphedema, and chorioretinal dysplasia from a consanguineous family. *American Journal of Medical Genetics, Part A*. **158 A**(7), pp.1686–1689.

He, X., Semenov, M., Tamai, K. and Zeng, X. 2004. LDL receptor-related proteins 5 and 6 in Wnt/beta-catenin signaling: arrows point the way. *Development*. **131**(8), pp.1663–77.

He, Y.-G., Wang, H., Zhao, B., Lee, J., Bahl, D. and McCluskey, J. 2010. Elevated vascular endothelial growth factor level in Coats' disease and possible therapeutic role of bevacizumab. *Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv für klinische und experimentelle Ophthalmologie*. **248**(10), pp.1519–21.

Hedegaard, J., Thorsen, K., Lund, M.K., Hein, A.M.K., Hamilton-Dutoit, S.J., Vang, S., Nordentoft, I., Birkenkamp-Demtröder, K., Kruhøffer, M., Hager, H., Knudsen, B., Andersen, C.L., Sørensen, K.D., Pedersen, J.S., Ørntoft, T.F. and Dyrskjøt, L. 2014. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. *PLoS ONE*. **9**(5).

Hejtmancik, J.F., Jiao, X., Li, A., Sergeev, Y. V, Ding, X., Sharma, A.K., Chan, C.-C., Medina, I. and Edwards, A.O. 2008. Mutations in KCNJ13 Cause Autosomal-Dominant Snowflake Vitreoretinal Degeneration. *The American Journal of Human Genetics*. **82**, pp.174–180.

Henikoff, S. and Henikoff, J.G. 1992. Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*. **89**(22), pp.10915–10919.

Henry, C.R., Sisk, R.A., Tzu, J.H., Albini, T.A., Davis, J.L., Murray, T.G. and Berrocal, A.M. 2015. Long-term follow-up of intravitreal bevacizumab for the treatment of pediatric retinal and choroidal diseases. *Journal of AAPOS*. **19**(6), pp.541–548.

Hey, P.J., Twells, R.C.J., Phillips, M.S., Nakagawa, Y., Brown, S.D., Kawaguchi, Y., Cox, R., Xie, G., Dugan, V., Hammond, H., Metzker, M.L., Todd, J.A. and Hess, F.J. 1998. Cloning of a novel member of the low-density lipoprotein receptor family. *Gene*. **216**, pp.103–111.

Heydt, C., Fassunke, J., Künstlinger, H., Ihle, M.A., König, K., Heukamp, L.C., Schildhaus, H.U., Odenthal, M., Büttner, R. and Merkelbach-Bruse, S. 2014. Comparison of pre-analytical FFPE sample preparation methods and their impact on massively parallel sequencing in routine diagnostics. *PLoS ONE*. **9**(8).

den Hollander, A.I., Davis, J., Van Der Velde-Visser, S.D., Zonneveld, M.N., Pierrottet, C.O., Koenekoop, R.K., Kellner, U., Van Den Born, L.I., Heckenlively, J.R., Hoyng, C.B., Handford, P.A., Roepman, R. and Cremers, F.P.M. 2004. CRB1 mutation spectrum in inherited retinal dystrophies. *Human Mutation*. **24**(5), pp.355–369.

den Hollander, A.I., Koenekoop, R.K., Yzer, S., Lopez, I., Arends, M.L., Voesenek, K.E.J., Zonneveld, M.N., Strom, T.M., Meitinger, T., Brunner, H.G., Hoyng, C.B., van den Born, L.I., Rohrschneider, K. and Cremers, F.P.M. 2006. Mutations in the CEP290 (NPHP6) Gene Are a Frequent Cause of Leber Congenital Amaurosis. *The American Journal of Human Genetics*. **79**(3), pp.556–561.

Hoon, M., Okawa, H., Della Santina, L. and Wong, R.O.L. 2014. Functional architecture of the retina: Development and disease. *Progress in Retinal and Eye Research*. **42**(i), pp.44–84.

Horvath, P., Barrangou, R., Horvath1, P. and Barrangou2, R. 2010. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Source:*

Science, New Series. **327**(5962), pp.167–170.

Hsieh, J.-C., Lee, L., Zhang, L., Wefer, S., Brown, K., DeRossi, C., Wines, M.E., Rosenquist, T., Holdener, B.C., Ang, S.L., Conlon, R.A., Jin, O., Rossant, J., Beddington, R.S.P., Beddington, R.S., Robertson, E.J., Brennan, J., Lu, C.C., Norris, D.P., Rodriguez, T.A., Beddington, R.S.P., Robertson, E., Bu, G., Rennke, S., Collingnon, J., Varlet, I., Robertson, E.J., Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., Robertson, E.J., Culi, J., Mann, R.S., Ding, J., Yang, L., Yan, Y.T., Chen, A., Desai, N., Wynshaw-Boris, A., Shen, M.M., Emanuelsson, O., Nielsen, H., Brunak, S., Heijne, G. von, Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., Pease, L.R., Hobbs, H.H., Brown, M.S., Goldstein, J.L., Hodges, R., Heaton, R., Parker, J., Molday, L., Molday, R., Holdener, B.C., Faust, C., Rosenthal, N., Magnuson, T., Holdener-Kenny, B., Sharan, S.K., Magnuson, T., Jorgensen, M.M., Jensen, O.N., Henrik, U.H., Hansen, J.-J., Corydon, T.J., Bross, P., Bolund, L., Gregersen, N., Lawson, K.A., Pedersen, R.A., Liu, P., Wakamiya, M., Shea, M.J., Albrecht, U., Behringer, R.R., Bradley, A., Lu, C.C., Brennan, J., Robertson, E.J., Manganas, L.N., Trimmer, J.S., Morris, S.M., Cooper, J.A., Morris, S.M., Tallquist, M.D., Rock, C.O., Cooper, J.A., Munro, S., Pelham, H.R.B., Nakai, K., Horton, P., Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J., Skarnes, W.C., Schedl, A., Ruppert, S., Kelsey, G., Thies, E., Niswander, L., Magnuson, T., Klebig, M.L., Rinchik, E.M., Schutz, G., Shumacher, A., Faust, C., Magnuson, T., Strickland, D.K., Gonias, S.L., Argraves, W.S., Tam, P.P., Behringer, R.R., Tam, P.P., Steiner, K.A., Tam, P.P., Gad, J.M., Kinder, S.J., Tsang, T.E., Behringer, R.R., Thomas, P., Beddington, R., Varlet, I., Collignon, J., Robertson, E.J., Waldrip, W.R., Bikoff, E.K., Hoodless, P.A., Wrana, J.L., Robertson, E.J., Weinstein, M., Yang, X., Li, C., Xu, X., Gotay, J., Deng, C.X., Willnow, T.E., Fohlmann, A., Horton, J., Otani, H., Braun, J.R., Hammer, R.E., Herz, J., Wines, M.E., Lee, L., Katari, M.S., Zhang, L., DeRossi, C., Shi, Y., Perkins, S., Feldman, M., McCombie, W.R., Holdener, B.C., Yang, D.-H., Smith, E.R., Roland, I.H., Sheng, Z., He, J., Martin, W.D., Hamilton, T.C., Lambeth, J.D. and Xu, X.-X. 2003. Mesd encodes an LRP5/6 chaperone essential for specification of mouse embryonic polarity. *Cell*. **112**(3), pp.355–67.

Hsieh, M., Boerboom, D., Shimada, M., Lo, Y., Parlow, A.F., Luhmann, U.F.O., Berger, W. and Richards, J.S. 2005. Mice Null for Frizzled4 (Fzd4−/−) Are Infertile and Exhibit Impaired Corpora Lutea Formation and Function1. *Biology of Reproduction*. **73**(6), pp.1135–1146.

Hu, H., Xiao, X., Li, S., Jia, X., Guo, X. and Zhang, Q. 2016. KIF11 mutations are a common cause of autosomal dominant familial exudative vitreoretinopathy. *British Journal of Ophthalmology*. **100**(2), pp.278–283.

Huang, S., Rutar, T., Bloomer, M. and Crawford, J. 2010. Analysis of Clinical Misdiagnoses in Children Treated With Enucleation. . **128**(8), pp.1009–1013.

Huang, X.-Y., Zhuang, H., Wu, J.-H., Li, J.-K., Hu, F.-Y., Zheng, Y., Tellier, L.C.A.M., Zhang, S.-H., Gao, F.-J., Zhang, J.-G. and Xu, G.-Z. 2017. Targeted next-generation sequencing analysis identifies novel mutations in families with severe familial exudative vitreoretinopathy. *Molecular vision*. **23**(February), pp.605–613.

Hughes, S., Yang, H. and Chan-Ling, T. 2000. Vascularization of the human fetal retina: Roles of vasculogenesis and angiogenesis. *Investigative Ophthalmology and Visual Science*. **41**(5), pp.1217–1228.

Humbert, P., Russell, S. and Richardson, H. 2003. Dlg, scribble and Lgl in cell polarity, cell proliferation and cancer. *BioEssays*. **25**(6), pp.542–553.

Hung, S.S.C., Chrysostomou, V., Li, F., Lim, J.K.H., Wang, J.H., Powell, J.E., Tu, L., Daniszewski, M., Lo, C., Wong, R.C., Crowston, J.G., Pebay, A., King, A.E., Bui, B. V., Liu, G.S. and Hewitt, A.W. 2016a. AAV-Mediated CRISPR/Cas Gene Editing of Retinal Cells in Vivo. *Investigative Ophthalmology and Visual Science*. **57**(7), pp.3470–3476.

Hung, S.S.C., McCaughey, T., Swann, O., Pébay, A. and Hewitt, A.W. 2016b. Genome engineering in ophthalmology: Application of CRISPR/Cas to the treatment of eye disease. *Progress in Retinal and Eye Research*. **53**, pp.1–20.

Ibeh, C. 1998. Handbook of Thermoset Plastics (Second Edition) *In*:, pp.23–71.

Imaizumi, A., Kusaka, S., Takaesu, S., Sawaguchi, S. and Shimomura, Y. 2016. Subretinal Fluid Drainage and Vitrectomy Are Helpful in Diagnosing and Treating Eyes with Advanced Coats' Disease. *Case Reports in Ophthalmology*. **7**(1), pp.223–229.

Itasaki, N. 2003. Wise, a context-dependent activator and inhibitor of Wnt signalling. *Development*. **130**(18), pp.4295–4305.

Jackson, D.P., Lewis, F.A., Taylor, G.R., Boylston, A.W. and Quirke, P. 1990. Tissue extraction of DNA and RNA and analysis by the polymerase chain reaction. *Journal of Clinical Pathology*. **43**(6), pp.499–504.

Jaffe, M.S., Shields, J.A., Canny, C.L., Eagle, R.C.J. and Fry, R.L. 1977. Retinoblastoma simulating Coats' disease: a clinicopathologic report. *Annals of ophthalmology*. **9**(7), pp.863–868.

Janecka, A., Adamczyk, A. and Gasińska, A. 2015. Comparison of eight commercially available kits for DNA extraction from formalin-fixed paraffin-embedded tissues. *Analytical Biochemistry*. **476**, pp.8–10.

Jeffery, G. 2001. Architecture of the optic chiasm and the mechanisms that sculpt its development. *Physiological reviews*. **81**(4), pp.1393–414.

Jiang, D., Hatahet, Z., Melamede, R.J., Kow, Y.W. and Wallace, S.S. 1997. Characterization of Escherichia coli Endonuclease VIII *. *The Journal of Biological Chemistry*. **272**(51), pp.32230–32239.

Jiang, Q., Shan, K., Qun-Wang, X., RM, Z., Yang, H., Liu, C., YJ, L., Yao, J., XM, L., Shen, Y., Cheng, H., Yuan, J., YY, Z. and Yan, B. 2016. Long non-coding RNA-MIAT promotes neurovascular remodeling in the eye and brain. *Oncotarget*. **7**(31), pp.49688–49698.

Jiang, X., Zhou, L., Cheng, J., Zhang, H., Wang, H., Chen, Z., Shi, F. and Zhu, C. 2014. A novel method for the sensitive detection of mutant proteins using a covalent-bonding tube-based proximity ligation assay. *Analytica Chimica Acta*. **841**, pp.17–23.

Jiao, X., Ventruto, V., Trese, M.T., Shastry, B.S. and Hejtmancik, J.F. 2004. Autosomal recessive familial exudative vitreoretinopathy is associated with mutations in LRP5. *American journal of human genetics*. **75**(5), pp.878–884.

Johnson, D.R., Levanat, S. and Bale, A.E. 1995. Direct molecular analysis of archival tumor tissue for loss of heterozygosity. *Biotechniques*. **19**(2),

p.190.
Jones, D. 1976. Aldehydes-Chemistry of fixation and preservation with aldehydes. *Monograph on Oceanographic Methodology*.
Junge, H.J., Yang, S., Burton, J.B., Paes, K., Shu, X., French, D.M., Costa, M., Rice, D.S. and Ye, W. 2009. TSPAN12 Regulates Retinal Vascular Development by Promoting Norrin- but Not Wnt-Induced FZD4/B-Catenin Signaling. *Cell*. **139**(2), pp.299–311.
Kalia, S.S., Adelman, K., Bale, S.J., Chung, W.K., Eng, C., Evans, J.P., Herman, G.E., Hufnagel, S.B., Klein, T.E., Korf, B.R., McKelvey, K.D., Ormond, K.E., Richards, C.S., Vlangos, C.N., Watson, M., Martin, C.L. and Miller, D.T. 2017. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. *Genetics in Medicine*. **19**(2), pp.249–255.
Karali, M. and Banfi, S. 2018. Non-coding RNAs in retinal development. *Human Genetics*. (1), pp.558–578.
Karjosukarso, D.W., van Gestel, S.H.C., Qu, J., Kouwenhoven, E.N., Duijkers, L., Garanto, A., Zhou, H. and Collin, R.W.J. 2018. An FEVR-associated mutation in ZNF408 alters the expression of genes involved in the development of vasculature. *Human Molecular Genetics*. **27**(20), pp.3519–3527.
Karlsen, F., Kalantari, M., Chitemerere, M., Johansson, B. and Hagmar, B. 1994. Modifications of human and viral deoxyribonucleic acid by formaldehyde fixation. *Laboratory investigation; a journal of technical methods and pathology*. **71**(4), pp.604–611.
Kashani, A.H., Learned, D., Nudleman, E., Drenser, K.A., Capone, A. and Trese, M.T. 2014. High prevalence of peripheral retinal vascular anomalies in family members of patients with familial exudative vitreoretinopathy. *Ophthalmology*. **121**(1), pp.262–268.
Kato, M., Patel, M.S., Levasseur, R., Lobov, I., Chang, B.H.J., Glass, D.A., Hartmann, C., Li, L., Hwang, T.H., Brayton, C.F., Lang, R.A., Karsenty, G. and Chan, L. 2002. Cbfa1-independent decrease in osteoblast proliferation, osteopenia, and persistent embryonic eye vascularization in mice deficient in Lrp5, a Wnt coreceptor. *Journal of Cell Biology*. **157**(2), pp.303–314.
Ke, J., Harikumar, K.G., Erice, C., Chen, C., Gu, X., Wang, L., Parker, N., Cheng, Z., Xu, W., Williams, B.O., Melcher, K., Miller, L.J. and Eric Xu, H. 2013. Structure and function of Norrin in assembly and activation of a Frizzled 4-Lrp5/6 complex. *Genes and Development*. **27**(21).
Kent, W.J., Sugnet, C.W., Furey, T.S. and Roskin, K.M. 2002. The Human Genome Browser at UCSC W. *Journal of medicinal chemistry*. **12**(6), pp.996–1006.
Khan, K., Logan, C. V., Mckibbin, M., Sheridan, E., Elçioglu, N.H., Yenice, O., Parry, D.A., Fernandez-fuentes, N., Abdelhamed, Z.I.A., Al-maskari, A., Poulter, J.A., Mohamed, M.D., Carr, I.M., Morgan, J.E., Jafri, H., Raashid, Y., Taylor, G.R., Johnson, C.A., Inglehearn, C.F., Toomes, C. and Ali, M. 2012. Next generation sequencing identifies mutations in atonal homolog 7 (ATOH7) in families with global eye developmental defects. *Human Molecular Genetics*. **21**(4), pp.776–783.
Kim, D.-H., Inagaki, Y., Suzuki, T., Ioka, R.X., Yoshioka, S.Z., Magoori, K.,

Kang, M.-J., Cho, Y., Nakano, A.Z., Liu, Q., Fujino, T., Suzuki, H., Sasano, H. and Yamamoto, T.T. 1998. A new low density lipoprotein receptor related protein, LRP5, is expressed in hepatocytes and adrenal cortex, and recognizes apolipoprotein E. *Journal of Biochemistry*. **124**(6), pp.1072–1076.

Kircher, M. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. **46**(3), pp.310–315.

Koenekoop, R.K., Wang, H., Majewski, J., Wang, X., Lopez, I., Ren, H., Chen, Y., Li, Y., Fishman, G.A., Genead, M., Schwartzentruber, J., Solanki, N., Traboulso, E.I., Cheng, J., Logan, C. V, Mckibbin, M., Hayward, B.E., Parry, D.A., Johnson, C.A., Nageeb, M., Canada, C.F. of R.D.G., Poulter, J.A., Mohamed, M.D., Jafri, H., Rashid, Y., Taylor, G.R., Keser, V., Mardon, G., Xu, H., Inglehearn, C.F., Fu, Q., Toomes, C. and Chen, R. 2012. Mutations in NMNAT1 cause Leber congenital amaurosis and identify a new disease pathway for retinal degeneration. *Nat Genet*. **44**(9), pp.1035–1039.

Kolb, H., Fernandez, E. and Nelson, R. 2005. *Simple Anatomy of the Retina*.

Komor, A.C., Badran, A.H. and Liu, D.R. 2017. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*. **168**(1–2), pp.20–36.

Kondo, H., Kusaka, S., Yoshinaga, A., Uchio, E., Tawara, A., Hayashi, K. and Tahira, T. 2011. Mutations in the TSPAN12 gene in Japanese patients with familial exudative vitreoretinopathy. *American Journal of Ophthalmology*. **151**(6), p.1095–1100.e1.

Kondo, H., Matsushita, I., Tahira, T., Uchio, E. and Kusaka, S. 2016. Mutations in *ATOH7* gene in patients with nonsyndromic congenital retinal nonattachment and familial exudative vitreoretinopathy. *Ophthalmic Genetics*. **37**(4), pp.462–464.

Kondo, H., Ohno, K., Tahira, T., Hayashi, H., Oshima, K. and Hayashi, K. 2001. Delineation of the critical interval for the familial exudative vitreoretinopathy gene by linkage and haplotype analysis. *Human Genetics*. **108**(5), pp.368–375.

Koshiba, M., Ogawa, K., Hamazaki, S., Sugiyama, T., Ogawa, O. and Kitajima, T. 1993. The Effect of Formalin Fixation on DNA and the Extraction of High-molecular-weight DNA from Fixed and Embedded Tissues. *Pathology Research and Practice*. **189**(1), pp.66–72.

Kovalenko, O. V., Metcalf, D.G., DeGrado, W.F. and Hemler, M.E. 2005. Structural organization and interactions of transmembrane domains in tetraspanin proteins. *BMC Structural Biology*. **5**, pp.1–20.

Kramer, G.D., Say, E.A.T. and Shields, C.L. 2015. Simultaneous Novel Mutations of LRP5 and TSPAN12 in a Case of Familial Exudative Vitreoretinopathy. , pp.1–5.

Krawczak, M., Reiss, J. and Cooper, D.N. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Human Genetics*. **90**, pp.41–54.

Kuechler, A., Willemsen, M.H., Albrecht, B., Bacino, C.A., Bartholomew, D.W., van Bokhoven, H., van den Boogaard, M.J.H., Bramswig, N., Büttner, C., Cremer, K., Czeschik, J.C., Engels, H., van Gassen, K., Graf, E., van Haelst, M., He, W., Hogue, J.S., Kempers, M., Koolen, D., Monroe, G., de Munnik, S., Pastore, M., Reis, A., Reuter, M.S., Tegay, D.H., Veltman, J., Visser, G., van Hasselt, P., Smeets, E.E.J., Vissers, L., Wieland, T.,

Wissink, W., Yntema, H., Zink, A.M., Strom, T.M., Lüdecke, H.J., Kleefstra, T. and Wieczorek, D. 2014. De novo mutations in beta-catenin (CTNNB1) appear to be a frequent cause of intellectual disability: expanding the mutational and clinical spectrum. *Human Genetics*. **134**(1), pp.97–109.

LaDuca, H., Farwell, K.D., Vuong, H., Lu, H.M., Mu, W., Shahmirzadi, L., Tang, S., Chen, J., Bhide, S. and Chao, E.C. 2017. Exome sequencing covers >98% of mutations identified on targeted next generation sequencing panels. *PLoS ONE*. **12**(2), pp.1–11.

Lai, M.B., Zhang, C., Shi, J., Johnson, V., Khandan, L., McVey, J., Klymkowsky, M.W., Chen, Z. and Junge, H.J. 2017. TSPAN12 Is a Norrin Co-receptor that Amplifies Frizzled4 Ligand Selectivity and Signaling. *Cell Reports*. **19**(13), pp.2809–2822.

Lam, H.D., Samuel, M.A., Rao, N.A. and Murphree, A.L. 2008. Retinoblastoma presenting as Coats' disease. *Eye*. **22**(9), pp.1196–1197.

Lang, R. a and Bishop, J.M. 1993. Macrophages are required for cell death and tissue remodeling in the developing mouse eye. *Cell*. **74**(3), pp.453–462.

Leber, T.H. 1912. Ueber eine durch Vorkommen multipler Miliaraneurysmen charakterisierte Form von Retinaldegeneration. *Albrecht von Graefes Archiv für Ophthalmologie*. **81**(1), pp.1–14.

Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D. and Radivojac, P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics*. **25**(21), pp.2744–2750.

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

Li, J.K., Fei, P., Li, Y., Huang, Q.J., Zhang, Q., Zhang, X., Rao, Y.Q., Li, J. and Zhao, P. 2016. Identification of novel KIF11 mutations in patients with familial exudative vitreoretinopathy and a phenotypic analysis. *Scientific Reports*. **6**(January), pp.1–7.

Liang, P., Xu, Y., Zhang, X., Ding, C., Huang, R., Zhang, Z., Lv, J., Xie, X., Chen, Y., Li, Y., Sun, Y., Bai, Y., Songyang, Z., Ma, W., Zhou, C. and Huang, J. 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein and Cell*. **6**(5), pp.363–372.

de Ligt, J., Willemsen, M.H., van Bon, B.W.M., Kleefstra, T., Yntema, H.G., Kroes, T., Vulto-van Silfhout, A.T., Koolen, D.A., de Vries, P., Gilissen, C., del Rosario, M., Hoischen, A., Scheffer, H., de Vries, B.B.A., Brunner, H.G., Veltman, J.A. and Vissers, L.E.L.M. 2012. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *New England Journal of Medicine*. **367**(20), pp.1921–1929.

Lin, C.J., Hwang, J.F., Chen, Y.T. and Chen, S.N. 2010. The effect of intravitreal bevacizumab in the treatment of Coats disease in children. *Retina*. **30**(4), pp.617–622.

Lin, Y., Gao, H., Chen, C., Zhu, Y., Li, T., Liu, B., Ma, C., Jiang, H., Li, Y., Huang, Y., Wu, Q., Li, H., Liang, X., Jin, C., Ye, J., Huang, X. and Lu, L. 2018. Clinical and next-generation sequencing findings in a Chinese family exhibiting severe familial exudative vitreoretinopathy. *International Journal of Molecular Medicine*. **41**(2), pp.773–782.

Lindahl, T., Ljungquist, S., Siegert, W., Nyberg, B. and Sperens, B. 1977. DNA

N-Glycosidases. *The Journal of biological chemistry*. **252**(10), pp.3286–3294.

Ling, T. and Stone, J. 1988. The development of astrocytes in the cat retina: evidence of migration from the optic nerve. *Developmental Brain Research*. **44**(1), pp.73–85.

Liu, C.C., Pearson, C. and Bu, G. 2009. Cooperative folding and ligand-binding properties of LRP6 β-propeller domains. *Journal of Biological Chemistry*. **284**(22), pp.15299–15307.

Livesey, F.J. and Cepko, C.L. 2001. Vertebrate Neural Cell-Fate Determination : Lessons From the Retina. *Nature Reviews Neuroscience*. **2**.

Lobov, I.B., Rao, S., Carroll, T.J., Vallance, J.E., Ito, M., Ondr, J.K., Kurup, S., Glass, D.A., Patel, M.S., Shu, W., Morrisey, E.E., McMahon, A.P., Karsenty, G. and Lang, R.A. 2005. WNT7b mediates macrophage-induced programmed cell death in patterning of the vasculature. *Nature*. **437**(7057), pp.417–421.

Lord, E.C., Poulter, J.A., Webster, A.R., Sergouniotis, P., Khan, K.N., Benke, P.J., Friedman, L., Ali, M., Inglehearn, C.F. and Toomes, C. 2017. Mutations in SLC38A8 and FOXD1 in patients with nystagmus and foveal hypoplasia. *Investigative Ophthalmology & Visual Science*. **58**(8), p.2786.

Ludyga, N., Grunwald, B., Azimzadeh, O., Englert, S., H??fler, H., Tapio, S. and Aubele, M. 2012. Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. *Virchows Archiv*. **460**(2), pp.131–140.

Lutty, G.A. and McLeod, D.S. 2018. Development of the hyaloid, choroidal and retinal vasculatures in the fetal human eye. *Progress in Retinal and Eye Research*. **62**, pp.58–76.

Macdonald, B.T. and He, X. 2012. Frizzled and LRP5 / 6 Receptors for Wnt / β - Catenin Signaling. *Cold Spring Harbor Perspectives in Biology*. **4**(12), pp.1–12.

Machkour, B., Denis, P. and Kodjikian, L. 2017. Unusual Presentation of Type 1 Idiopathic Macular Telangiectasia. *Ophthalmological Medicine*., pp.1–7.

Mafee, M.F., Goldberg, M.F., Cohen, S.B., Gotsis, E.D., Safran, M., Chekuri, L. and Raofi, B. 1989. Magnetic resonance imaging versus computed tomography of leukocoric eyes and use of in vitro proton magnetic resonance spectroscopy of retinoblastoma. *Ophthalmology*. **96**(7), pp.965–75.

Mannu, G.S. 2014. Retinal phototransduction. *Neurosciences*. **19**(4), pp.275–280.

Mao, J., Wang, J., Liu, B., Pan, W., Farr, G.H., Flynn, C., Yuan, H., Takada, S., Kimelman, D., Li, L., Wu, D. and York, N. 2001. Low-Density Lipoprotein Receptor-Related Protein-5 Binds to Axin and Regulates the Canonical Wnt Signaling Pathway. *Molecular Cell*. **7**, pp.801–809.

Maquat, L.E. 1996. Defects in RNA splicing and the consequence of shortened translational reading frames. *American journal of human genetics*. **59**(2), pp.279–286.

Maquat, L.E. 2004. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. *Nature Reviews Molecular Cell Biology*. **5**(2), pp.89–99.

Margolis, R., Couvillion, S.S., Mavrofrides, E.C., Hess, D. and Murray, T.G.

2004. Progression of Familial Exudative Vitreoretinopa- thy After Laser Treatment. *Archives of Ophthalmology*. **22**, pp.1717–1719.

Marosy, B.A., Craig, B.D., Hetrick, K.N., Witmer, P.D., Ling, H., Griffith, S.M., Myers, B., Ostrander, E.A., Stanford, J.L., Brody, L.C., Doheny, K.F., Marosy, B.A., Craig, B.D., Hetrick, K.N., Witmer, P.D., Ling, H., Griffith, S.M., Myers, B., Ostrander, E.A., Stanford, J.L., Brody, L.C. and Doheny, K.F. 2017. Generating Exome Enriched Sequencing Libraries from Formalin-Fixed, Paraffin-Embedded Tissue DNA for Next-Generation Sequencing. *Current Protocols in Human Genetics*. **18**(10), pp.1–25.

Marquardt, T. and Gruss, P. 2002. Generating neuronal diversity in the retina: One for nearly all. *Trends in Neurosciences*. **25**(1), pp.32–38.

Matsumine, A., Ogai, A., Senda, T., Okumura, N., Satoh, K., Baeg, G.-H., Kawahara, T., Kobayashi, S., Okada, M., Toyoshima, K. and Akiyama, T. 1996. Binding of APC to the Human Homolog of the Drosophila Discs Large Tumor Suppressor Protein. *Science*. **272**(5264), pp.1020–1023.

Mazur, P. 1970. Cryobiology : The Freezing of Biological Systems. *Science*. **168**(3934), pp.939–949.

McClements, M.E. and Maclaren, R.E. 2017. Adeno-associated virus (AAV) dual vector strategies for gene therapy encoding large transgenes. *Yale Journal of Biology and Medicine*. **90**(4), pp.611–623.

McGettrick, P.M. and Loeffler, K.U. 1987. Bilateral Coats' disease in an infant (a clinical, angiographic, light and electron microscopic study). *Eye*. **1**(1), pp.136–45.

McLeod, D.S., Hasegawa, T., Prow, T., Merges, C. and Lutty, G. 2006. The initial fetal human retinal vasculature develops by vasculogenesis. *Developmental Dynamics*. **235**(12), pp.3336–3347.

Mears, K., Bakall, B., Harney, L.A., Penticoff, J.A. and Stone, E.M. 2015. Autosomal Dominant Microcephaly Associated With Congenital Lymphedema and Chorioretinopathy Due to a Novel Mutation in KIF11 Pigmented Limbal Nodule Consistent With a Ciliary Body Nevus in an Organ Donor. *JAMA ophthalmology*. **133**(6), pp.720–721.

Meindl, A., Berger, W., Meitinger, T., van de Pol, D., Achatz, H., Dorner, C., Haasemann, M., Hellebrand, H., Gal, A., Cremers, F & Ropers, H.H. 1992. Norrie disease is caused by mutations in an extracellular protein resembling C-terminal globular domain of mucins. *Nat Genet.* **2**(2), pp.139–143.

Meitinger, T., Meindl, A., Bork, P., Rost, B., Sander, C., Haasemann, M. & Murken, J. 1993. Molecular modelling of the Norrie disease protein predicts a cysteine knot growth factor tertiary structure. *Nature Genetics*. **5**(4), pp.376–380.

Mendez, P., Fang, L.T., Jablons, D.M. and Kim, I.J. 2017. Systematic comparison of two whole-genome amplification methods for targeted next-generation sequencing using frozen and FFPE normal and cancer tissues. *Scientific Reports*. **7**(1), pp.1–10.

Meyer-Schwickerath, G. 1959. *Lichtkoagulation*. F. Enke.

Miller, K.E., Willis, M.J. and McClatchey, S.K. 2015. A case of familial exudative vitreoretinopathy identified after genetic testing. *Journal of AAPOS*. **19**(2), pp.178–180.

Mirzaa, G.M., Enyedi, L., Parsons, G., Collins, S., Medne, L., Adams, C., Ward, T., Davitt, B., Bicknese, A., Zackai, E., Toriello, H., Dobyns, W.B.

and Christian, S. 2014. Congenital microcephaly and chorioretinopathy due to de novo heterozygous KIF11 mutations: Five novel mutations and review of the literature. *American Journal of Medical Genetics, Part A*. **164**(11), pp.2879–2886.

Molenaar, M., Van De Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. and Clevers, H. 1996. XTcf-3 transcription factor mediates b-catenin-induced axis formation in xenopus embryos. *Cell*. **86**, pp.391–399.

Moore, A.T. 2017. Genetic Testing for Inherited Retinal Disease. *Ophthalmology*. **124**(9), pp.1254–1255.

Mori, K., Iwao, K., Miyoshi, Y., Nakagawara, A., Kofu, K., Akiyama, T., Arita, N., Hayakawa, T. and Nakamura, Y. 1998. Identification of brain-specific splicing variants of the hDLG1 gene and altered splicing in neuroblastoma cell lines. *Journal of Human Genetics*. **43**(2), pp.123–127.

Morris, B., Foot, B. and Mulvihill, A. 2010. A population-based study of Coats disease in the United Kingdom II: investigation, treatment, and outcomes. *Eye (London, England)*. **24**(12), pp.1802–7.

Mudhar, H.S., Pollock, R. a, Wang, C., Stiles, C.D. and Richardson, W.D. 1993. PDGF and its receptors in the developing rodent retina and optic nerve. *Development*. **118**(2), pp.539–52.

Musada, G.R., Syed, H., Jalali, S., Chakrabarti, S. and Kaur, I. 2016. Mutation spectrum of the FZD-4, TSPAN12 AND ZNF408 genes in Indian FEVR patients. *BMC ophthalmology*. **16**, p.90.

Narumi, S., Numakura, C., Shiihara, T., Seiwa, C., Nozaki, Y., Yamagata, T., Momoi, M.Y., Watanabe, Y., Yoshino, M., Matsuishi, T., Nishi, E., Kawame, H., Akahane, T., Nishimura, G., Emi, M. and Hasegawa, T. 2010. Various types of LRP5 mutations in four patients with osteoporosis-pseudoglioma syndrome: Identification of a 7.2-kb microdeletion using oligonucleotide tiling microarray. *American Journal of Medical Genetics, Part A*. **152**(1), pp.133–140.

Nathans, J. 1999. The Evolution and Physiology of Human Review Color Vision: Insights from Molecular Genetic Studies of Visual Pigments. *Neuron*. **24**, pp.299–312.

Ng, P.C. and Henikoff, S. 2001. Predicting Deleterious Amino Acid Substitutions Predicting Deleterious Amino Acid Substitutions. *Genome Research*. **11**, pp.863–874.

Nichols, C.W. and Lambertsen, C.J. 1969. Effects of high oxygen pressures on the eye. *New England Journal of Medicine*. **281**(1), pp.1361–1367.

Nickle, B. and Robinson, P.R. 2007. The opsins of the vertebrate retina: Insights from structural, biochemical, and evolutionary studies. *Cellular and Molecular Life Sciences*. **64**(22), pp.2917–2932.

Nikopoulos, K., Gilissen, C., Hoischen, A., Erik van Nouhuys, C., Boonstra, F.N., Blokland, E.A.W., Arts, P., Wieskamp, N., Strom, T.M., Ayuso, C., Tilanus, M.A.D., Bouwhuis, S., Mukhopadhyay, A., Scheffer, H., Hoefsloot, L.H., Veltman, J.A., Cremers, F.P.M. and Collin, R.W.J. 2010. Next-Generation Sequencing of a 40 Mb Linkage Interval Reveals TSPAN12 Mutations in Patients with Familial Exudative Vitreoretinopathy. *American Journal of Human Genetics*. **86**(2), pp.240–247.

Nikopoulos, K., Venselaar, H., Collin, R.W.J., Riveiro-Alvarez, R., Boonstra, F.N., Hooymans, J.M.M., Mukhopadhyay, A., Shears, D., Van Bers, M.,

De Wijs, I.J., Van Essen, A.J., Sijmons, R.H., Tilanus, M.A.D., Van Nouhuys, C.E., Ayuso, C., Hoefsloot, L.H. and Cremers, F.P.M. 2010. Overview of the mutation spectrum in familial exudative vitreoretinopathy and Norrie disease with identification of 21 novel variants in FZD4, LRP5, and NDP. *Human Mutation*. **31**(6), pp.656–666.

Norrie, G. 1927. Causes of blindness in children. *Acta Ophthalmologica*. **5**(1–3), pp.357–386.

van Nouhuys, C.E. 1991. Signs, complications, and platelet aggregation in familial exudative vitreoretinopathy. *American journal of ophthalmology*. **111**(1), pp.34–41.

Ober, R.R., Bird, A.C., Hamilton, A.M. and Sehmi, K. 1980. Autosomal dominant exudative vitreoretinopathy. . (64), pp.112–120.

Okamoto, N., Hatsukawa, Y., Shimojima, K. and Yamamoto, T. 2011. Submicroscopic deletion in 7q31 encompassing CADPS2 and TSPAN12 in a child with autism spectrum disorder and PHPV. *American Journal of Medical Genetics, Part A*. **155**(7), pp.1568–1573.

Ormondroyd, E., Mackley, M.P., Blair, E., Craft, J., Knight, J.C., Taylor, J.C., Taylor, J. and Watkins, H. 2018. 'Not pathogenic until proven otherwise': Perspectives of UK clinical genomics professionals toward secondary findings in context of a Genomic Medicine Multidisciplinary Team and the 100,000 Genomes Project. *Genetics in Medicine*. **20**(3), pp.320–328.

Ostergaard, P., Simpson, M.A., Mendola, A., Vasudevan, P., Connell, F.C., Van Impel, A., Moore, A.T., Loeys, B.L., Ghalamkarpour, A., Onoufriadis, A., Martinez-Corral, I., Devery, S., Leroy, J.G., Van Laer, L., Singer, A., Bialer, M.G., McEntagart, M., Quarrell, O., Brice, G., Trembath, R.C., Schulte-Merker, S., Makinen, T., Vikkula, M., Mortimer, P.S., Mansour, S. and Jeffery, S. 2012. Mutations in KIF11 cause autosomal-dominant microcephaly variably associated with congenital lymphedema and chorioretinopathy. *American Journal of Human Genetics*. **90**(2), pp.356–362.

Panagiotou, E. 2018. *Using NGS to identify new genes and modifiers underlying Familial Exudative Vitreoretinopathy (FEVR)*. Ph.D. thesis, University of Leeds.

Panagiotou, E.S., Sanjurjo Soriano, C., Poulter, J.A., Lord, E.C., Dzulova, D., Kondo, H., Hiyoshi, A., Chung, B.H.-Y., Chu, Y.W.-Y., Lai, C.H.Y., Tafoya, M.E., Karjosukarso, D., Collin, R.W.J., Topping, J., Downey, L.M., Ali, M., Inglehearn, C.F. and Toomes, C. 2017. Defects in the Cell Signaling Mediator β-Catenin Cause the Retinal Vascular Condition FEVR. *American Journal of Human Genetics*. **100**(6), pp.960–968.

Park, S., Cho, H.J., Lee, D.W., Kim, C.G. and Kim, J.W. 2016. Intravitreal bevacizumab injections combined with laser photocoagulation for adult-onset Coats disease. *Graefe's Archive for Clinical and Experimental Ophthalmology*. **254**(8), pp.1511–1517.

Park, T.S., Bhutto, I., Zimmerlin, L., Huo, J.S., Nagaria, P., Miller, D., Rufaihah, A.J., Talbot, C., Aguilar, J., Grebe, R., Merges, C., Reijo-Pera, R., Feldman, R.A., Rassool, F., Cooke, J., Lutty, G. and Zambidis, E.T. 2014. Vascular Progenitors from Cord Blood-Derived iPSC Possess Augmented Capacity for Regenerating Ischemic Retinal Vasculature. *Circulation.* **129**(3), pp.359–372.

Pasquay, C., Wang, L.F., Lorenz, B. and Preising, M.N. 2015. Bestrophin 1 -

Phenotypes and Functional Aspects in Bestrophinopathies. *Ophthalmic Genetics*. **36**(3), pp.193–212.

Pefkianaki, M., Hasanreisoglu, M., Suchy, S.F. and Shields, C.L. 2016. Familial exudative vitreoretinopathy with a novel LRP5 mutation. *Journal of pediatric ophthalmology and strabismus*. **53**, pp.39–42.

Philippakis, A.A., Azzariti, D.R., Beltran, S., Brookes, A.J., Brownstein, C.A., Brudno, M., Brunner, H.G., Buske, O.J., Carey, K., Doll, C., Dumitriu, S., Dyke, S.O.M., den Dunnen, J.T., Firth, H. V., Gibbs, R.A., Girdea, M., Gonzalez, M., Haendel, M.A., Hamosh, A., Holm, I.A., Huang, L., Hurles, M.E., Hutton, B., Krier, J.B., Misyura, A., Mungall, C.J., Paschall, J., Paten, B., Robinson, P.N., Schiettecatte, F., Sobreira, N.L., Swaminathan, G.J., Taschner, P.E., Terry, S.F., Washington, N.L., Züchner, S., Boycott, K.M. and Rehm, H.L. 2015. The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery. *Human Mutation*. **36**(10), pp.915–921.

Pikor, L.A., Enfield, K.S.S., Cameron, H. and Lam, W.L. 2011. DNA Extraction from Paraffin Embedded Material for Genetic and Epigenetic Analyses. *Journal of Visualized Experiments*. (49), pp.15–17.

Pinson, K.I., Brennan, J., Monkley, S., Avery, B.J. and Skarnes, W.C. 2000. An LDL receptor-related protein mediates Wnt signaling in mice. *Nature*. **407**, pp.535–538.

Plagnol, V., Curtis, J., Epstein, M., Mok, K.Y., Stebbings, E., Grigoriadou, S., Wood, N.W., Hambleton, S., Burns, S.O., Thrasher, A.J., Kumararatne, D., Doffinger, R. and Nejentsev, S. 2012. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. **28**(21), pp.2747–2754.

Planutis, K., Planutiene, M. and Holcombe, R.F. 2014. A novel signaling pathway regulates colon cancer angiogenesis through Norrin. *Scientific Reports*. **4**, pp.2–6.

Poulter, J.A., Ali, M., Gilmour, D.F., Rice, A., Kondo, H., Hayashi, K., Mackey, D.A., Kearns, L.S., Ruddle, J.B., Craig, J.E., Pierce, E.A., Downey, L.M., Mohamed, M.D., Markham, A.F., Inglehearn, C.F. and Toomes, C. 2010. Mutations in TSPAN12 Cause Autosomal-Dominant Familial Exudative Vitreoretinopathy. *American Journal of Human Genetics*. **86**(2), pp.248–253.

Poulter, J.A., Davidson, A.E., Ali, M., Gilmour, D.F., Parry, D.A., Mintz-Hittner, H.A., Carr, I.M., Bottomley, H.M., Long, V.W., Downey, L.M., Sergouniotis, P.I., Wright, G.A., MacLaren, R.E., Moore, A.T., Webster, A.R., Inglehearn, C.F., Toomes, C., Rice, A., Kondo, H., Hayashi, K., Mackey, D.A., Kearns, L.S., Ruddle, J.B., Craig, J.E., Pierce, E.A., Downey, L.M., Mohamed, M.D., Markham, A.F., Inglehearn, C.F. and Toomes, C. 2012. Recessive mutations in TSPAN12 cause retinal dysplasia and severe familial exudative vitreoretinopathy (FEVR). *Investigative Ophthalmology and Visual Science*. **53**(6), pp.2873–2879.

Prasov, L., Masud, T., Khaliq, S., Mehdi, S.Q., Abid, A., Oliver, E.R., Silva, E.D., Lewanda, A., Brodsky, M.C., Borchert, M., Kelberman, D., Sowden, J.C., Dattani, M.T. and Glaser, T. 2012. ATOH7 mutations cause autosomal recessive persistent hyperplasia of the primary vitreous. *Human Molecular Genetics*. **21**(16), pp.3681–3694.

Provis, J.M. 2001. Development of the primate retinal vasculature. *Progress*

in Retinal and Eye Research. **20**(6), pp.799–821.

Pruitt, K.D., Brown, G.R., Hiatt, S.M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C.M., Hart, J., Landrum, M.J., McGarvey, K.M., Murphy, M.R., O'Leary, N.A., Pujar, S., Rajput, B., Rangwala, S.H., Riddick, L.D., Shkeda, A., Sun, H., Tamez, P., Tully, R.E., Wallin, C., Webb, D., Weber, J., Wu, W., Dicuccio, M., Kitts, P., Maglott, D.R., Murphy, T.D. and Ostell, J.M. 2014. RefSeq: An update on mammalian reference sequences. *Nucleic Acids Research*. **42**(D1), pp.756–763.

Punchihewa, C., Ferreira, A.M., Cassell, R., Rodrigues, P. and Fujii, N. 2009. Sequence requirement and subtype specificity in the high-affinity interaction between human frizzled and dishevelled proteins. *Protein science : a publication of the Protein Society*. **18**(5), pp.994–1002.

Qin, M., Hayashi, H., Oshima, K., Tahira, T., Hayashi, K. and Kondo, H. 2005. Complexity of the genotype-phenotype correlation in familial exudative vitreoretinopathy with mutations im the LRP5 and/or FZD4 genes. *Human Mutation*. **26**(2), pp.104–112.

Qin, M., Kondo, H., Tahira, T. and Hayashi, K. 2008. Moderate reduction of Norrin signaling activity associated with the causative missense mutations identified in patients with familial exudative vitreoretinopathy. *Human Genetics*. **122**(6), pp.615–623.

Quach, N., Goodman, M.F. and Shibata, D. 2004. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. *BMC Clinical Pathology*. **5**, pp.1–5.

Quinodoz, M., Royer-Bertrand, B., Cisarova, K., Di Gioia, S.A., Superti-Furga, A. and Rivolta, C. 2017. DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders. *American Journal of Human Genetics*. **101**(4), pp.623–629.

Quiram, P.A., Drenser, K.A., Lai, M.M., Capone, A. and Trese, M.T. 2008. Treatment of vascularly active familial exudative vitreoretinopathy with pegaptanib sodium (Macugen). *Retina*. **28**(3 SUPPL.), pp.8–12.

Rabiolo, A., Marchese, A., Sacconi, R., Cicinelli, M.V., Grosso, A., Querques, L., Querques, G. and Bandello, F. 2017. Refining Coats' disease by ultra-widefield imaging and optical coherence tomography angiography. *Graefe's Archive for Clinical and Experimental Ophthalmology*. **255**(10), pp.1881–1890.

Rajendran, P., Rengarajan, T., Thangavel, J., Nishigaki, Y., Sakthisekaran, D., Sethi, G. and Nishigaki, I. 2013. The vascular endothelium and human diseases. *International Journal of Biological Sciences*. **9**(10), pp.1057–1069.

Ranchod, T.M., Ho, L.Y., Drenser, K.A., Capone, A. and Trese, M.T. 2011. Clinical presentation of familial exudative vitreoretinopathy. *Ophthalmology*. **118**(10), pp.2070–2075.

Rao, F.Q., Cai, X.B., Cheng, F.F., Cheng, W., Fang, X.L., Li, N., Huang, X.F., Li, L.H. and Jin, Z.B. 2017. Mutations in LRP5, FZD4, TSPAN12, NDP, ZNF408, or KIF11 genes account for 38.7% of Chinese patients with familial exudative vitreoretinopathy. *Investigative Ophthalmology and Visual Science*. **58**(5), pp.2623–2629.

Rao, S., Chun, C., Fan, J., Kofron, M.J., Yang, M.B., Hegde, R.S., Ferrara, N., Copenhagen, D.R. and Lang, R.A. 2013. A direct and melanopsin-dependent fetal light response regulates mouse eye development.

Nature. **494**(7436), pp.243–246.

Reese, A.B. 1956. Telangiectasis of the Retina and Coats' Disease. *American Journal of Ophthalmology*. **42**(1), pp.1–2.

Rehm, H.L., Zhang, D.-S., Brown, M.C., Burgess, B., Halpin, C., Berger, W., Morton, C.C., Corey, D.P. and Chen, Z.-Y. 2002. Vascular defects and sensorineural deafness in a mouse model of Norrie disease. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. **22**(11), pp.4286–4292.

Reichstein, D.A. and Recchia, F.M. 2011. Coats disease and exudative retinopathy. *International ophthalmology clinics*. **51**(1), pp.93–112.

Reinartz, S., Finkernagel, F., Adhikary, T., Rohnalter, V., Schumann, T., Schober, Y., Nockher, W.A., Nist, A., Stiewe, T., Jansen, J.M., Wagner, U., Müller-Brüsselbach, S. and Müller, R. 2016. A transcriptome-based global map of signaling pathways in the ovarian cancer microenvironment associated with clinical outcome. *Genome biology*. **17**(1), p.108.

Richter, M., Gottanka, J., May, C.A., Welge-Lüßen, U., Berger, W. and Lütjen-Drecoll, E. 1998. Retinal vasculature changes in Norrie disease mice. *Investigative Ophthalmology and Visual Science*. **39**(12), pp.2450–2457.

Rivera-Vega, M.R., Chiñas-Lopez, S., Jimenez Vaca, A.L., Arenas-Sordo, M.L., Kofman-Alfaro, S., Messina-Baas, O. and Cuevas-Covarrubias, S.A. 2005. Molecular analysis of the NDP gene in two families with Norrie disease. *Acta Ophthalmologica Scandinavica*. **83**(2), pp.210–214.

Rivera, J.C., Dabouz, R., Noueihed, B., Omri, S., Tahiri, H. and Chemtob, S. 2017. Ischemic retinopathies: Oxidative stress and inflammation. *Oxidative Medicine and Cellular Longevity*., pp.1–16.

Robitaille, J., MacDonald, M.L.E., Kaykas, A., Sheldahl, L.C., Zeisler, J., Dubé, M.-P., Zhang, L.-H., Singaraja, R.R., Guernsey, D.L., Zheng, B., Siebert, L.F., Hoskin-Mott, A., Trese, M.T., Pimstone, S.N., Shastry, B.S., Moon, R.T., Hayden, M.R., Goldberg, Y.P. and Samuels, M.E. 2002. Mutant frizzled-4 disrupts retinal angiogenesis in familial exudative vitreoretinopathy. *Nature Genetics*. **32**(2), pp.326–330.

Robitaille, J.M., Wallace, K., Zheng, B., Beis, M.J., Samuels, M., Hoskin-Mott, A. and Guernsey, D.L. 2009. Phenotypic overlap of familial exudative vitreoretinopathy (FEVR) with persistent fetal vasculature (PFV) caused by FZD4 mutations in two distinct pedigrees. *Ophthalmic Genetics*. **30**(1), pp.23–30.

Robitaille, J.M., Zheng, B., Wallace, K., Beis, M.J., Tatlidil, C., Yang, J., Sheidow, T.G., Siebert, L., Levin, A. V, Lam, W.-C., Arthur, B.W., Lyons, C.J., Jaakkola, E., Tsilou, E., Williams, C. a, Weaver, R.G., Shields, C.L. and Guernsey, D.L. 2011. The role of Frizzled-4 mutations in familial exudative vitreoretinopathy and Coats disease. *The British journal of ophthalmology*. **95**(4), pp.574–579.

Robitaille, J.M., Gillett, R.M., LeBlanc, M.A., Gaston, D., Nightingale, M., Mackley, M.P., Parkash, S., Hathaway, J., Thomas, A., Ells, A., Traboulsi, E.I., Heon, E., Roy, M., Shalev, S., Fernandez, C. V, MacGillivray, C., Wallace, K., Fahiminiya, S., Majewski, J., McMaster, C.R. and Bedard, K. 2014. Phenotypic overlap between familial exudative vitreoretinopathy and microcephaly, lymphedema, and chorioretinal dysplasia caused by KIF11 mutations. *JAMA ophthalmology*. **132**(12), pp.1393–1399.

Roest, P.A.M., Roberts, R.G., van der Tuijn, A.C., Heikoop, J.C., van Ommen,

G.J.B. and den Dunnen, J.T. 1993. Protein truncation test (PTT) to rapidly screen the DMD gene for translation terminating mutations. *Neuromuscular Disorders*. **3**(5–6), pp.391–394.

Roscioli, T., Kamsteeg, E.J., Buysse, K., Maystadt, I., Van Reeuwijk, J., Van Den Elzen, C., Van Beusekom, E., Riemersma, M., Pfundt, R., Vissers, L.E.L.M., Schraders, M., Altunoglu, U., Buckley, M.F., Brunner, H.G., Grisart, B., Zhou, H., Veltman, J.A., Gilissen, C., Mancini, G.M.S., Delrée, P., Willemsen, M.A., Ramadža, D.P., Chitayat, D., Bennett, C., Sheridan, E., Peeters, E.A.J., Tan-Sindhunata, G.M.B., De Die-Smulders, C.E., Devriendt, K., Kayserili, H., El-Hashash, O.A.E.F., Stemple, D.L., Lefeber, D.J., Lin, Y.Y. and Van Bokhoven, H. 2012. Mutations in ISPD cause Walker-Warburg syndrome and defective glycosylation of α-dystroglycan. *Nature Genetics*. **44**(5), pp.581–585.

Royer-Bertrand, B. and Rivolta, C. 2015. Whole genome sequencing as a means to assess pathogenic mutations in medical genetics and cancer. *Cellular and Molecular Life Sciences*. **72**(8), pp.1463–1471.

Ruan, G.X., Barry, E., Yu, D., Lukason, M., Cheng, S.H. and Scaria, A. 2017. CRISPR/Cas9-Mediated Genome Editing as a Therapeutic Approach for Leber Congenital Amaurosis 10. *Molecular Therapy*. **25**(2), pp.331–341.

Rubinstein, E., Naour, F. Le, Lagaudrière-Gesbert, C., Billard, M., Conjeaud, H. and Boucheix, C. 1996. CD9, CD63, CD81, and CD82 are components of a surface tetraspan network connected to HLA-DR and VLA integrins. *European Journal of Immunology*. **26**(11), pp.2657–2665.

Rudenko, G. 2002. Structure of the LDL Receptor Extracellular Domain at Endosomal pH. *Science*. **298**(5602), pp.2353–2358.

Salvo, J., Lyubasyuk, V., Xu, M., Wang, H., Wang, F., Nguyen, D., Wang, K., Luo, H., Wen, C., Shi, C., Lin, D., Zhang, K. and Chen, R. 2015. Next-generation sequencing and novel variant determination in a cohort of 92 familial exudative vitreoretinopathy patients. *Investigative Ophthalmology and Visual Science*. **56**(3), pp.1937–1946.

Sangermano, R., Bax, N.M., Bauwens, M., Van Den Born, L.I., De Baere, E., Garanto, A., Collin, R.W.J., Goercharn-Ramlal, A.S.A., Den Engelsman-Van Dijk, A.H.A., Rohrschneider, K., Hoyng, C.B., Cremers, F.P.M. and Albert, S. 2016. Photoreceptor Progenitor mRNA Analysis Reveals Exon Skipping Resulting from the ABCA4 c.5461-10T→C Mutation in Stargardt Disease. *Ophthalmology*. **123**(6), pp.1375–1385.

Schagat, T., Paguio, A. and Kopish, K. 2007. Normalizing genetic reporter assays: approaches and considerations for increasing consistency and statistical significance. *Cell Notes*. **17**, pp.9–12.

Schander, C., Schander, C., Halanych, K.M. and Halanych, K.M. 2003. DNA, PCR and formalinized animal tissue: a short review and protocols. *Organisms, Diversity & Evolution*. **3**, pp.195–205.

Schlögel, M.J., Mendola, A., Fastré, E., Vasudevan, P., Devriendt, K., De Ravel, T.J., Van Esch, H., Casteels, I., Arroyo Carrera, I., Cristofoli, F., Fieggen, K., Jones, K., Lipson, M., Balikova, I., Singer, A., Soller, M., Mercedes Villanueva, M., Revencu, N., Boon, L.M., Brouillard, P. and Vikkula, M. 2015. No evidence of locus heterogeneity in familial microcephaly with or without chorioretinopathy, lymphedema, or mental retardation syndrome. *Orphanet Journal of Rare Diseases*. **10**(1), pp.1–11.

Schnitzer, J. 1987. Retinal astrocytes: their restriction to vascularized parts of the mammalian retina. *Neuroscience Letters*. **78**(1), pp.29–34.

Schuback, D.E., Chen, Z.Y., Craig, I.W., Breakefield, X.O. and Sims, K.B. 1995. Mutations in the Norrie disease gene. *Human Mutation*. **5**(4), pp.285–292.

Schweiger, M.R., Kerick, M., Timmermann, B., Albrecht, M.W., Borodina, T., Parkhomchuck, D., Zatloukal, K. and Lehrach, H. 2009. Genome-wide massively parallel sequencing of formaldehyde fixed-paraffin embedded (FFPE) tumor tissues for copy-number-and mutation-analysis. *PLoS ONE*. **4**(5), pp.3–9.

Semënov, M. V., Tamai, K., Brott, B.K., Kühl, M., Sokol, S. and He, X. 2001. Head inducer dickkopf-1 is a ligand for Wnt coreceptor LRP6. *Current Biology*. **11**(12), pp.951–961.

Semënov, M., Tamai, K. and He, X. 2005. SOST is a ligand for LRP5/LRP6 and a Wnt signaling inhibitor. *Journal of Biological Chemistry*. **280**(29), pp.26770–26775.

Semenov, M. V. and He, X. 2006. LRP5 mutations linked to high bone mass diseases cause reduced LRP5 binding and inhibition by SOST. *Journal of Biological Chemistry*. **281**(50), pp.38276–38284.

Sengüven, B., Baris, E., Oygur, T. and Berktas, M. 2014. Comparison of methods for the extraction of DNA from formalin-fixed, paraffin-embedded archival tissues. *International Journal of Medical Sciences*. **11**(5), pp.494–499.

Seo, S.H., Yu, Y.S., Park, S.W., Kim, J.H., Kim, H.K., Cho, S.I., Park, H., Lee, S.J., Seong, M.W., Park, S.S. and Kim, J.Y. 2015. Molecular characterization of FZD4, LRP5, and TSPAN12 in familial exudative vitreoretinopathy. *Investigative Ophthalmology and Visual Science*. **56**(9), pp.5143–5151.

Seo, S.H., Kim, M.J., Park, S.W., Kim, J.H., Yu, Y.S., Song, J.Y., Cho, S.I., Ahn, J.H., Oh, Y.H., Lee, J.-S., Lee, S., Seong, M.-W., Park, S.S. and Kim, J.Y. 2016. Large Deletions of TSPAN12 cause Familial Exudative Vitreoretinopathy (FEVR). *Investigative Opthalmology and Visual Science*. **57**(15), pp.6902–6908.

Shaikh, S. and Trese, M.T. 2002. Retinal reattachment facilitated by short-term perfluorocarbon liquid tamponade in a case of fevr and rhegmatogenous retinal detachment. *Retina*. **22**(5), pp.674–676.

Shields, J.A., Shields, C.L., Honavar, S.G., Demirci, H. and Cater, J. 2001a. Disease : The 2000 Proctor Lecture. *American Journal of Ophthalmology*. **131**, pp. 572–583.

Shields, J.A., Shields, C.L., Honavar, S.G. and Demirci, H. 2001b. The 2000 Sanford Gifford Memorial Lecture. *American Journal of Ophthalmology*. **131,** pp.561–571.

Shukla, D., Singh, J., Sudheer, G., Soman, M., John, R.K., Ramasamy, K. and Perumalsamy, N. 2003. Original Article Familial Exudative Vitreoretinopathy (FEVR). *Indian J Ophthalmol*. **51**, pp.323–328.

Sigler, E.J., Randolph, J.C., Calzada, J.I., Wilson, M.W. and Haik, B.G. 2014. Current management of Coats disease. *Survey of Ophthalmology*. **59**(1), pp.30–46.

Sisk, R.A. 2012. Intravenous Sodium Fluorescein 10% for Laser Ablation of Subtle Retinal Neovascularization in FEVR. *Journal of pediatric*

ophthalmology and strabismus. **49**(2), pp.126–127.

Skage, M. and Schander, C. 2007. DNA from formalin-fixed tissue: extraction or repair? That is the question. *Marine Biology Research*. **3**(5), pp.289–295.

Small, K.W., DeLuca, A.P., Whitmore, S., Rosenberg, T., Silva-Garcia, R., Udar, N., Puech, B., Garcia, C.A., Rice, T.A., Fishman, G.A., Heon, E., Folk, J.C., Streb, L.M., Haas, C.M., Wiley, L.A., Scheetz, T.E., Fingert, J.H., Mullins, R.F., Tucker, B.A. and Stone, E.M. 2016. North Carolina Macular Dystrophy is caused by dysregulation of the retinal transcription factor PRDM13. *Ophthalmology*. **123**(1), pp.9–18.

Smallwood, P.M., Williams, J., Xu, Q., Leahy, D.J. and Nathans, J. 2007. Mutational analysis of Norrin-Frizzled4 recognition. *Journal of Biological Chemistry*. **282**(6), pp.4057–4068.

Smits, A.H. and Vermeulen, M. 2016. Characterizing Protein – Protein Interactions Using Mass Spectrometry : Challenges and Opportunities. *Trends in Biotechnology*. **34**(10), pp.825–834.

Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G. and Landegren, U. 2006. Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nature Methods*. **3**(12), pp.995–1000.

Soliman, S.E., Wan, M.J., Heon, E., Hazrati, L.-N. and Gallie, B. 2016. Retinoblastoma versus advanced Coats' disease: Is enucleation the answer? *Ophthalmic Genetics*. **6810**(July), pp.1–3.

Soriano, C.S. 2017. *Functional characterisation of FEVR-related LGR4 missense mutations. Implications in Norrin-β-Catenin signalling pathway and angiogenesis.* Ph.D. thesis, University of Leeds.

Spencer, D.H., Sehn, J.K., Abel, H.J., Watson, M.A., Pfeifer, J.D. and Duncavage, E.J. 2013. Comparison of clinical targeted next-generation sequence data from formalin-fixed and fresh-frozen tissue specimens. *Journal of Molecular Diagnostics*. **15**(5).

Spitznas, M., Joussen, F. and Wessing, A. 1976. Treatment of coats' disease with photocoagulation. *Albrecht v. Graefes Arch Klin Exp Ophthal*. **199**, pp.31–37.

Springer, T.A. 1998. An extracellular β-propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal growth factor precursor, and extracellular matrix components. *Journal of Molecular Biology*. **283**(4), pp.837–862.

Stamos, J.L. and Weis, W.I. 2013. The β-catenin destruction complex. *Cold Spring Harbor Perspectives in Biology*. **5**(1), pp.1–16.

Steffensen, A.Y., Dandanell, M., Jønson, L., Ejlertsen, B., Gerdes, A.M., Nielsen, F.C. and Hansen, T. v. O. 2014. Functional characterization of BRCA1 gene variants by mini-gene splicing assay. *European journal of human genetics : EJHG*. **22**(12), pp.1362–1368.

Steidl, S.M; Hirose, T.; Sang, D; Harnett, M.E. 1996. Difficulties in excluding the diagnosis of ret- inoblastoma in cases of advanced Coats' dis- ease: a clinicopathologic report. *Ophthalmologica*. (210), pp.336–340.

Stenkamp, D.L. 2015. Development of the Vertebrate Eye and Retina. *Prog Mol Biol Transl Sci.* **134**, pp.397–414.

Stipp, C.S., Kolesnikova, T. V. and Hemler, M.E. 2003. Functional domains in

tetraspanin proteins. *Trends in Biochemical Sciences*. **28**(2), pp.106–112.

Streeten, E.A., McBride, D., Puffenberger, E., Hoffman, M.E., Pollin, T.I., Donnelly, P., Sack, P. and Morton, H. 2008. Osteoporosis-pseudoglioma syndrome: Description of 9 new cases and beneficial response to bisphosphonates. *Bone*. **43**(3), pp.584–590.

Suzuki, T., Ohsumi, S. and Makino, K. 1994. Mechanistic studies on depurination and apurinic site chain breakage in oligodeoxyribonucleotides. *Nucleic Acids Research*. **22**(23), pp.4997–5003.

Tagami, M., Kusuhara, S., Honda, S., Tsukahara, Y. and Negi, A. 2008. Rapid regression of retinal hemorrhage and neovascularization in a case of familial exudative vitreoretinopathy treated with intravitreal bevacizumab. *Graefe's Archive for Clinical and Experimental Ophthalmology*. **246**(12), pp.1787–1789.

Tam, B.M. and Moritz, O.L. 2009. The Role of Rhodopsin Glycosylation in Protein Folding, Trafficking, and Light-Sensitive Retinal Degeneration. *Journal of Neuroscience*. **29**(48), pp.15145–15154.

Tamai, K., Semenov, M., Kato, Y., Spokony, R., Liu, C., Katsuyama, Y., Hess, F., Saint-Jeannet, J. and He, X. 2000. LDL receptor-related proteins in Wnt signal transduction. *Nature*. **407**(407), pp.530–535.

Tamai, K., Zeng, X., Liu, C., Zhang, X., Harada, Y., Chang, Z. and He, X. 2004. A Mechanism for Wnt Coreceptor Activation. *Molecular Cell*. **13**(1), pp.149–156.

Tan, R., Wang, Y., Kleinstein, S.E., Liu, Y., Zhu, X., Guo, H., Jiang, Q., Allen, A.S. and Zhu, M. 2014. An Evaluation of Copy Number Variation Detection Tools from Whole-Exome Sequencing Data. *Human Mutation*. **35**(7), pp.899–907.

Tang, M., Sun, L., Hu, A., Yuan, M., Yang, Y., Peng, X. and Ding, X. 2017. in Chinese Patients With Familial Exudative Vitreoretinopathy. *Retina*. (58), pp.5949–5957.

Tao, C. and Zhang, X. 2014. Development of astrocytes in the vertebrate eye. *Developmental Dynamics*. **243**(12), pp.1501–1510.

Tarkkanen, A. and Laatikainen, L. 1983. Coat's disease: clinical, angiographic, histopathological findings and clinical management. *The British journal of ophthalmology*. **67**(11), pp.766–76.

Thanos, A., Todorich, B. and Trese, M.T. 2016. A Novel Approach to Understanding Pathogenesis and Treatment of Capillary Dropout in Retinal Vascular Diseases. *Ophthalmic Surgery, Lasers and Imaging Retina*. **47**(3), pp.288–292.

Toomes, C., Downey, L.M., Bottomley, H.M., Scott, S., Woodruff, G., Trembath, R.C. and Inglehearn, C.F. 2004a. Identification of a fourth locus (EVR4) for familial exudative vitreoretinopathy (FEVR). *Molecular vision*. **10**, pp.37–42.

Toomes, C., Bottomley, H.M., Jackson, R.M., Towns, K. V, Scott, S., Mackey, D. a, Craig, J.E., Jiang, L., Yang, Z., Trembath, R., Woodruff, G., Gregory-Evans, C.Y., Gregory-Evans, K., Parker, M.J., Black, G.C.M., Downey, L.M., Zhang, K. and Inglehearn, C.F. 2004b. Mutations in LRP5 or FZD4 underlie the common familial exudative vitreoretinopathy locus on chromosome 11q. *American journal of human genetics*. **74**(4), pp.721–730.

Toomes, C., Bottomley, H.M., Scott, S., Mackey, D.A., Craig, J.E., Appukuttan, B., Stout, J.T., Flaxel, C.J., Zhang, K., Black, G.C.M., Fryer, A., Downey, L.M. and Inglehearn, C.F. 2004c. Spectrum and frequency of FZD4 mutations in familial exudative vitreoretinopathy. *Investigative Ophthalmology and Visual Science*. **45**(7), pp.2083–2090.

Toomes, C. and Downey, L. 2005. Familial exudative vitreoretinopathy, autosomal dominant. , Last revision: September, 2011.

Toomes, C., Downey, L.M., Bottomley, H.M., Mintz-Hittner, H.A. and Inglehearn, C.F. 2005. Further evidence of genetic heterogeneity in familial exudative vitreoretinopathy; exclusion of EVR1, EVR3, and EVR4 in a large autosomal dominant pedigree. *British Journal of Ophthalmology*. **89**(2), pp.194–197.

Trkova, M., Krutilkova, V., Smetanova, D., Becvarova, V., Hlavova, E., Jencikova, N., Hodacova, J., Hnykova, L., Hroncova, H., Horacek, J. and Stejskal, D. 2015. ISPD gene homozygous deletion identified by SNP array confirms prenatal manifestation of Walker-Warburg syndrome. *European Journal of Medical Genetics*. **58**(8), pp.372–375.

Turashvili, G., Yang, W., Mckinney, S., Kalloger, S., Gale, N., Ng, Y., Chow, K., Bell, L., Lorette, J., Carrier, M., Luk, M., Aparicio, S., Huntsman, D. and Yip, S. 2012. Nucleic acid quantity and quality from paraffin blocks : Defining optimal fixation, processing and DNA/RNA extraction techniques. *Experimental and Molecular Pathology*. **92**(1), pp.33–43.

Uezato, A., Yamamoto, N., Iwayama, Y., Hiraoka, S., Hiraaki, E., Umino, A., Haramo, E., Umino, M., Yoshikawa, T. and Nishikawa, T. 2015. Reduced cortical expression of a newly identified splicing variant of the DLG1 gene in patients with early-onset schizophrenia. *Translational Psychiatry*. **5**(10), pp.e654-8.

Uezato, A., Yamamoto, N., Jitoku, D., Haramo, E., Hiraaki, E., Iwayama, Y., Toyota, T., Umino, M., Umino, A., Iwata, Y., Suzuki, K., Kikuchi, M., Hashimoto, T., Kanahara, N., Kurumaji, A., Yoshikawa, T. and Nishikawa, T. 2017. Genetic and molecular risk factors within the newly identified primate-specific exon of the SAP97/DLG1 gene in the 3q29 schizophrenia-associated locus. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics*. **174**(8), pp.798–807.

Valenta, T., Hausmann, G. and Basler, K. 2012. The many faces and functions of Î 2-catenin. *EMBO Journal*. **31**(12), pp.2714–2736.

VdAuwera, G. What is a GVCF and How is it Different from a ‘Regular’ VCF. *http://gatkforums.broadinstitute.org/gatk/discussion/4017/what-is-a-gvcf-and-how-is-it-different-from-a-regular-vcf.* accessed 31 October 2017.

Vervoort, R., Lennon, A., Bird, A.C., Tulloch, B., Axton, R., Miano, M.G., Meindl, A., Meitinger, T., Ciccodicola, A. and Wright, A.F. 2000. Mutational hot spot within a new RPGR exon in X-linked retinitis pigmentosa. *Nature Genetics*. **25**(4), pp.462–466.

Vesnaver, G., Chang, C.N., Eisenberg, M., Grollman, A.P. and Breslauer, K.J. 1989. Influence of abasic and anucleosidic sites on the stability, conformation, and melting behavior of a DNA duplex: correlations of thermodynamic and structural data. *Proceedings of the National Academy of Sciences of the United States of America*. **86**(10), pp.3614–8.

Villegas, V.M., Gold, A.S., Berrocal, A.M. and Murray, T.G. 2014. Advanced Coats’ disease treated with intravitreal bevacizumab combined with laser

vascular ablation. *Clinical Ophthalmology*. **8**, pp.973–976.
Walker, J.F. 1964. Formaldehyde (3rd ed.). New York NY: Reinhold Publishing Company. *American Chemical Society Monograph Series.*
Wang, Y., Rattner, A., Zhou, Y., Williams, J., Smallwood, P.M. and Nathans, J. 2012. Norrin/Frizzled4 signaling in retinal vascular development and blood brain barrier plasticity. *Cell*. **151**(6), pp.1332–1344.
Warburg, M. 1961. Norrie's Disease. *Acta Ophthalmologica*. **39**(5), pp.757–772.
Warburg, M. 1966. Norrie's disease. A congenital progressive oculo-acoustico-cerebral degeneration. *Acta Ophthalmol (Copenh)*. **89**, pp.1–47.
Warburg, O. 1928. Über die Klassifizierung tierischer Gewebe nach ihrem Stoffwechsel *In*: *Über die Katalytischen Wirkungen der Lebendigen Substanz*. *Springer*, pp.510–514.
Watanabe, T. and Raff, M.C. 1988. Retinal astrocytes are immigrants from the optic nerve. *Nature*. **332**(6167), pp.834–837.
Wechsler-Reya, R.J. and Barres, B.A. 1997. Retinal development: Communication helps you see the light. *Current Biology*. **7**(7), pp.433–436.
Wehrli, M., Dougan, S.T., Caldwell, K., O'Keefe, L., Schwartz, S., Valzel-Ohayon, D., Schejter, E., Tomlinson, A. and DiNardo, S. 2000. Arrow encodes an LDL-receptor-related protein essential for Wingless signalling. *Nature*. **407**(6803), pp.527–530.
Willer, T., Lee, H., Lommel, M., Yoshida-moriguchi, T., Bernabe, B.V. De, Venzke, D., Cirak, S., Schachter, H., Vajsar, J., Voit, T., Muntoni, F., Loder, A.S., Dobyns, W.B., Moore, S. a and Campbell, K.P. 2012. ISPD loss-of-function mutations disrupt dystroglycan O- mannosylation and cause Walker-Warburg syndrome. *Nature Genetics*. **44**(5), pp.575–580.
Willnow, T.E., Nykjaer, A. and Herz, J. 1999. Lipoprotein receptors: new roles for ancient proteins. *Nature cell biology*. **1**(6), pp.E157-62.
Wong, S.Q., Li, J., Salemi, R., Sheppard, K.E., Hongdo, D., Tothill, R.W., McArthur, G.A. and Dobrovic, A. 2013. Targeted-capture massively-parallel sequencing enables robust detection of clinically informative mutations from formalin-fixed tumours. *Scientific Reports*. **3**(3), pp.1–10.
Wong, S.Q., Li, J., Y-C Tan, A., Vedururu, R., Pang, J.-M.B., Do, H., Ellul, J., Doig, K., Bell, A., MacArthur, G.A., Fox, S.B., Thomas, D.M., Fellowes, A., Parisot, J.P. and Dobrovic, A. 2014. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Medical Genomics*. **7**(1), pp.1–10.
Wu, J.H., Liu, J.H., Ko, Y.C., Wang, C.T.C.T., Chung, Y.C., Chu, K.C., Liu, T.T., Chao, H.M., Jiang, Y.J., Chen, S.J. and Chung, M.Y. 2016. Haploinsufficiency of RCBTB1 is associated with Coats disease and familial exudative vitreoretinopathy. *Human Molecular Genetics*. **25**(8), pp.1637–1647.
Xia, C.H., Liu, H., Cheung, D., Wang, M., Cheng, C., Du, X., Chang, B., Beutler, B. and Gong, X. 2008. A model for familial exudative vitreoretinopathy caused by LPR5 mutations. *Human Molecular Genetics*. **17**(11), pp.1605–1612.
Xia, C. hong, Yablonka-Reuveni, Z. and Gong, X. 2010. LRP5 is required for

vascular development in deeper layers of the retina. *PLoS ONE*. **5**(7), pp.1–7.

Xia, C. hong, Lu, E., Zeng, J. and Gong, X. 2013. Deletion of LRP5 in VLDLR Knockout Mice Inhibits Retinal Neovascularization. *PLoS ONE*. **8**(9), pp.1–11.

Xu, Q., Wang, Y., Dabdoub, A., Smallwood, P.M., Williams, J., Woods, C., Kelley, M.W., Jiang, L., Tasman, W., Zhang, K. and Nathans, J. 2004. Vascular Development in the Retina and Inner Ear. *Cell*. **116**(6), pp.883–895.

Yamane, T., Yokoi, T., Nakayama, Y., Nishina, S. and Azuma, N. 2014. Surgical outcomes of progressive tractional retinal detachment associated with familial exudative vitreoretinopathy. *American Journal of Ophthalmology*. **158**(5), pp.1049–1055.

Yang, H., Li, S., Xiao, X., Wang, P., Guo, X. and Zhang, Q. 2012. Identification of FZD4 and LRP5 mutations in 11 of 49 families with familial exudative vitreoretinopathy. *Molecular vision*. **18**, pp.2438–46.

Ye, X., Wang, Y., Cahill, H., Yu, M., Badea, T.C., Smallwood, P.M., Peachey, N.S. and Nathans, J. 2009. Norrin, Frizzled-4, and Lrp5 Signaling in Endothelial Cells Controls a Genetic Program for Retinal Vascularization. *Cell*. **139**(2), pp.285–298.

Ye, X., Wang, Y. and Nathans, J. 2010. The Norrin/Frizzled4 signaling pathway in retinal vascular development and disease. *Trends Mol Medicine*. **16**(9), pp.417–425.

Yoshikawa, Y., Yamada, T., Nagara, I.T., Okabe, K., Kitagawa, Y., Ema, M. and Kubota, Y. 2016. Developmental regression of hyaloid vasculature is triggered by neurons. *The Journal of Experimental Medicine*. **213**(7), pp.1175–1183.

Yoshizumi, M.O., Kreiger, A.E., Lewis, H., Foxman, B. and Hakakha, B.A. 1995. Vitrectomy techniques in late-stage Coats'-like exudative retinal detachment. *Documenta Ophthalmologica*. **90**(4), pp.387–394.

Yost, S.E., Smith, E.N., Schwab, R.B., Bao, L., Jung, H., Wang, X., Voest, E., Pierce, J.P., Messer, K., Parker, B.A., Harismendy, O. and Frazer, K.A. 2012. Identification of high-confidence somatic mutations in whole genome sequence of formalin-fixed breast cancer specimens. *Nucleic Acids Research*. **40**(14), pp.1–12.

Zhang, C.-L., Zou, Y., Yu, R.T., Gage, F.H. and Evans, R.M. 2006. Nuclear receptor TLX prevents retinal dystrophy and recruits the corepressor atrophin1. *Genes and Development*. **20**, pp.1308–1320.

Zhang, C., Lai, M.B., Khandan, L., Lee, L.A., Chen, Z. and Junge, H.J. 2017. Norrin-induced Frizzled4 endocytosis and endo-lysosomal trafficking control retinal angiogenesis and barrier function. *Nature Communications*. **8**, p.16050.

Zhang, K., Harada, Y., Wei, X., Shukla, D., Rajendran, A., Tawansy, K., Bedell, M., Lim, S., Shaw, P.X., He, X. and Yang, Z. 2011. An essential role of the cysteine-rich domain of FZD4 in Norrin/Wnt signaling and familial exudative vitreoretinopathy. *Journal of Biological Chemistry*. **286**(12), pp.10210–10215.

Zhang, L., Yang, Y., Li, S., Tai, Z., Huang, L., Liu, Y., Zhu, X., Di, Y., Qu, C., Jiang, Z., Li, Y., Zhang, G., Kim, R., Sundaresan, P., Yang, Z. and Zhu, X. 2016. Whole Exome Sequencing Analysis Identifies Mutations in LRP5

in Indian Families with Familial Exudative Vitreoretinopathy. *Genetic Testing and Molecular Biomarkers*. **20**(7), pp.1–6.

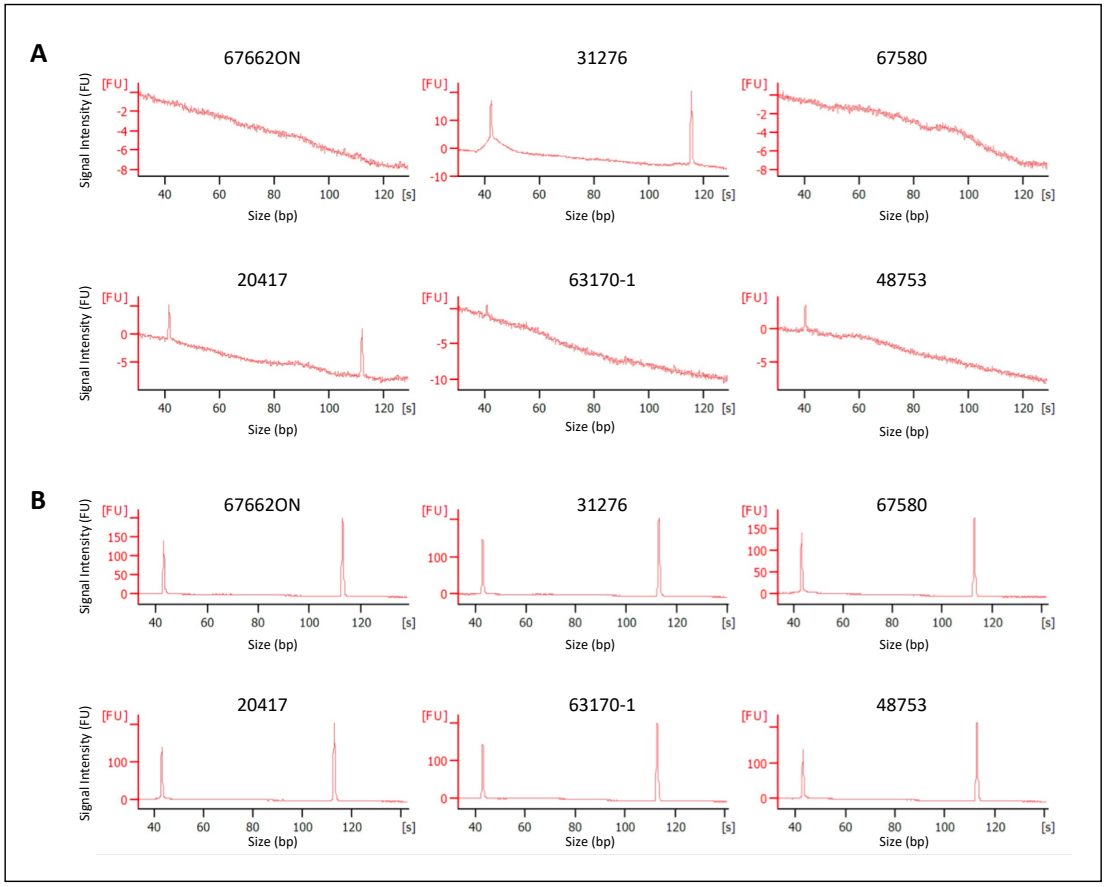
Zhou, R.M., Wang, X.Q., Yao, J., Shen, Y., Chen, S.N., Yang, H., Jiang, Q. and Yan, B. 2015. Identification and characterization of proliferative retinopathy-related long noncoding RNAs. *Biochemical and Biophysical Research Communications*. **465**(3), pp.324–330.

Zhou, Y., Wang, Y., Tischfield, M., Williams, J., Smallwood, P.M., Rattner, A., Taketo, M.M. and Nathans, J. 2014. Canonical WNT signaling components in vascular development and barrier formation. *The Journal of Clinical Investigation*. **124**(9), pp.3825–3846.

Zhu, M., Madigan, M.C., Van Driel, D., Maslim, J., Billson, F.A., Provis, J.M. and Penfold, P.L. 2000. The human hyaloid system: Cell death and vascular regression. *Experimental Eye Research*. **70**(6), pp.767–776.

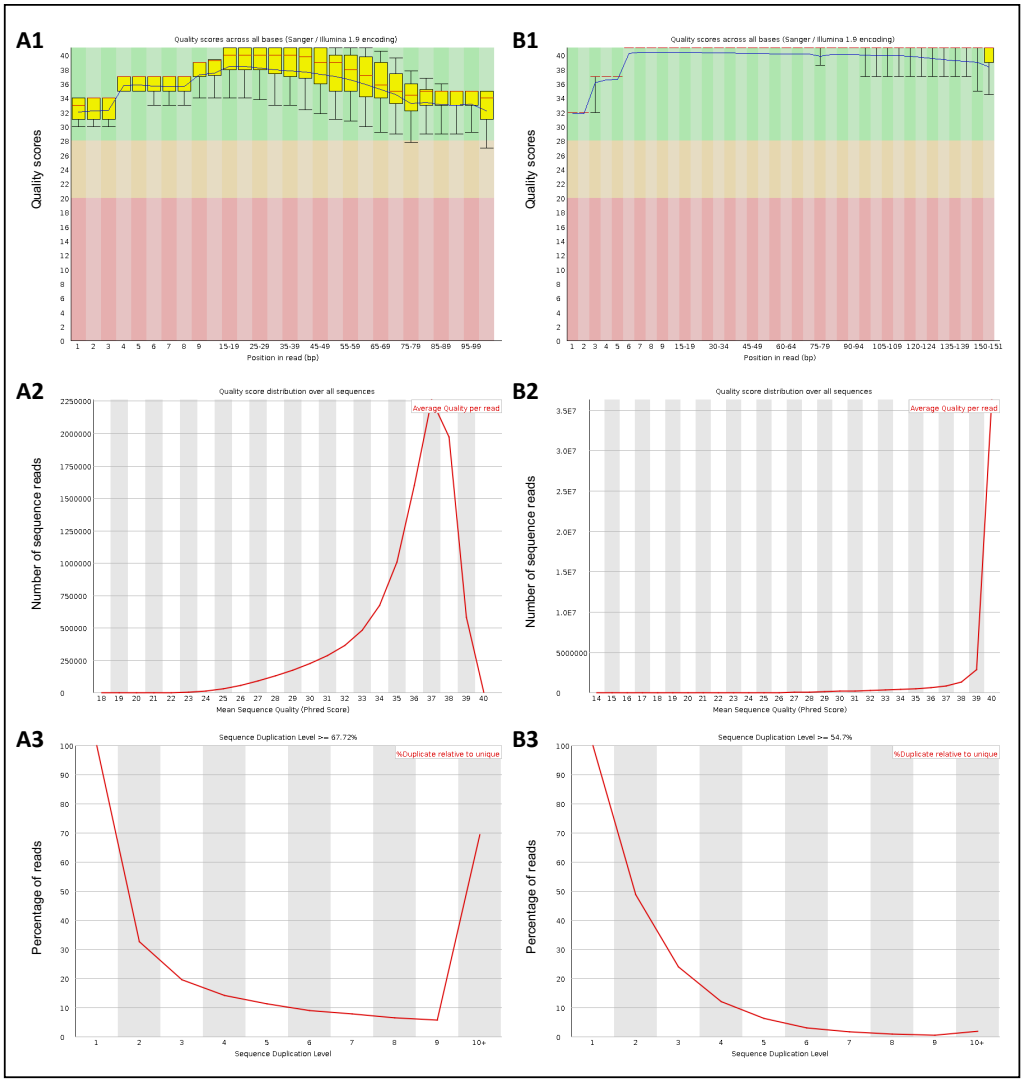
7 Chapter: Appendices

7.1 Bioanalyser traces for FFPE patients



Bioanalyser traces for the remainder of the Coats disease patients for the first round of WES prior to shearing and post-shearing before the library preparation.
(A) All samples are unsheared and mask the ladders and were not able to be quantified. **(B)** All samples are post shearing. X-axis shows the size in base pairs (bp) and the y-axis represents the signal intensity of detected DNA.

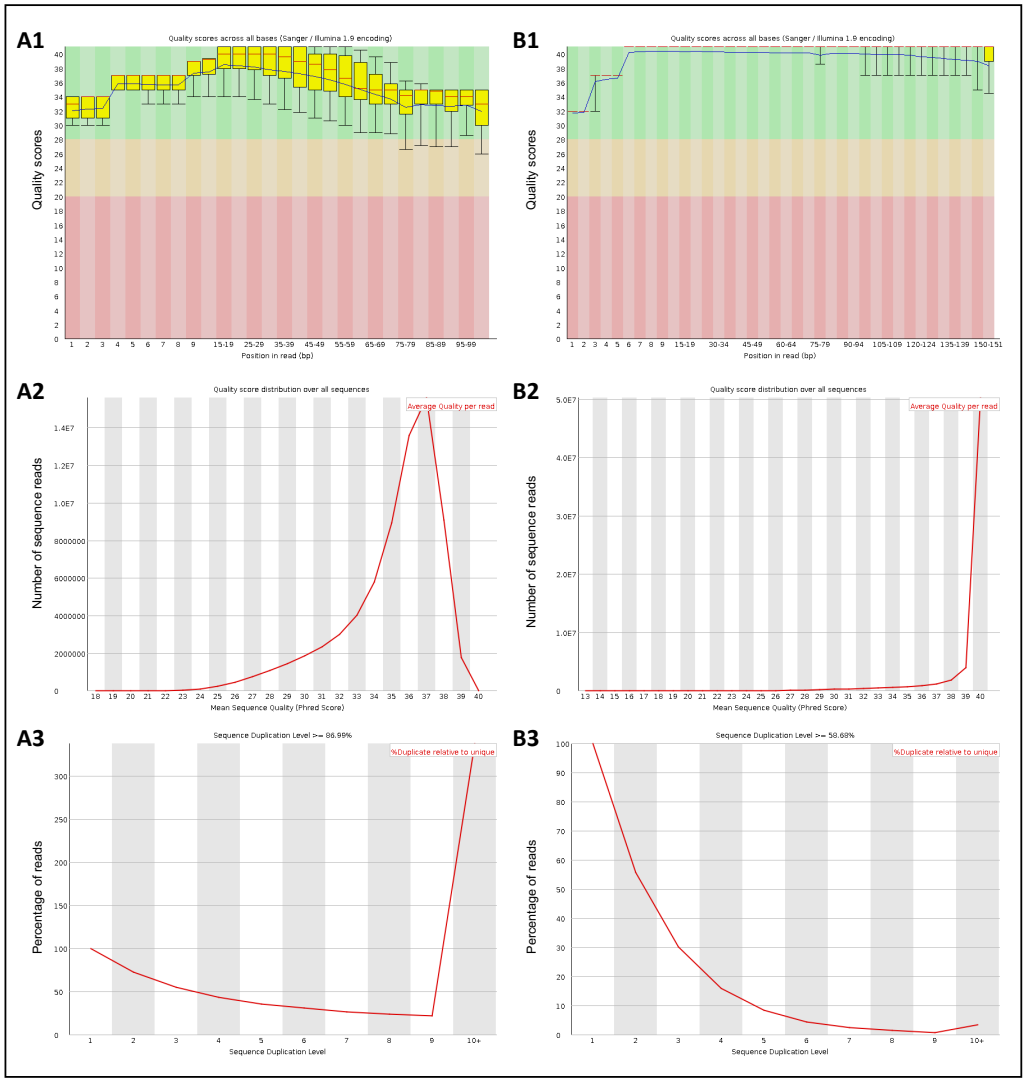
7.2 Quality of raw data QC checks for remainder of Coats disease patients



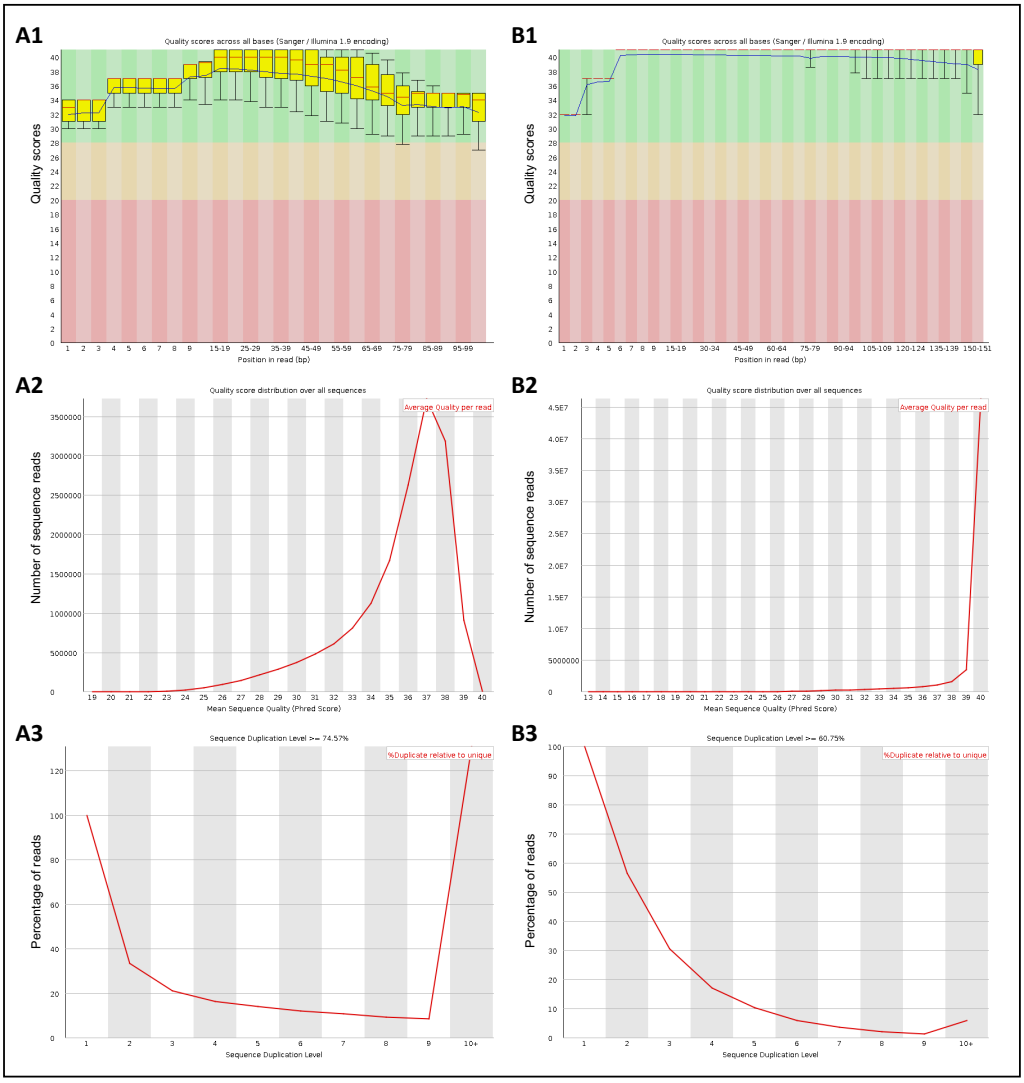
Comparison of the QC between first (A) and second (B) run of WES for patient 20417.

(A1) & (B1) A representative of the QC of the raw fastq file for patient 20417 using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 20417. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 67.7% of all sequences were duplicated between 2 to 10+ times with a big spike between 9-10 times. 32.3% of all sequences were only read once. However, from the second run of WES (B3), the

percentage of duplicates was improved to 54.7% between 2 to 10+ times. 45.3% of all sequences were only read once.

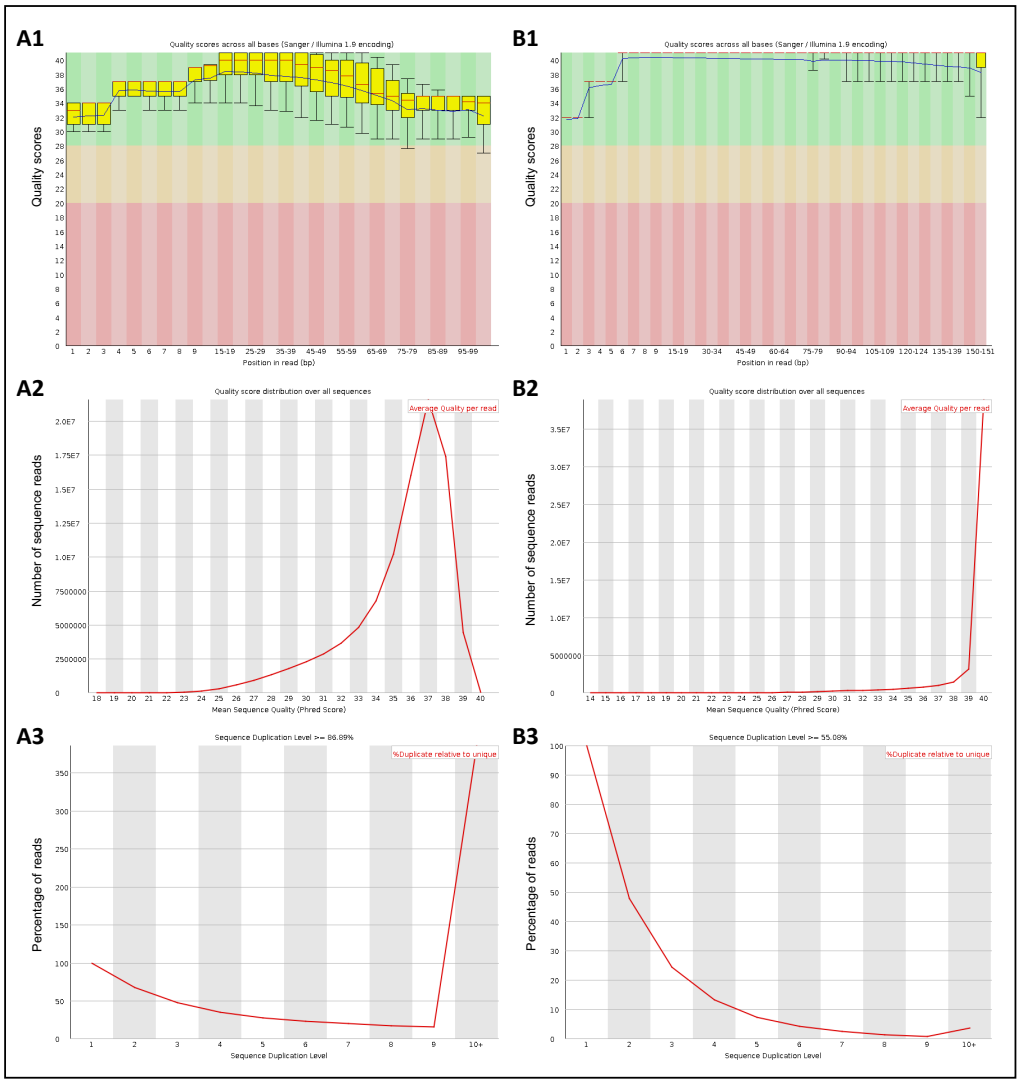


Comparison of the QC between first (A) and second (B) run of WES for patient 31276.
(A1) & (B1) Representative examples of the QC of the raw fastq file for patient 31276 using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 31276. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 86.99% of all sequences were duplicated between 2 to 10+ times with a big spike between 9-10 times. 13.01% of all sequences were only read once. However, from the second run of WES (B3), the percentage of duplicates was improved to 58.68% between 2 to 10+ times. 41.32% of all sequences were only read once.



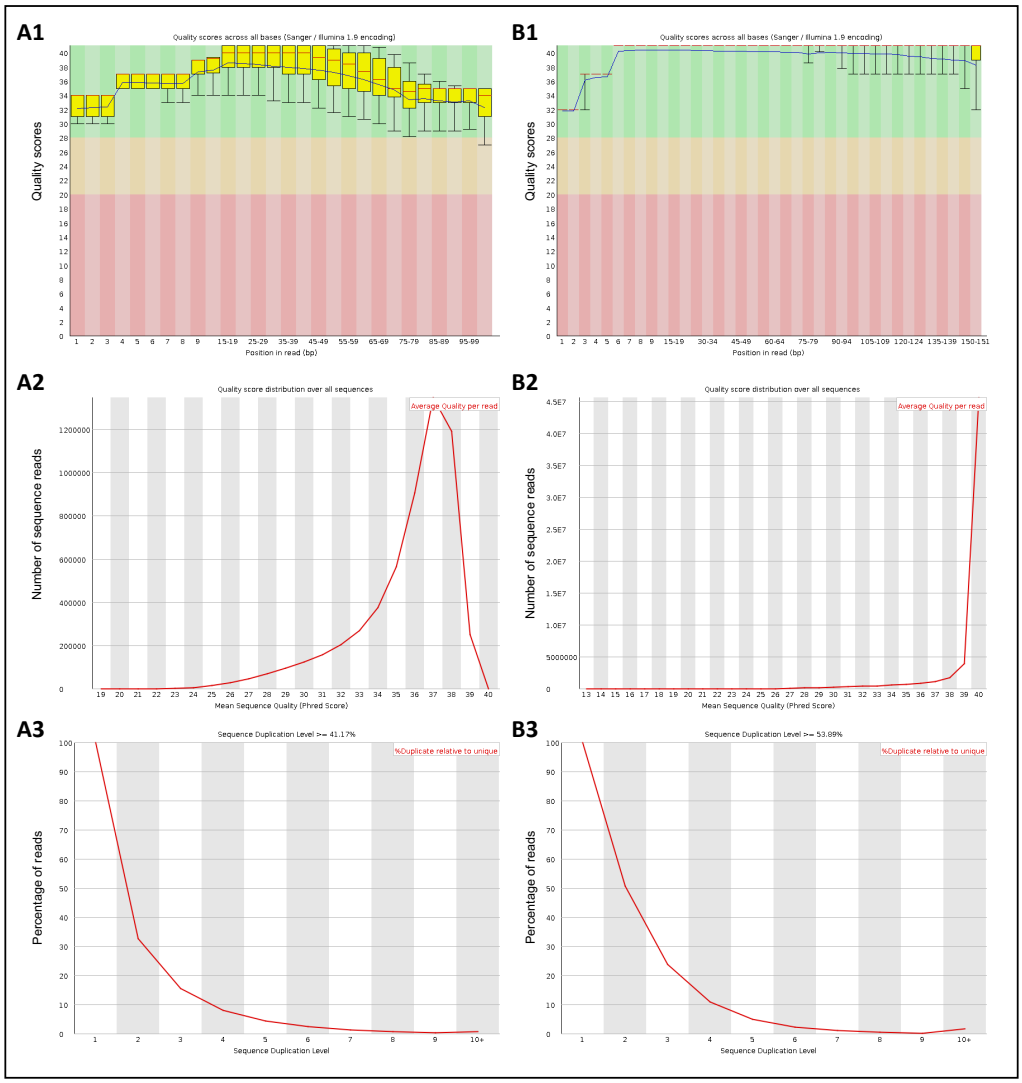
Comparison of the QC between first (A) and second (B) run of WES for patient 48753.

(A1) & (B1) Representative examples of the QC of the raw fastq file for patient 48753 using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 48753. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 74.57% of all sequences were duplicated between 2 to 10+ times with a big spike between 9-10 times. 25.33% of all sequences were only read once. However, from the second run of WES (B3), the percentage of duplicates was improved to 60.75% between 2 to 10+ times. 39.25% of all sequences were only read once.



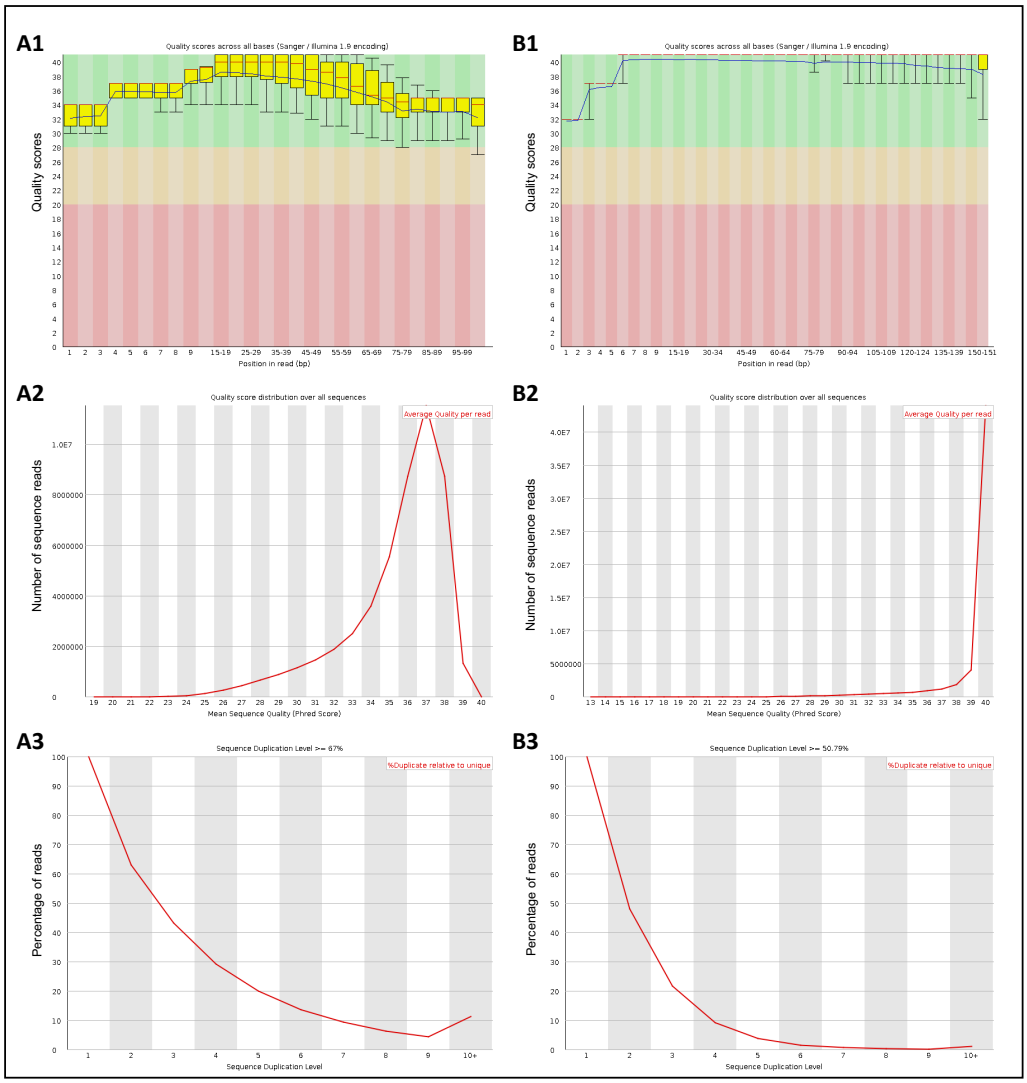
Comparison of the QC between first (A) and second (B) run of WES for patient 63170-1.

(A1) & (B1) Representative examples of the QC of the raw fastq file for patient 63170-1 using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 63170-1. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 86.89% of all sequences were duplicated between 2 to 10+ times with a big spike between 9-10 times. 13.11% of all sequences were only read once. However, from the second run of WES (B3), the percentage of duplicates was improved to 55.08% between 2 to 10+ times. 44.92% of all sequences were only read once.



Comparison of the QC between first (A) and second (B) run of WES for patient 67580.

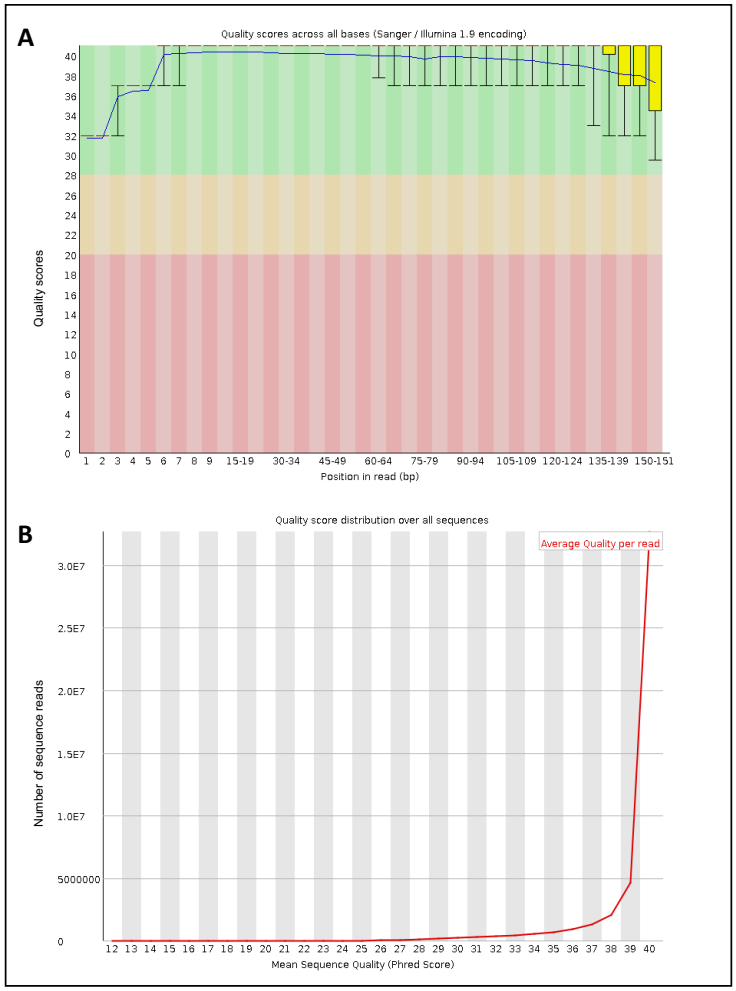
(A1) & (B1) Representative examples of the QC of the raw fastq file for patient 67580 using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 67580. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 41.17% of all sequences were duplicated between 2 to 10+ times. 57.83% of all sequences were only read once. However, from the second run of WES (B3), the percentage of duplicates was improved to 53.89% between 2 to 10+ times. 46.11% of all sequences were only read once.



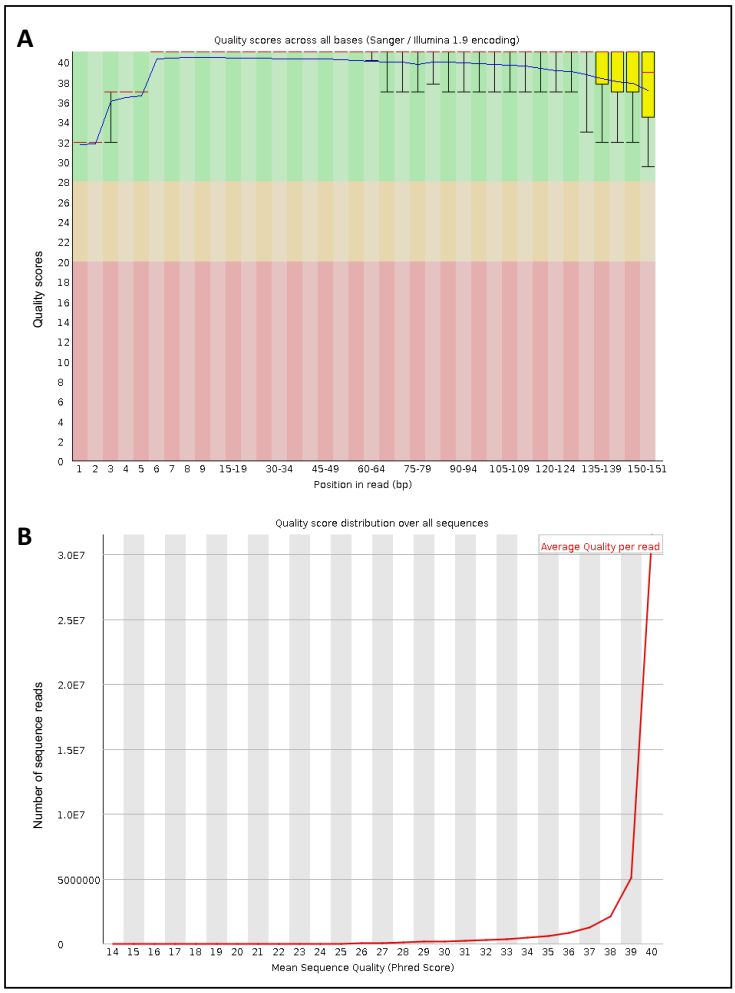
Comparison of the QC between first (A) and second (B) run of WES for patient 67662ON.

(A1) & (B1) Representative examples of the QC of the raw fastq file for patient 67662ON using a BoxWhisker plot. **(A2) & (B2)** Representative example showing the quality score per read distribution over sequences of the fastq file for sample 67662ON. The increase of data quality can be seen between the two runs of WES for the same Coats patient. B1 image shows increased quality in comparison to A1 based on the quality scores, where the Phred score is noted to be above 40. The quality score distribution was also shown to have increased in quality for the second run of WES (B2) in comparison to the first (A2). The average quality per read was noted at 40 Phred score for the second run in comparison to 37 from the first WES. **(A3) & (B3)** Representative example of sequence duplication level. Percentage of reads is displayed on the y-axis whilst the sequence duplication level is on the x-axis. The results for first run of WES (A3) showed that 67% of all sequences were duplicated between 2 to 10+ times with a smaller spike between 9-10 times. 33% of all sequences were only read once. However, from the second run of WES (B3), the percentage of duplicates was improved to 50.79% between 2 to 10+ times. 49.21% of all sequences were only read once.

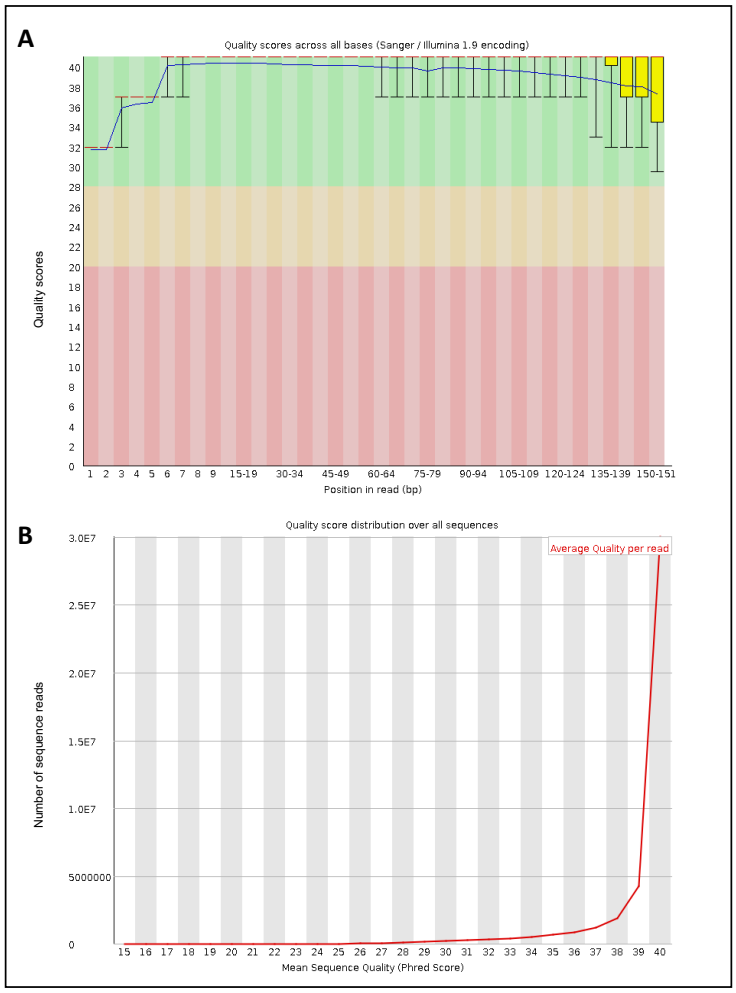
7.3 Quality of raw data QC checks for FEVR patients with identified mutations



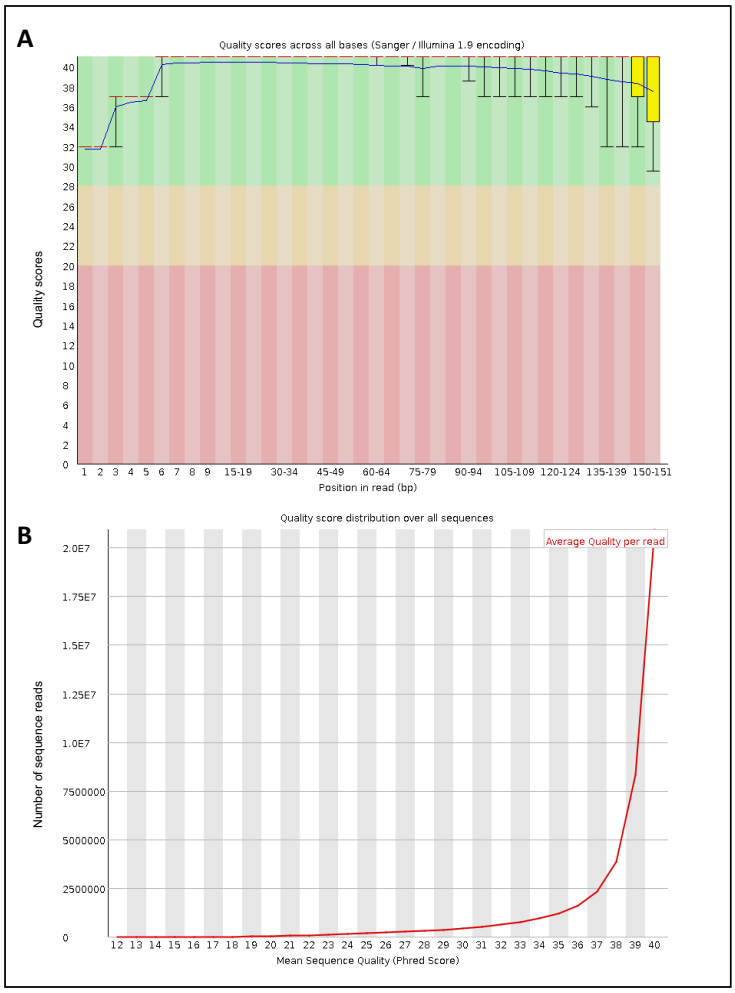
BoxWhisker plot of the quality control (QC) for the first read val_1 for patient F1057_1072. (A) A representative of the QC of the raw fastq file for F1057_1072 using a BoxWhisker plot. Quality scores across 150bp read generated by Illumina HiSeq 2500 throughout the WES run. The x-axis represents the position in read depth in bp and the y-axis shows the quality scores. The scores below 20 of the BoxWhisker plot represents poor quality calls, the scores below 28 represent calls of reasonable quality and the scores below 40 show very good quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality. **(B)** Representative of quality score distribution over sequences of the fastq file for a patient F057_1072. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads.



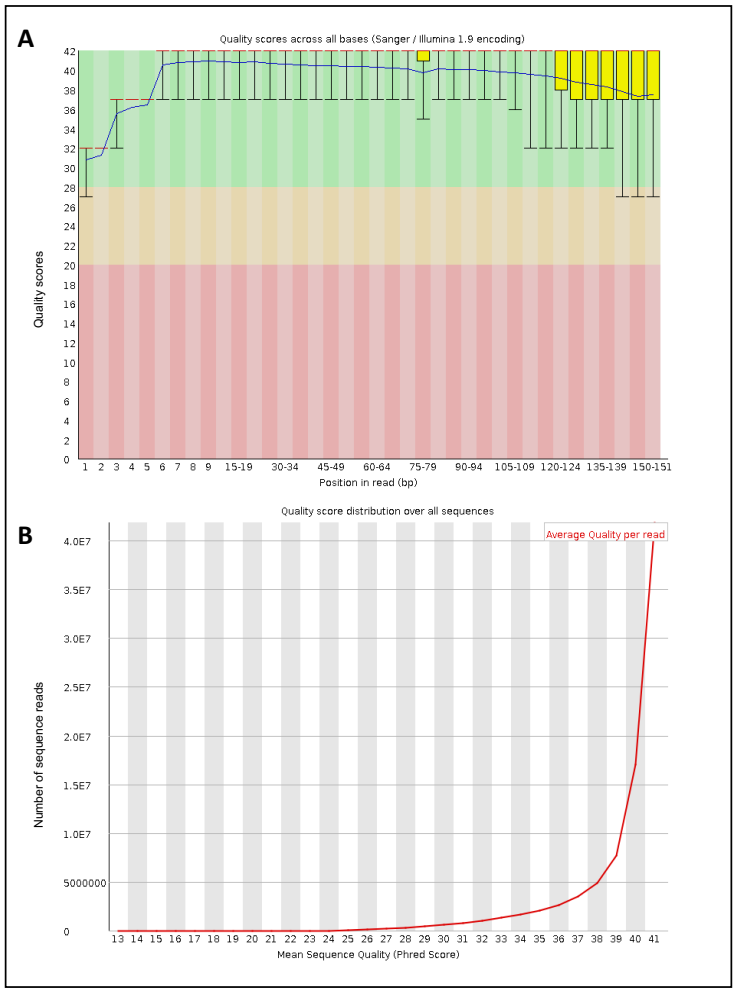
BoxWhisker plot of the quality control (QC) for the first read val_1 for patient F1375. (A) A representative of the QC of the raw fastq file for F1375 using a BoxWhisker plot. Quality scores across 150bp read generated by Illumina HiSeq 2500 throughout the WES run. The x-axis represents the position in read depth in bp and the y-axis shows the quality scores. The scores below 20 of the BoxWhisker plot represents poor quality calls, the scores below 28 represent calls of reasonable quality and the scores below 40 show very good quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality. **(B)** Representative of quality score distribution over sequences of the fastq file for a patient F1375. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads.



BoxWhisker plot of the quality control (QC) for the first read val_1 for patient F164. (A) A representative of the QC of the raw fastq file for F164 using a BoxWhisker plot. Quality scores across 150bp read generated by Illumina HiSeq 2500 throughout the WES run. The x-axis represents the position in read depth in bp and the y-axis shows the quality scores. The scores below 20 of the BoxWhisker plot represents poor quality calls, the scores below 28 represent calls of reasonable quality and the scores below 40 show very good quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality. **(B)** Representative of quality score distribution over sequences of the fastq file for a patient F164. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads.



BoxWhisker plot of the quality control (QC) for the first read val_1 for patient F1308. (A) A representative of the QC of the raw fastq file for F1308 using a BoxWhisker plot. Quality scores across 150bp read generated by Illumina HiSeq 2500 throughout the WES run. The x-axis represents the position in read depth in bp and the y-axis shows the quality scores. The scores below 20 of the BoxWhisker plot represents poor quality calls, the scores below 28 represent calls of reasonable quality and the scores below 40 show very good quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality. **(B)** Representative of quality score distribution over sequences of the fastq file for a patient F1308. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads.



BoxWhisker plot of the quality control (QC) for the first read val_1 for patient F561. (A) A representative of the QC of the raw fastq file for F561 using a BoxWhisker plot. Quality scores across 150bp read generated by Illumina HiSeq 2500 throughout the WES run. The x-axis represents the position in read depth in bp and the y-axis shows the quality scores. The scores below 20 of the BoxWhisker plot represents poor quality calls, the scores below 28 represent calls of reasonable quality and the scores below 40 show very good quality calls. The yellow boxes show the inter-quartile range (25-75%), whereas the whiskers on either side show the 10% and 90% range. The red line across the yellow boxes is the median and the blue continuous line represents the mean quality. **(B)** Representative of quality score distribution over sequences of the fastq file for a patient F561. The x-axis shows the mean sequence quality and the y-axis represents the total number of sequence reads.

7.4 Primer sequences to amplify *LRP5* c.2951A>G variant

LRP5 Y984 – F (120bp)

5’ CACAGCCCGGATCTCATC 3’

LRP5 Y984 – R (120bp)

5’ CTTGGCTCGCTTGATGTTCT 3’

LRP5 Y984 – 2F (170bp)

5’CACCACCTTCTTGCTGTTCA 3’

LRP5 Y984 – 2R (170bp)

5’ TGATGTTCTGGCG.

7.5 Command for the chromosomal location of genes

$ perl /home/vcfhacks/getVariantsByLocation.pl -i input.vcf -o output.vcf -b FEVR_genes.bed

*FEVR_genes.bed file is a file made using the Table Browser tool of the UCSC Genome Browser using the coordinates for all exons (+/-) of the genes of interest: *LRP5, FZD4, NDP, TSPAN12, ATOH7, ZNF408, KIF11.*

7.6 Command lines for 1st and 2nd round of WES for Coats disease patients

- Trimming of adaptors and run FastQC

$ trim_galore -q 20 --fastqc_args "--outdir /data/medddz/Patient_20417" --illumina --gzip -o /data/medddz/CutAdaptor/ --length 20 --paired /data/medddz/Patient_20417/20417_S27_L003_R1_001.fastq.gz /data/medddz/Patient_20417/20417_S27_L003_R2_001.fastq.gz

- Alignment

$ bwa mem -t 12 -M /home/ref/b37/human_g1k_v37.fasta
/data/medddz/CutAdaptor/20417_S27_L003_R1_001_val_1.fq.gz
/data/medddz/CutAdaptor/20417_S27_L003_R2_001_val_2.fq.gz
-v 1
-R
'@RG\tID:Coats\tSM:20417_WES\tPL:Illumina\tPU:HiSeq3000\tLB:20417_WES_Coats'
-M | samtools view -Sb - > /data/medddz/Align/20417_WES.bwamem.bam

- Sort SAM file

$ java -Xmx8g -jar /home/picard/picard-tools-1.129/picard.jar SortSam
I=/data/medddz/Align/20417_WES.bwamem.bam
O=/data/medddz/Sort/20417_WES.bwamem.sort.bam
SO=coordinate CREATE_INDEX=TRUE

- Mark duplicates

```
$ java -Xmx8g -jar /home/picard/picard-tools-1.129/picard.jar MarkDuplicates
I=/data/medddz/Sort/20417_WES.bwamem.sort.bam
O=/data/medddz/MarkDup/20417_WES.bwamem.sort.dedup.bam
M=/data/medddz/MarkDup/20417_WES.bwamem.sort.dedup.metrics
CREATE_INDEX=TRUE
```

- Create indel realigner targets

```
$ java -Xmx6g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T RealignerTargetCreator
-R /home/ref/b37/human_g1k_v37.fasta
-known /home/ref/b37/1000G_phase1.indels.b37.vcf
-known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf
-I /data/medddz/MarkDup/20417_WES.bwamem.sort.dedup.bam
-o /data/medddz/Realigner/20417_WES.bwamem.sort.dedup.bam.intervals
```

- Perform indel realignment

```
$ java -Xmx6g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T IndelRealigner
-R /home/ref/b37/human_g1k_v37.fasta
-known /home/ref/b37/1000G_phase1.indels.b37.vcf
-known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf
-I /data/medddz/MarkDup/20417_WES.bwamem.sort.dedup.bam
-targetIntervals
/data/medddz/Realigner/20417_WES.bwamem.sort.dedup.bam.intervals
-o
/data/medddz/IndelRealigner/20417_WES.bwamem.sort.dedup.indelrealn.ba
m
```

- Get base recalibration model

```
$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T BaseRecalibrator
-R /home/ref/b37/human_g1k_v37.fasta
-knownSites /home/ref/b37/dbSnp146.b37.vcf.gz
-knownSites /home/ref/b37/1000G_phase1.indels.b37.vcf
-knownSites
/home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf
-I
/data/medddz/IndelRealigner/20417_WES.bwamem.sort.dedup.indelrealn.ba
m
-o
/data/medddz/BaseRecal/20417_WES.bwamem.sort.dedup.indelrealn.recal.
grp
-nct 6
```

- Apply recalibration (print reads)

```
$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T PrintReads \
-R /home/ref/b37/human_g1k_v37.fasta
-I
/data/medddz/IndelRealigner/20417_WES.bwamem.sort.dedup.indelrealn.ba
m
-BQSR
/data/medddz/BaseRecal/20417_WES.bwamem.sort.dedup.indelrealn.recal.
grp
-o
/data/medddz/FinalBam/20417_WES.bwamem.sort.dedup.indelrealn.recal.ba
m
--disable_indel_quals
```

- Create genomic VCF (gVCF) and call all variants

```
$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T HaplotypeCaller --emitRefConfidence GVCF
-R /home/ref/b37/human_g1k_v37.fasta -D
/home/ref/b37/dbSnp146.b37.vcf.gz
-stand_call_conf 30 -stand_emit_conf 10
-I
/data/medddz/FinalBam/20417_WES.bwamem.sort.dedup.indelrealn.recal.ba
m
-o /data/medddz/gVCF/20417_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-L /home/ref/ SureSelectAllExonV5/S04380110 _Regions_b37.bed -ip 30
```

And for the 2nd WES

```
-L /home/ref/SureSelectAllExonV6/S07604514_Regions_b37.bed -ip 30
```

- Combine gVCF

```
$ java -Xmx18g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T GenotypeGVCFs -R /home/ref/b37/human_g1k_v37.fasta
-D /home/ref/b37/dbSnp146.b37.vcf.gz -stand_call_conf 30
-stand_emit_conf 10
-V /data/medddz/gVCF/20417_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/gVCF/31276_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/gVCF/48753_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/gVCF/63170_1_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/gVCF/67580_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-V
/data/medddz/gVCF/67662_ON_WES.sort.dedup.indelrealn.recal.HC.g.vcf
```

-V /data/medddz/gVCF/71208_M1_WES.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/control.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/control.sort.dedup.indelrealn.recal.HC.g.vcf
-o /data/medddz/OnegVCF/all.genotype.vcf -nda --showFullBamList -nt 8

+ a further of 55 HC.g.vcfs as controls for 2nd WES

- Select variants (SNPs)

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar
-T SelectVariants
-R /home/ref/b37/human_g1k_v37.fasta
-selectType SNP
--variant /data/medddz/OnegVCF/all.genotype.vcf
-o /data/medddz/SelectSNPVariants/all.genotype.raw-snps.vcf

- Select Variants (INDELs)

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar
-T SelectVariants
-R /home/ref/b37/human_g1k_v37.fasta
--variant /data/medddz/OnegVCF/all.genotype.vcf
-selectType INDEL -selectType MNP
-o /data/medddz/SelectINDELVariants/all.genotype.raw-indels.vcf

- Variant filtration for SNPs and INDELs separately

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar
-T VariantFiltration
-R /home/ref/b37/human_g1k_v37.fasta
-V /data/medddz/SelectSNPVariants/all.genotype.raw-snps.vcf
--filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 ||
MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0"
--filterName "snp_hard_filter"
-o /data/medddz/VariantFiltrationSNP/all.genotype.filtered.snps.vcf

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar
-T VariantFiltration
-R /home/ref/b37/human_g1k_v37.fasta
-V /data/medddz/SelectINDELVariants/all.genotype.raw-indels.vcf
--filterExpression "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0"
--filterName "indel_hard_filter"
-o /data/medddz/VariantFiltrationINDEL/all.genotype.filtered.indels.vcf

- Combine SNPs and INDELs

```
$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T CombineVariants -R /home/ref/b37/human_g1k_v37.fasta
--variant /data/medddz/VariantFiltrationSNP/all.genotype.filtered.snps.vcf
--variant /data/medddz/VariantFiltrationINDEL/all.genotype.filtered.indels.vcf
-o /data/medddz/Combine/combined.genotype.filtered.combinedvar.vcf
--genotypemergeoption UNSORTED
```

- Annotate to dbSNP (≥1%)

```
$ perl /home/vcfhacks-v0.2.0/annotateSnps.pl
-d /home/ref/b37/dbSnp142.b37.vcf.gz
/home/ref/b37/clinvar_20150330.vcf.gz -f 0.01 -pathogenic
-i /data/medddz/Combine/combined.genotype.filtered.combinedvar.vcf
-o
/data/medddz/AnnotatedbSNP/combined.0.01.filtered.combinedvar.1pcdbsnp
.vcf -t 6
```

- Filter variants in EVS (≥1%)

```
$ perl /home/vcfhacks/filterOnEvsMaf.pl -d /home/ref/evs/ -f 1 --progress -i
fltd-combinedvars.split.1pcdbsnp.vcf -o fltd-
combinedvars.split.1pcdbsnp.1pcEVS.vcf
```

*EVS carried out for the 1st WES

- Filter on ExAC (≥1% or ≥0.01%)

```
$ perl /home/vcfhacks-v0.2.0/filterVcfOnVcf.pl
-i
/data/medddz/AnnotatedbSNP/combined.0.01.filtered.combinedvar.1pcdbsnp
.vcf
-f /home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz
-o
/data/medddz/ExAC/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vcf -
w -y 0.01 or 0.0001
```

*ExAC carried out for 2nd WES with 57 controls

- Variant effect predictor (VEP)

```
$ perl /home/variant_effect_predictor/variant_effect_predictor.pl
--offline --vcf --everything
--dir_cache /home/variant_effect_predictor/vep_cache
--dir_plugins /home/variant_effect_predictor/vep_cache/Plugins
--plugin
Condel,/home/variant_effect_predictor/vep_cache/Plugins/config/Condel/conf
ig/
--plugin ExAC,/home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz
--plugin SpliceConsensus
```

```
--fasta
/home/variant_effect_predictor/fasta/Homo_sapiens.GRCh38.dna.primary_as
sembly.fa.gz
-i
/data/medddz/ExAC/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vcf
-o
/data/medddz/VEP/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vep.v
cf
--fork 6
```

- Filter on sample

```
$ perl /home/vcfhacks-v0.2.0/filterOnSample.pl
-i /project_folder/geno/genotype.fltd-combinedvars.1pcdbsnp.1pcEVS.vep.vcf
-s Sample_ID_SM -x (or -n2)
-o /project_folder/filtered/Sample_ID.vcf
```

*x = only keep variants if present in all samples

*n2, n3, n4, n5, n6 = only keep variants if present in x number of samples

FILTERS:

- Get Functional variants

```
$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl \
-i
/data/medddz/VEP/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vep.v
cf
-s all -n1 (or -n2)
-o
/data/medddz/GetFunctional/combined.0.01.gene.filtered.combinedvar.1pcdb
snp.exac.getfunctional.vep.vcf
```

*n1 = only keep variants if present in the 1 sample

*n2, n3, n4, n5, n6, n7 = only keep variants if present in x number of samples

- Get Functional variants (based on shared genes with any variants)

```
$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl \
-i
/data/medddz/VEP/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vep.v
cf
-s all -f -n1 (or -n2)
-o
/data/medddz/GetFunctional/combined.0.01.gene.filtered.combinedvar.1pcdb
snp.exac.getfunctional.vep.vcf
```

*-f = find shared genes, output that makes up functional variants in the same genes for the specified samples, will also return list of all functional variants

- Find Bi-allelic variants

$ perl /home/vcfhacks-v0.2.0/findBiallelic.pl
-i /project_folder/filtered/Sample_ID.vcf
-s Sample_ID (-x or -r)
-o /project_folder/filtered/Sample_ID.findBiallelic.vcf

*-x = reject variants present in all samples except these

*-r = samples to exclude variants from

- Get Heterozygous variants

$ perl /home/vcfhacks-v0.2.0/getHetVariants.pl -i
/project_folder/filtered/Sample_ID.vcf
-s Sample_ID -o /project_folder/filtered/Sample_ID.getHetVariants.vcf

- Get Homozygous variants (getHetVariants.pl but reversed)

$ perl /home/vcfhacks-v0.2.0/getHetVariants.pl -i
/project_folder/filtered/Sample_ID.vcf
-s Sample_ID
-r -o /project_folder/filtered/Sample_ID.getHetVariants.Hom.vcf

*-r = reverse the command and look for homozygous variants only

- Obtain CADD scores

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl
-c /data/shared/cadd/v1.2/*.gz
-i
/data/medddz/GetFunctional/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.exac.getfunctional.vep.vcf
-o
/data/medddz/CADD/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.exac.vep.getfunctional.not_cadd_ranked.vcf -n cadd_not_found.tvs -d --progress

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl
-d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb
-i
/data/medddz/CADD/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.exac.vep.getfunctional.not_cadd_ranked.vcf
-o
/data/medddz/GeneAnnotator/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.exac.vep.getfuntional.not_cadd_ranked.gene_anno

- Convert to Excel

```
$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl
-i
/data/medddz/GeneAnnotator/combined.0.01.gene.filtered.combinedvar.1pcd
bsnp.exac.vep.getfuntional.not_cadd_ranked.gene_anno
--vep --gene_anno
-o
/data/medddz/Excel/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.exa
c.vep.getfuntional.not_cadd_ranked.gene_anno.simple.xlsx
```

7.7 Command line for FEVR WES

- Trimming of adaptors and run FastQC

```
$ trim_galore -q 20 --fastqc_args "--outdir /data/medddz/Patient_20417" --
illumina --gzip -o /data/medddz/CutAdaptor/ --length 20 --paired
/data/medddz/Patient_ID/Patient_ID_R1.fastq.gz
/data/medddz/Patient_ID/Patient_ID_R2.fastq.gz
```

- Alignment

```
$ bwa mem -t 12 -M /home/ref/b37/human_g1k_v37.fasta
/data/medddz/Cut_Adaptor/Patient_ID_R1.fastq.gz
/data/medddz/Cut_Adaptor/Patient_ID_R2.fastq.gz -v 1
-R
'@RG\tID:FEVR\tSM:Patient_ID\tPL:Illumina\tPU:HiSeq3000\tLB:Patient_ID'
-M | samtools view -Sb - > /data/medddz/Align/Patient_ID.bwamem.bam
```

- Sort SAM file

```
$ java -Xmx8g -jar /home/picard/picard-tools-1.129/picard.jar SortSam
I=/data/medddz/Align/Patient_ID.bwamem.bam
O=/data/medddz/Sort/Patient_ID.bwamem.sort.bam
SO=coordinate CREATE_INDEX=TRUE
```

- Mark duplicates

```
$ java -Xmx8g -jar /home/picard/picard-tools-1.129/picard.jar MarkDuplicates
I=/data/medddz/Sort/Patient_ID.bwamem.sort.bam
O=/data/medddz/MarkDup/Patient_ID.bwamem.sort.dedup.bam
M=/data/medddz/MarkDup/Patient_ID.bwamem.sort.dedup.metrics
CREATE_INDEX=TRUE
```

- Create indel realigner targets

```
$ java -Xmx6g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
```

```
-T RealignerTargetCreator
-R /home/ref/b37/human_g1k_v37.fasta
-known /home/ref/b37/1000G_phase1.indels.b37.vcf
-known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf
-I /data/medddz/MarkDup/Patient_ID.bwamem.sort.dedup.bam
-o /data/medddz/Realigner/Patient_ID.bwamem.sort.dedup.bam.intervals
```

- Perform indel realignment

```
$ java -Xmx6g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T IndelRealigner
-R /home/ref/b37/human_g1k_v37.fasta
-known /home/ref/b37/1000G_phase1.indels.b37.vcf
-known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf
-I /data/medddz/MarkDup/Patient_ID.bwamem.sort.dedup.bam
-targetIntervals
/data/medddz/Realigner/Patient_ID.bwamem.sort.dedup.bam.intervals
-o
/data/medddz/IndelRealigner/Patient_ID.bwamem.sort.dedup.indelrealn.bam
```

- Get base recalibration model

```
$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T BaseRecalibrator
-R /home/ref/b37/human_g1k_v37.fasta
-knownSites /home/ref/b37/dbSnp146.b37.vcf.gz
-knownSites /home/ref/b37/1000G_phase1.indels.b37.vcf
-knownSites
/home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf
-I
/data/medddz/IndelRealigner/Patient_ID.bwamem.sort.dedup.indelrealn.bam
-o
/data/medddz/BaseRecal/Patient_ID.bwamem.sort.dedup.indelrealn.recal.grp
-nct 6
```

- Apply recalibration (print reads)

```
$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T PrintReads \
-R /home/ref/b37/human_g1k_v37.fasta
-I
/data/medddz/IndelRealigner/Patient_ID.bwamem.sort.dedup.indelrealn.bam
-BQSR
/data/medddz/BaseRecal/Patient_ID.bwamem.sort.dedup.indelrealn.recal.grp
-o
/data/medddz/FinalBam/Patient_ID.bwamem.sort.dedup.indelrealn.recal.bam
--disable_indel_quals
```

- Create genomic VCF (gVCF) and call all variants

```
$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T HaplotypeCaller --emitRefConfidence GVCF
-R /home/ref/b37/human_g1k_v37.fasta -D
/home/ref/b37/dbSnp146.b37.vcf.gz
-stand_call_conf 30 -stand_emit_conf 10
-I
/data/medddz/FinalBam/Patient_ID.bwamem.sort.dedup.indelrealn.recal.bam
-o /data/medddz/gVCF/Patient_ID.sort.dedup.indelrealn.recal.HC.g.vcf
-L /home/ref/SureSelectAllExonV6/S07604514_Regions_b37.bed -ip 30
```

- Combine gVCF

```
$ java -Xmx18g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T GenotypeGVCFs -R /home/ref/b37/human_g1k_v37.fasta
-D /home/ref/b37/dbSnp146.b37.vcf.gz -stand_call_conf 30
-stand_emit_conf 10
-V /data/medddz/gVCF/Patient_ID.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/gVCF/Patient_ID.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/control.sort.dedup.indelrealn.recal.HC.g.vcf
-V /data/medddz/control.sort.dedup.indelrealn.recal.HC.g.vcf
-o /data/medddz/OnegVCF/all.genotype.vcf -nda --showFullBamList -nt 8
```

*for all 20 patients combined, and later additional 10 FEVR -ve exomes too
*patients were also analysed individually

- Select variants (SNPs)

```
$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T SelectVariants
-R /home/ref/b37/human_g1k_v37.fasta
-selectType SNP
--variant /data/medddz/OnegVCF/all.genotype.vcf
-o /data/medddz/SelectSNPVariants/all.genotype.raw-snps.vcf
```

- Select Variants (INDELs)

```
$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T SelectVariants
-R /home/ref/b37/human_g1k_v37.fasta
--variant /data/medddz/OnegVCF/all.genotype.vcf
-selectType INDEL -selectType MNP
-o /data/medddz/SelectINDELVariants/all.genotype.raw-indels.vcf
```

- Variant filtration for SNPs and INDELs separately

```
$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T VariantFiltration
-R /home/ref/b37/human_g1k_v37.fasta
-V /data/medddz/SelectSNPVariants/all.genotype.raw-snps.vcf
--filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 ||
MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0"
--filterName "snp_hard_filter"
-o /data/medddz/VariantFiltrationSNP/all.genotype.filtered.snps.vcf

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T VariantFiltration
-R /home/ref/b37/human_g1k_v37.fasta
-V /data/medddz/SelectINDELVariants/all.genotype.raw-indels.vcf
--filterExpression "QD < 2.0 || FS > 200.0 || ReadPosRankSum < -20.0"
--filterName "indel_hard_filter"
-o /data/medddz/VariantFiltrationINDEL/all.genotype.filtered.indels.vcf
```

- Combine SNPs and INDELs

```
$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-
0/GenomeAnalysisTK.jar
-T CombineVariants -R /home/ref/b37/human_g1k_v37.fasta
--variant /data/medddz/VariantFiltrationSNP/all.genotype.filtered.snps.vcf
--variant /data/medddz/VariantFiltrationINDEL/all.genotype.filtered.indels.vcf
-o /data/medddz/Combine/combined.genotype.filtered.combinedvar.vcf
--genotypemergeoption UNSORTED
```

- Annotate to dbSNP (≥1%)

```
$ perl /home/vcfhacks-v0.2.0/annotateSnps.pl
-d /home/ref/b37/dbSnp142.b37.vcf.gz
/home/ref/b37/clinvar_20150330.vcf.gz -f 0.01 -pathogenic
-i /data/medddz/Combine/combined.genotype.filtered.combinedvar.vcf
-o
/data/medddz/AnnotatedbSNP/combined.0.01.filtered.combinedvar.1pcdbsnp
.vcf -t 6
```

- Filter on ExAC (≥1% or ≥0.01%)

```
$ perl /home/vcfhacks-v0.2.0/filterVcfOnVcf.pl
-i
/data/medddz/AnnotatedbSNP/combined.0.01.filtered.combinedvar.1pcdbsnp
.vcf
-f /home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz
```

-o
/data/medddz/ExAC/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vcf -
w -y 0.01 or 0.0001

- Variant effect predictor (VEP)

$ perl /home/variant_effect_predictor/variant_effect_predictor.pl
--offline --vcf --everything
--dir_cache /home/variant_effect_predictor/vep_cache
--dir_plugins /home/variant_effect_predictor/vep_cache/Plugins
--plugin
Condel,/home/variant_effect_predictor/vep_cache/Plugins/config/Condel/conf
ig/
--plugin ExAC,/home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz
--plugin SpliceConsensus
--fasta
/home/variant_effect_predictor/fasta/Homo_sapiens.GRCh38.dna.primary_as
sembly.fa.gz
-i
/data/medddz/ExAC/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vcf
-o
/data/medddz/VEP/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vep.v
cf
--fork 6

- Filter on sample

$ perl /home/vcfhacks-v0.2.0/filterOnSample.pl
-i /project_folder/geno/genotype.fltd-combinedvars.1pcdbsnp.1pcEVS.vep.vcf
-s Sample_ID_SM -n1 (or -n2)
-o /project_folder/filtered/Sample_ID.vcf

*n1 = only keep variants if present in 1 number of samples

*n2 = only keep variants if present in 2 number of samples

FILTERS:

- Get Functional variants

$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl \
-i
/data/medddz/VEP/combined.0.01.filtered.combinedvar.1pcdbsnp.exac.vep.v
cf
-s all -n1 (or -n2)
-o
/data/medddz/GetFunctional/combined.0.01.gene.filtered.combinedvar.1pcdb
snp.exac.getfunctional.vep.vcf

*n1 = only keep variants if present in 1 number of samples

*n2 = only keep variants if present in 2 number of samples

- Find Bi-allelic variants

$ perl /home/vcfhacks-v0.2.0/findBiallelic.pl
-i /project_folder/filtered/Sample_ID.vcf
-s Sample_ID
-o /project_folder/filtered/Sample_ID.findBiallelic.vcf

- Get Heterozygous variants

$ perl /home/vcfhacks-v0.2.0/getHetVariants.pl -i
/project_folder/filtered/Sample_ID.vcf
-s Sample_ID -o /project_folder/filtered/Sample_ID.getHetVariants.vcf

- Get Homozygous variants (getHetVariants.pl but reversed)

$ perl /home/vcfhacks-v0.2.0/getHetVariants.pl -i
/project_folder/filtered/Sample_ID.vcf
-s Sample_ID
-r -o /project_folder/filtered/Sample_ID.getHetVariants.Hom.vcf

*-r = reverse the command and look for homozygous variants only

- Obtain CADD score

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl
-c /data/shared/cadd/v1.2/*.gz
-i
/data/medddz/GetFunctional/combined.0.01.gene.filtered.combinedvar.1pcdb
snp.exac.getfunctional.vep.vcf
-o
/data/medddz/CADD/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.ex
ac.vep.getfunctional.not_cadd_ranked.vcf -n cadd_not_found.tvs -d --
progress

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl
-d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb
-i
/data/medddz/CADD/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.ex
ac.vep.getfunctional.not_cadd_ranked.vcf
-o
/data/medddz/GeneAnnotator/combined.0.01.gene.filtered.combinedvar.1pcd
bsnp.exac.vep.getfuntional.not_cadd_ranked.gene_anno

- Convert to Excel

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl

-i
/data/medddz/GeneAnnotator/combined.0.01.gene.filtered.combinedvar.1pcd
bsnp.exac.vep.getfuntional.not_cadd_ranked.gene_anno
--vep --gene_anno
-o
/data/medddz/Excel/combined.0.01.gene.filtered.combinedvar.1pcdbsnp.exa
c.vep.getfuntional.not_cadd_ranked.gene_anno.simple.xlsx

7.8 Depth of coverage

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.3-
0/GenomeAnalysisTK.jar -T DepthOfCoverage -R
/home/ref/b37/human_g1k_v37.fasta -I rmdups_indelrealign_recal.bam -o
cov_depth.txt -L
/home/ref/SureSelectAllExonV5/S04380110_Padded_b37.bed -ct 5 -ct 20

Or for 2nd WES

-L /home/ref/SureSelectAllExonV6/S07604514_Regions_b37.bed -ct 5 -ct 20

7.9 ExomeDepth pipeline

- Define the group of BAM files including controls and test samples

bam_files <- c("sample1.bam", "sample2.bam", ...)

- Create count data for autosomal chromosomes from all BAM files

name_counts <- getBamCounts(bed.frame=exons.hg19,
bam.files=bam_files, include.chr=FALSE,
referenceFasta="/home/ref/b37/human_g1k_v37.fasta")

- Convert counts into a data frame which is the input format

name_counts.dafr <- as(name_counts[, colnames(name_counts)],
'data.frame') print(head(name_counts.dafr))

- Define the test sample

test_sample <- name_counts$test_sample.bam

- Define the reference samples

ref_samples < -c('sample1.bam', 'sample2.bam'...)

ref_set < -as.matrix(counts.dafr[,ref_samples])

- Optimise the choice of aggregate reference set

```
choice <- select.reference.set(test.counts=test_sample,
reference.counts=ref_set, bin.length=(name_counts.dafr$end-
counts.dafr$start)/1000, n.bins.reduced=10000)

print (name_choice[ [1] ])
```

- Construct the reference set

```
matrix <- (name_counts.dafr[,choice$reference.choice])

reference.selected <- apply(X=matrix, MAR=1, FUN=sum)
```

- Fit the beta-binomial model on a data frame

```
all.exons <- new('ExomeDepth', test=test_sample,
reference=reference.selected, formula='cbind(test, reference)~1')
```

- Call CNVs

```
all.exons <- CallCNVs(x=all.exons, transition.probability=10^-4,
chromosome=name_counts.dafr$space, start=name_counts.dafr$start,
end=name_counts.dafr$end, name=name_counts.dafr$names)

head (name_all.exons@CNV.calls)
```

- Load the set of common CNVs identified by (Conrad et al., 2010)

```
data(Conrad.hg19)

head(Conrad.hg19.common.CNVs)
```

- Annotate CNV calls

```
all.exons <- AnnotateExtra(x=name_all.exons,
reference.annotation=Conrad.hg19.common.CNVs, min.overlap=0.5,
column.name='Conrad.hg19')

print(head(name_all.exons@CNV.calls))
```

- Add exon information

```
exons.hg19.Granges <-
GenomicRanges::GRanges(seqnames=exons.hg19$chromosome,
IRanges::IRanges(start=exons.hg19$start, end=exons.hg19$end),
names=exons.hg19$name)
```

```
all.exons<-AnnotateExtra(x=name_all.exons,
reference.annotation=exons.hg19.GRanges, min.overlap=0.0001,
column.name='exons.hg19')
```

- Output CNV calls to a .csv file

```
output.file <- 'CNVs.csv'

write.csv(file=output.file, x=all.exons@CNV.calls, row.names=FALSE)
```

7.10 Sequencing primers of expression constructs pDEST40_LRP5 and pDEST47_LRP5

Primer ID	Sequence 5'- 3'
SSCP 2-F	CAAGCAGACCTACCTGAACC
SSCP 3-F	CGGATTGAGCGGGCAGGGAT
SSCP 3-R	GGATGAAGCTGAGCTTGGCGTC
SSCP 6-F	CGACCCGCTAGAGGGCTATGT
SSCP 6-R	GTCGACCGCGATGCCATCGG
SSCP 9-F	GTGCCTGAGGCCTTCTTGGTCT
SSCP 9-R	CCGTGAGCGGGATGGCCACG
SSCP 12-F	CTAGCGGCCGGAACCGCA
SSCP 12-R	CATCACGAAGTCCAGGTGG
cDNA 14-F	GACCTCTCTGAGCCAAGGCC
cDNA 15-R	GTGAAGAGGACCTCGCGCTC
cDNA 16-F	CAAGCATCTCTACTGGATCG

cDNA 18-R	GTGACGGCTTTCCCGAGTGC
cDNA 19-F	CAGTGTGTCCTCATCAAACAG
cDNA 20-R	CTATGAAATTGAGGGGCACG
cDNA 22-F	CTACTCTTCAAACATTCCGG
Ex13-F	CACAGCCCGGATCTCATC
Ex13-R	CTTGGCTCGCTTGATGTTCT

7.11 SDM primer sequences

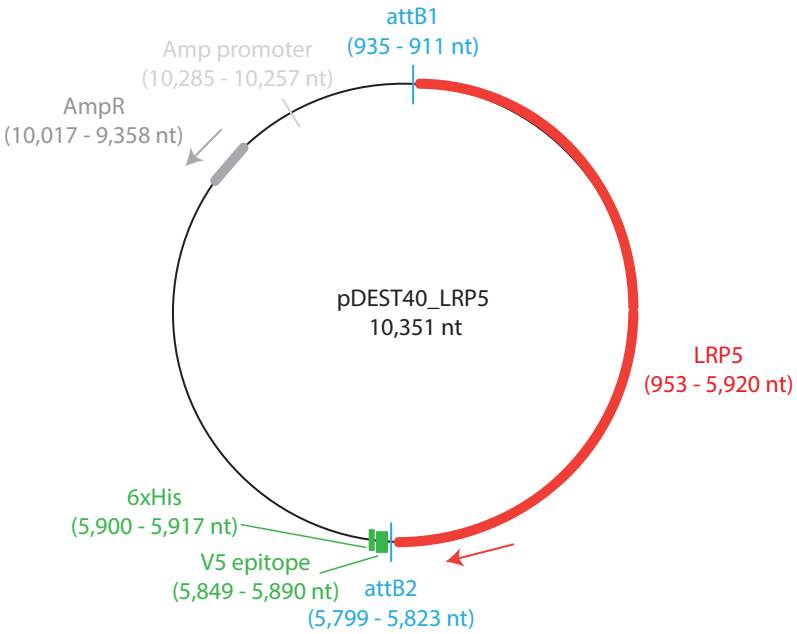
LRP5 c.2951A>G – F

5’ CCAGTGGGTCACAGTCGATGGCTTTGACGTTCC 3’

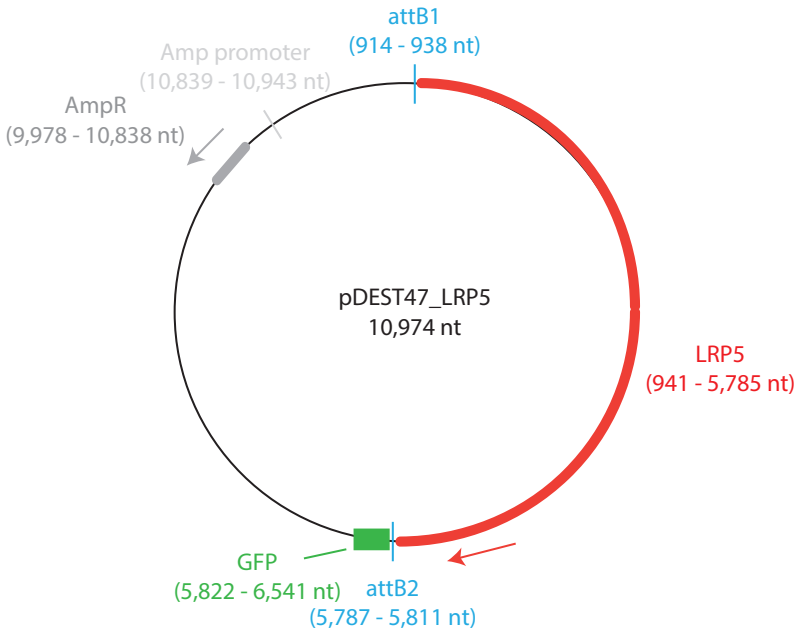
LRP5 c.2951A>G – R

5’ GGAACGTCAAAGCCATCGACTGTGACCCACTGG 3’

7.12 Maps of expression constructs



Schematic representation of the pDEST40_LRP5 expression construct. LRP5 ORF is highlighted in red. AttB recombination sites are highlighted in blue flanking each side of LRP5. GFP tag (in green) is located at the C-terminal of LRP5. AmpR and Amp promotor are highlighted in shades of grey.



Schematic representation of the pDEST47_LRP5 expression construct. LRP5 ORF is highlighted in red. AttB recombination sites are highlighted in blue flanking each side of LRP5. GFP tag (in green) is located at the C-terminal of LRP5. AmpR and Amp promotor are highlighted in shades of grey.

7.13 *P53* primer sequences

p53-F

5’ AGATATTCCCCTGCCCTCAACA 3’

p53-R

5’ CTGGAGTCTTCCAGTGTGAT 3’

7.14 Primer sequences for *LRP5* transcripts

Transcript 1:

LRP5ex1RT – F

5’ ACTCAGAGACCAACCGCATC 3’

LRP5ex3RT – R

5’ TCCATTGGGCCAGTAAATGT 3’

Transcript 2 (new transcript):

LRP5 cDNA ex8a – 2F

5’ CAGGGACTCTGTGTCCTGCC 3’

LRP5 cDNA ex9 – R

5’ GGGTGTGAAGAAGCACAGGT 3’

7.15 *LRP5* primer sequences for Ex8a

LRP5 cDNA ex8a – F

5’ GGCAGGGACAGGCCTGGGGC 3’

LRP5 cDNA ex8a – R

5’ GCCCCAGGCCTGTCCCTGCC 3’

7.16 *DLG1* primer sequences for transcript 4

DLG1v4 ex1 – F

5’ CGATCATTCACCACTCAGCA 3’

DLG1v4 ex1 – R

5’ CACATCCAAGGTTTCAAGGC 3’

7.17 *DLG1* primer sequences for cDNA panel

DLG1 cDNA – 1F

5’ GTATCAGGATGAAGATACACCT 3’

DLG1 cDNA – 1R

5’ ATCTGTGTTGACCAGTACTGG 3’

7.18 *ATOH7* primer sequences

ATOH 7 46delG – F

5’ GATGAAGTCCTGCAAGCCCAG 3’

ATOH 7 46delG – R

5’ ACGTGCCCGCGCACTCGGT 3’

ATOH7 – F Coats

5’ ATGAAGTCCTGCAAGCCCAG 3’

ATOH7 – R Coats

5’ AGTGTTGAGCCCCTGCATGC 3’

7.19 *ISPD* primer sequences

ISPD – F

5’ CCCACCCCGAAGCAATTC 3’

ISPD – R

5’ CCTGTAGGGTGTAGCTGATGA 3’