The molecular genetics of foveal hypoplasia

Emma Catherine Lord

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

School of Medicine

December 2018

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

Acknowledgements

Firstly, I would like to thank my lead supervisor, Dr Carmel Toomes, for her support, guidance and ever enthusiastic approach to the project. My thanks also go to Prof Chris Inglehearn and Dr Manir Ali for their supervision, support and guidance. I have also been extremely lucky to work alongside Dr James Poulter, who has been a great supervisor. I greatly appreciate the opportunity to undertake the PhD and would like to thank The University of Leeds for funding my scholarship. I would like to extend my gratitude to the UK Inherited Retinal Disease Consortium, funded by Fight for Sight and RP Fighting Blindness, for their support. I would like to thank Dr Chris Watson for his help analysing the genomes in this study and also Phoebe Cambridge (MSc) who helped with the analysis of the F1288 variants and designed the primers.

I am very grateful to the Leeds Vision Research Group, in particular, Jo, Claire, Evi, Carla and Layal, alongside all members of Level 8 who have made my time at Leeds much happier. A very special thanks goes to Denisa, who has been there throughout this journey and without whom my time in Leeds would not have been the same – you are a true friend.

Most importantly, my acknowledgements would not be complete without thanking my greatest support network, my family. Thank you to my parents, Martin and Debbie, for the countless opportunities they have given me and

their ability to always put a smile on my face – this thesis also belongs to you. A big thank you also goes to my brother, Jonathan, and my grandparents, who have always been so supportive and who never fail to brighten my day. Finally, I would like to thank Alexander Dudok de Wit, for his endless patience, love and for making me so happy.

Abstract

Foveal hypoplasia with optic nerve decussation defects in the absence of pigmentation deficits (FHONDA) is a rare recessive disorder which usually presents with congenital nystagmus. To date, *SLC38A8* is the only gene known to be mutated in this disease. However, many patients with the FHONDA phenotype do not have detectable mutations in *SLC38A8*, suggesting that this disorder is either genetically heterogeneous or that a proportion of mutations in *SLC38A8* are undetectable by Sanger screening. This study describes the genetic analysis of a cohort of foveal hypoplasia patients to expand the mutation spectrum of *SLC38A8*, investigate if mutations in new gene(s) cause this disorder, or if *SLC38A8* mutations are being missed.

Two novel mutations were identified in *SLC38A8*. In one patient, a homozygous missense mutation was identified (c.848A>C, p.(D283A)). In a second patient, a heterozygous missense mutation was identified (c.2T>C, p.?(M1?)), but a second mutation was not detected. Whole genome sequencing (WGS) was performed on this patient, and application of a 'midigene' splicing assay was used to analyse a candidate intronic splicing variant, but no mis-splicing was observed and the second mutation remains outstanding. In a third case, a homozygous inversion of exons 6, 7 and 8 of *SLC38A8* was detected by WGS (g.84015974_84027074inv). ExomeDepth analysis in a fourth case revealed a homozygous duplication in exon 6 of *SLC38A8*, but the breakpoints were not defined.

The lack of detectable *SLC38A8* mutations in patients with FHONDA syndrome, and haplotype analysis, indicated genetic heterogeneity. Whole exome sequencing of *SLC38A8*-negative cases identified mutations in *PAX6* in two patients and variants in candidate genes requiring further investigation, including *LAMP1*, *VSX2* and *EPHA2*.

This study is the first to characterise structural mutations in *SLC38A8*. Furthermore, the results of this study will impact on FHONDA families by facilitating molecular testing and genetic counselling.

Abbreviations

ABCA4	ATP-binding cassette, sub-family A (ABC1), member 4
AIED	Aland Island eye disease
ALDH3A1	Aldehyde dehydrogenase 3A1
AMP	Ampicillin
OA1	Ocular albinism protein 1
A-P	Antero-posterior
AP3B1	Adapter related protein complex 3 beta 1
AR	Autosomal recessive
ASD	Anterior segment dysgenesis
ATOH7	Atonal homolog 7
BAC	Bacterial artificial chromosome
BAM	Binary alignment/map
BF	Bayes factor
BLOC1S3/6	Biogenesis of lysosomal organelles complex 1 subunit 3/6
β-ME	β-mercaptoethanol
BMPs	Bone morphogenetic proteins
BSA	Bovine serum albumin
Buffer RLT	guanidine-thiocyanate-containing buffer
BWA	Burrow's-Wheeler aligner
CACNAF1	Calcium voltage-gated channel subunit alpha F1
CADD	Combined annotation-dependent depletion
CAT	chloramphenicol acetyltransferase
cDNA	Complementary DNA

CEP290	Centrosomal protein of 290 kDa
cGMP	Cyclic guanosine monophosphate
CHS	Chediak-Higashi syndrome
CIP	Calf intestinal alkaline phosphatase
CNS	Central nervous system
CNV	Copy number variation
COS7	African green monkey kidney cells
CXCR-4	C-X-C chemokine receptor type 4
dbSNP	Database of single nucleotide polymorphisms and multiple small-scale variations
ddH_2O	double-distilled water
DEPC	Diethylpyrocarbonate
dH_2O	de-ionised water
DHS	DNaseI hypersensitive site
DNA	Deoxyribonucleic acid
DNAAF1	Dynein, axonemal, assembly factor 1
dLGN	Dorsal lateral geniculate nucleus
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulphoxide
dNTPs	Deoxyribonucleic triphosphate
DPBS	Dulbecco's phosphate buffered saline
dsDNA	Double-stranded DNA
DTNBP1	Dystrobrevin binding protein 1
D-V	Dorso-ventral
EDTA	Ethylenediaminetetraacetic acid
EFTFs	Eye field transcription factors

ENCODE	Encyclopedia of DNA elements
ERDC	European Retinal Disease Consortium
ESE	Exonic splicing enhancers
ESS	Exonic splicing silencers
EVS	Exome variant server
ExAC	Exome aggregation consortium
ey	Eyeless
EYS	Eyes shut homolog
FA	Fluorescein angiography
FATHMM	Functional analysis through hidden Markov models
FAZ	Foveal avascular zone
FCS	Fetal calf serum
FEVR	Familial exudative vitreoretinopathy
FGF	Fibroblast growth factor
FH	Foveal hypoplasia
FHONDA	Foveal hypoplasia, optic-nerve decussation defects and anterior segment dysgenesis in the absence of albinism
FoSTeS	Forkhead stalling and template switching
FOXD1	Forkhead box D1
FRMD7	FERM domain containing 7
FZD4	Frizzled class receptor 4
GABA	Gamma-aminobutyric acid
GATK	Genome analysis toolkit
GCL	Ganglion cell layer
gDNA	Genomic DNA
GeL	Genomics England

GLAST	Glutamate aspartate transporter
gnomAD	Genome aggregation database
GPCR	G-protein coupled receptor
GPR143	G-Protein coupled receptor 143
GS	Griscelli syndrome
H_2O	Water
HCl	Hydrogen chloride
HEK293T	Human embryonic kidney cells, highly transfectable
HPS	Hermansky-Pudlak syndrome
HSF	Human splice finder
HSMM	HotShot mastermix
IGV	Integrative genomics viewer
ILM	Inner limiting membrane
INL	Inner nuclear layer
Indels	Insertion/deletion mutations
IPL	Inner plexiform layer
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRD	Inherited retinal disease
ITGB8	Integrin, beta 8
KAN	Kanamycin
KCl	Potassium chloride
LAMP1	Lysosomal membrane protein 1
LAR II	Luciferase assay reagent
LB	Luria-Bertani
LCA	Lebers congenital amaurosis
LCA5	Leber congenital amaurosis 5

LGN	Lateral geniculate nucleus
LIDA	Leeds Institute for Data Analytics
LOD	Logarithm of the odds
LOVD	Leiden open variant database
LRAT	Lecithin retinol acyltransferase
LRP5	Low-density lipoprotein receptor-related protein 5
LTR	Long terminal repeat
LYST	Lysosomal trafficking regulator
MATP	Membrane-associated transporter protein
$MgCl_2$	Magnesium chloride
MGI	Mouse Genome Informatics
$MgSO_4$	Magnesium sulphate
MITF	Melanogenesis associated transcription factor
M-L	Medial-lateral
MMBIR	Microhomology-mediated break-induced replication
MMLV	Moloney murine leukemia virus
MNV	Multi-nucleotide variant
NaCl	Sodium chloride
NCAM	Neural cell adhesion molecule
NCBI	National Centre for Biotechnology Information
NCMD	North Carolina macular dystrophy
NCRNA/NSRNA	Non-syndromic congenital retinal detachment
NDP	Norrie disease (pseudoglioma)
NFL	Nerve fibre layer
NGS	Next generation sequencing
NHEJ	Non-homologous end joining

NPPB	Natriuretic peptide precursor B
N-T	Naso-temporal
OA	Ocular albinism
OCA	Oculocutaneous albinism
OCT	Optical coherence tomography
OLM	Outer limiting membrane
ONL	Outer nuclear layer
OMIM	Online mendelian inheritance in man
OPL	Outer plexiform layer
PAX6	Paired box 6
PCR	Polymerase chain reaction
PEDF	Pigment epithelium derived factor
pen/strep	Penicillin/streptomycin
pRL-TK	Renilla luciferase plasmid
PROVEAN	Protein variant effect analyzer
PSORTII	Prediction of protein sorting signals and localization sites in amino acid sequences
PST	Proline, serine, threonine
RFU	Relative luciferase units
RGC	Retinal ganglion cell
RHO	Rhodopsin
RIFS	Reference into function
RNA	Ribonucleic acid
Robo	Roundabout
RPE	Retinal pigment epithelium
RFLPs	Restriction fragment length polymorphisms

RLM-RACE	RNA ligase-mediated rapid amplification of cDNA ends
RP	Retinitis pigmentosa
RT	Room temperature
RT-PCR	Reverse transcriptase-polymerase chain reaction
SC	Superior colliculus
SDM	Site directed mutagenesis
sey	Smalleye
SHH	Sonic hedgehog
SIFT	Sorting intolerant from tolerant
SINE	Short interspersed nuclear element
SIX3	Sine oculis homeobox 3
SLC38A8	Solute carrier family 38, member 8
SMART	Simple modular architecture research tool
SMRT	Single-molecule real-time sequencing
SNAT	Sodium-coupled neutral amino acid transporter
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variant
SP	Spectinomycin
STRPs	Short tandem repeat polymorphisms
SV	Structural variation
TAE	Tris-acetate-EDTA
TAP	Tobacco acid pyrophosphatase
TBE	Tris-borate-EDTA
TD-PCR	Touch down-polymerase chain reaction
TE	Tris-Ethylenediaminetetraacetic
TGF-β	Transforming growth factor-beta

TIC	Translation initiation codon
TRP-1/2	Tyrosinase protein 1/2
TSPAN12	Tetraspanin-12
TYR	Tyrosinase
USH	Usher syndrome
USH2A	Usherin
UV	Ultraviolet
VCF	Variant call format
VEGF	Vascular endothelial growth factor
VEP	Variant effect predictor
UKIRDC	United Kingdom Inherited Retinal Disease Consortium
USA	United States of America
UTR	Untranslated region
VSX2	Visual system homeobox 2
V-T	Ventro-temporal
WES	Whole exome sequencing
WGS	Whole genome sequencing
WT	Wild-type
ZMWs	Zero-mode wavelengths
ZNF408	Zinc finger protein 408

Contents

Acknowledgements ..... **II**
Abstract ..... **IV**
Abbreviations ..... **VI**
Contents ..... **XIV**
List of Figures ..... **XXI**
List of Tables ..... **XXV**

List of Figures

List of Tables

Chapter 1 Introduction

1.1 General overview of retinal disease

Inherited retinal diseases (IRDs) are a diverse group of Mendelian inherited genetic disorders predominantly causing visual impairment. Inherited retinal disease has an estimated prevalence of 1 in 2,000 (Krumpaszky *et al.*, 1999; Rattner, Sun and Nathans, 1999; Francis, 2006; Hamel, 2006; Cremers *et al.*, 2018; Sohocki *et al.*, 2001), and a high impact due to the social and healthcare burden, alongside the difficulties and low quality-of-life experienced by blind or visually impaired people (Williams, Pathak-Ray and Austin, 2007; Köberlein *et al.*, 2013; Chakravarthy *et al.*, 2017; Liew, Michaelides and Bunce, 2014; Francis, 2006). IRDs are characterised by both clinical and genetic heterogeneity, thus ensuring many challenges for investigating teams (Cremers *et al.*, 2018; Astuti *et al.*, 2018). There is also high phenotypic variability both within and between retinal conditions, including variation in age of onset, severity and progression. IRDs can occur as non-syndromic forms such as LCA (MIM 204000) (Leber, 1869), Stargardt disease (MIM 248200) (Stargardt, 1913) or North Carolina macular dystrophy (NCMD) (MIM 136550) (Lefler, Wadsworth and Sidbury, 1971; Frank *et al.*, 1974) or can be diagnosed as part of a syndrome, such as Usher syndrome (MIM 276900) (Usher, 1909), Joubert syndrome (MIM 213300) (Joubert *et al.*, 1969) and Bardet-Biedl syndrome (MIM 209900) (Harnett *et al.*, 1988; Green *et al.*, 1989), to name a few. There is a clear need to understand the genetics of IRDs in order to support patients in the long-term, leading to improved counselling and a better understanding of the likely prognosis

(Sohocki *et al.*, 2001). Further to understanding a patient's molecular diagnosis, the development of many gene-specific and mutation-specific treatments for IRDs are underway (Cremers *et al.*, 2018). The number of genes implicated in retinal disease has increased steadily over the years (Figure 1-1), in line with more meticulous phenotyping, availability of the human genome sequence and the advancement of new sequencing technologies, including next generation (NGS) and third generation sequencing (Eisenberger *et al.*, 2013).

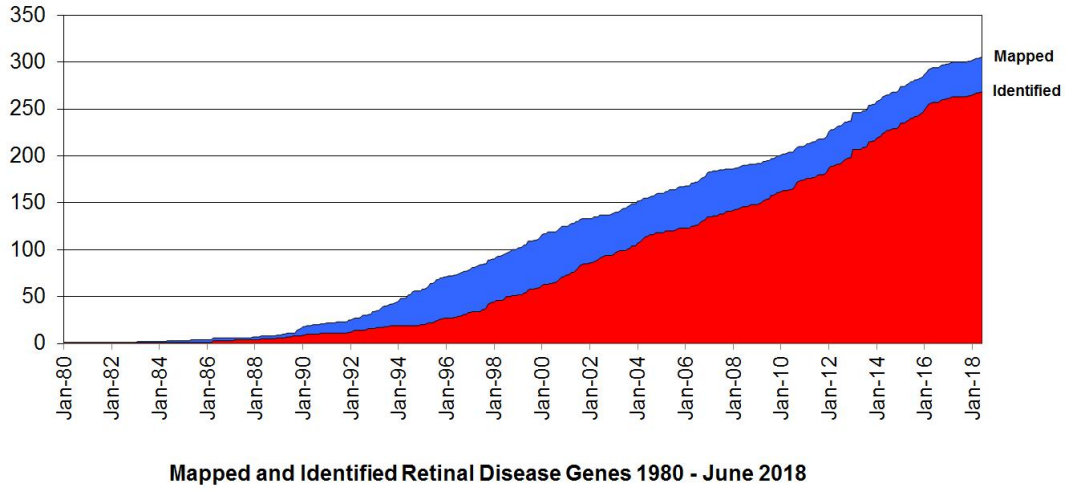


Figure 1-1 RetNet mapped and identified genes found to be defective in IRDs 1980-2017.

There is a high level of genetic heterogeneity in IRDs, with 269 genes implicated in IRDs to date and at least another 39 chromosomal loci identified. Figure used with permission, RetNet Retinal Information Network, ©1996-2018, Stephen P. Daiger, PhD and The University of Texas Health Science Center, Houston, Texas (https://sph.uth.edu/retnet/).

However, the new technologies, including whole exome sequencing (WES) and whole genome sequencing (WGS), bring both data management challenges and diagnostic challenges, as the narrowing down of candidate variants to find the causal mutation requires detailed interpretation of the results (Katsanis and Katsanis, 2013; Frebourg, 2014; Matthijs *et al.*, 2016). To date, mutations in 269 genes have been implicated in retinal disease, with a further 39 genetic loci mapped (Figure 1-1)

(RetNet, 2018). Even with so much detailed knowledge of the causes of IRDs, only approximately 55-60% of IRD cases are solved by WES (Cremers *et al.*, 2018). This is most likely due to difficulties detecting mutations that are not as easily identifiable, such as structural variations (SVs) or non-coding mutations (Eisenberger *et al.*, 2013; Bujakowska *et al.*, 2016).

This thesis focuses on the study of a rare recessively inherited IRD, foveal hypoplasia. Foveal hypoplasia is a developmental defect characterised by an absent or poorly defined foveal pit in the centre of the retina (macula) and is associated with poor visual acuity (described in more detail in Sections 1.2.4, 1.2.6 and 1.3). It is often found in cases of aniridia (MIM 106210), achromatopsia (MIM 216900) and retinopathy of prematurity and is also a key feature of ocular (MIM 300500) and oculocutaneous (MIM 201300) albinism when present with optic nerve decussation defects (when axons are abnormally and disproportionately misrouted to cross the optic chiasm to innervate the contralateral cortex of the brain) (Michaelides, Jeffery and Moore, 2012). All patients included in this study have been diagnosed with foveal hypoplasia. A number of the cohort have also been diagnosed with optic nerve misrouting, but no signs of albinism. This phenotype has previously been reported as foveal hypoplasia, optic nerve decussation defects and anterior segment dysgenesis syndrome (FHONDA) (MIM 609218), caused by mutations in the gene *solute carrier family 38* (*SLC38A8)* (MIM 615585) (Al-Araimi *et al.*, 2013).

1.2 Basic anatomy and function of the human eye

The human eye (Figure 1-2) is composed of three main layers, the fibrous tunic, the vascular tunic and the neural and pigmented retina.

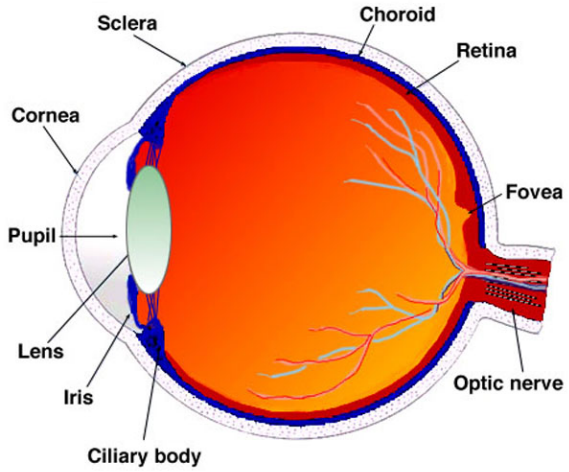


Figure 1-2 Gross anatomy of the human eye.

Sagittal section of the mature human eye highlighting the gross structures. Figure adapted with permission from Webvision (www.webvision.med.utah.edu). Image used with a non-exclusive right under an Attribution, Noncommercial, No Derivative Works Creative Commons license.

The external layer, or fibrous tunic, consists of the sclera and cornea (Figure 1-2), which are both highly collagenous fibrous tissues that envelop the eye globe and maintain its strength and shape. The cornea is an avascular, transparent structure with transparency accomplished by the regulated spacing and fibril growth of mostly collage type I, but also collagen type V (Hassell and Birk, 2010). Light refraction takes place at the corneal surface, in collaboration with the lens, focusing light on the photosensitive retina and reducing light scatter (Maurice, 1957; Benedek, 1971; Farrell, McCally and Tatham, 1973). The cornea also functions to reduce damage

caused by ultra-violet light through the expression of aldehyde dehydrogenase 3A1 (ALDH3A1), also proven to be protective against stress-induced apoptosis (Manzer *et al.*, 2003; Pappa *et al.*, 2005; Lassen *et al.*, 2006). The sclera is a dense tissue composed of collagen fibrils embedded in a matrix of proteoglycans and non-collagenous glycoproteins (Harper and Summers, 2015).

The vascular tunic is the second layer of the eye and includes the choroid, ciliary body, iris and pupil (Figure 1-2). The anterior chamber is the space between the cornea and the iris. The iris, which is anchored peripherally to the ciliary body, serves to control the levels of light entering the eye. Behind the iris is the transparent lens, through which light passes and focuses on the retina. The posterior chamber is the space between the iris and the lens. The choroid is located between the sclera and the retina and is a highly vascularised and pigmented tissue, with major roles in supplying nourishment to the outer retinal layers and preventing uncontrolled reflection of light. The ciliary body connects the choroid with the iris and contains ciliary processes and muscle fibres (Delamere, 2005). The ciliary body is the site of aqueous humour secretion that outflows mainly through the trabecular meshwork at the iridiocorneal angle (the angular recess between the cornea and the anterior surface of the attached margin of the iris), supplying both the anterior and posterior chambers.

1.3 The histology and function of the retina

The internal layer of the eye encompasses the retina, which plays a pivotal role in vision, converting light into an electrical signal transmitted to the visual cortex of the

brain. The retina contains the retinal pigment epithelium (RPE) and the neurosensory retina.

The RPE, which is the outermost layer of the retina, is a pigmented monolayer of hexagonal epithelial cells. The RPE has multiple functions; one major function is to interact with the photoreceptor cell layer in phagocytosing shed photoreceptor outer segments and recycling the all-*trans*-retinol (a vitamin A derivative) back into 11-*cis*-retinal (Bok, 1993; Gu *et al.*, 2012). The RPE is also rich in melanosomes, or pigment granules, that aid in the absorption of scattered light and diminish oxidative stress (Strauss, 2005). Melanosomes are generated *in utero* and remain relatively constant throughout the life-span of the RPE (Mann, 1969).

The neurosensory layer contains the following layers (starting from the vitreous side and moving to the choroidal side) (Figure 1-3): the inner limiting membrane (ILM), the nerve fibre layer (NFL), the retinal ganglion cell (RGC) layer, the inner plexiform layer (IPL), the inner nuclear layer (INL), the outer plexiform layer (OPL), the outer nuclear layer (ONL), the outer limiting membrane (OLM) and the rod and cone photoreceptor layer (Graw, 2010). The neural retina contains seven major neuronal cell types (Figure 1-3); the rod and cone photoreceptor cells, the horizontal cells, the bipolar cells, the interplexiform cells, the amacrine cells and retinal ganglion cells (RGCs). RGC bodies make up the inner-most nuclear layer of the neural retina and their axons become myelinated after they exit the eye via the lamina cribrosa (Section 1.6). The fovea is a depression that lies in the centre of the macula area to the temporal side of the optic nerve head (Figure 1-3). Cone photoreceptors are

concentrated at maximum density in the foveal pit and it is therefore the region with the most acute vision in the retina. Many proteins are involved in the process of phototransduction, including the recycling of all-*trans*-retinol back into 11-*cis*-retinal. This includes the photoreceptor specific ATP-binding cassette transporter gene (*ABCA4*), mutations in which cause Stargardts macular dystrophy (Allikmets *et al.*, 1997). Mutations in the gene encoding lecithin retinol acyltransferase (*LRAT*) (MIM 604863), an important molecular switch involved in Vitamin A esterification, have been identified in cases of early-onset retinal dystrophy including LCA type 14 and juvenile retinitis pigmentosa (RP) (MIM 613341) (Thompson *et al.*, 2001). Retinal pigment epithelium-specific 65-kDa protein (RPE65), is an important isomerase that is encoded by the *RPE65* gene (MIM 180069). Mutations in *RPE65* have been identified in patients with autosomal recessive LCA type 2 (MIM 204100) and RP (MIM 613794) (Morimura *et al.*, 1998).

Visual information is processed by both horizontal and vertical pathways within the retina. In the OPL, the rod and cone photoreceptor cells are tightly stacked together and their axons form synaptic connections with the horizontal cells and bipolar cells (Figure 1-3). The nuclei of the horizontal, bipolar, amacrine, interplexiform and Müller cells are located in the INL (Figure 1-3). Horizontal cell bodies and amacrine cells located in the INL are responsible for modulation of bipolar and RGC response through lateral connections (Dyer and Cepko, 2001; Euler *et al.*, 2014). The modulation of the retinal circuitry is important for adaptation under conditions of low and high light and contrast sensitivity, so that the eye can see optimally in variable light conditions (Hildebrand, 2011). The inner retina is provided with oxygen and nutrients via the vasculature layers known as the intraretinal beds, which lie on either

side of the INL, and also from the superficial primary vascular plexus within the NFL (Hildebrand, 2011). The photoreceptor cells also interact with the bipolar cells directly, as the bipolar cells transmit signals to the IPL. There are 2 distinct types of bipolar cell that are distinguished by their response to light on the centres of their receptive fields (Euler *et al.*, 2014). Whether bipolar cells are ON- (metabotropic receptors, glutamate hyperpolarisation) or OFF- centre (ionotropic receptors, glutamate depolarisation) depends on their response to glutamate release by the cone photoreceptors (Euler *et al.*, 2014). Visual information is ultimately transmitted to the RGC layer of the inner retina. The RGC axons travel towards the optic nerve head within the NFL (Figure 1-3). All axons unite at the optic nerve to transmit the visual signal to higher brain regions including the lateral geniculate nucleus (LGN), the superior colliculus and the hypothalamus (Hildebrand, 2011).

Four glial cell types are found in the retina: Müller cells, astrocytes, microglia and oligodendrocytes (occasionally). Müller cells are the predominant type of retinal glial cells and they span almost the full width of the neural retina, with their cell bodies located in the INL and their processes projecting into the OLM and ILM (de Melo Reis *et al.*, 2008). They provide physical and metabolic support for the retinal neurons and are formed alongside the other neurons, whereas astrocytes migrate into the retina from the optic nerve (Dyer and Cepko, 2001; Vecino *et al.*, 2016; Turner and Cepko, 1987).

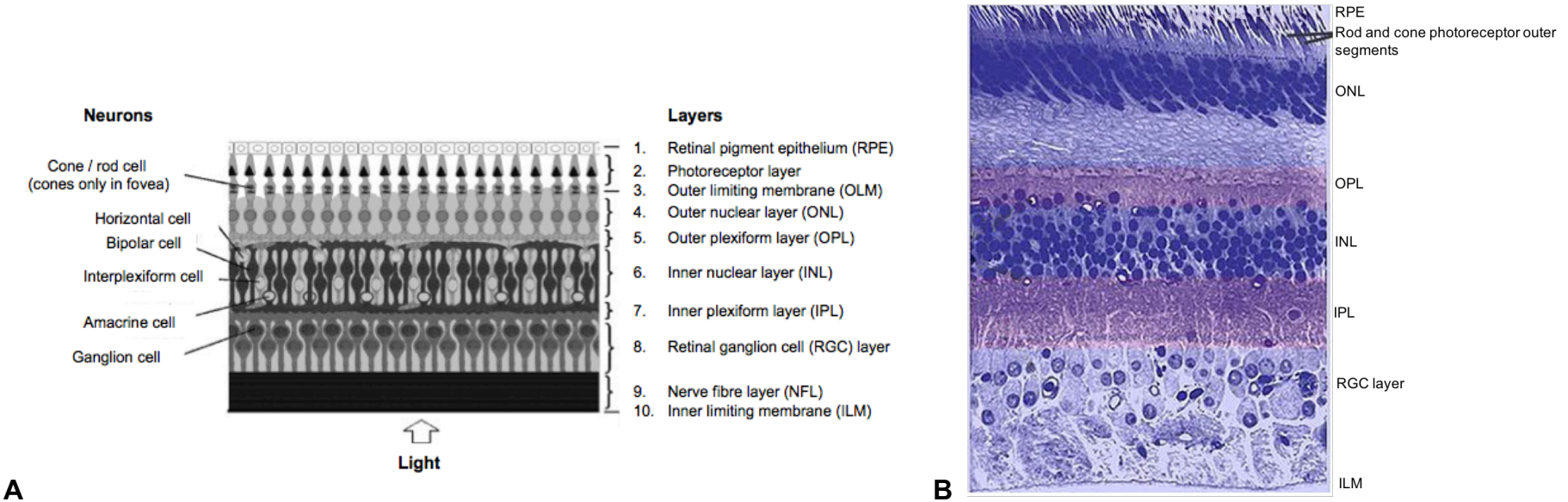


Figure 1-3 The lamination of the mature human retina and organisation of the neurons.

A diagram of a vertical section through the retina detailing all ten layers and the neurons located in each region. Figure adapted with permission from (Martínez *et al.*, 2008). B Light micrograph of a vertical section through the central region of the human mature retina. Image adapted with permission from Webvision (www.webvision.med.utah.edu). Image used with a non-exclusive right under an Attribution, Non-commercial, No Derivative Works Creative Commons licence.

Astrocytes are mostly limited to the NFL and function in a similar manner to Müller glia (Vecino *et al.*, 2016). Microglia enter the retina alongside the mesenchymal retinal blood vessel precursors during early development (Chan-Ling, 1994; Provis, 2001). The innermost layer of the retina is the ILM, which as a basement membrane provides a physical barrier between the neural retina and the vitreous humour and is vital for maintaining the integrity of the RGC layer (Halfter *et al.*, 2005).

1.4 Embryology of the vertebrate eye

A number of regulatory factors, known as eye-field transcription factors (EFTFs), act as inductive signals to initiate and mediate development of the eye field, as they are expressed in the anterior region of the vertebrate neural plate. EFTFs, such as sine oculis homeobox homolog 3 (*Six3)*, paired box 6 (*PAX6)* and melanogenesis associated transcription factor (*MITF)*, are paramount in the development of the early eye. Mutations in these genes are known to cause developmental eye malformations, including holoprosencephaly (MIM 236100) (Wallis *et al.*, 1999) Aniridia (MIM 106210) (Ton *et al.*, 1991; Jordan *et al.*, 1992) (Section 1.8) and Waardenburg syndrome type 2 (MIM 193510) (Tassabehji, Newton and Read, 1994).

In the developing prosencephalon (forebrain), the human eye field emerges during neural plate patterning, from around embryonic day 22 (Chow and Lang, 2001; Zuber *et al.*, 2003) (Figure 1-4). The neural plate is the site of neurulation as it bends dorsally, separating the eye field laterally and forming the optic sulci, or optic

grooves in the lateral neuroectoderm (Barber, 1955; Mann, 1969) (Figure 1-4, A, B). As the forebrain grows, closure of the neural tube results in deeper evagination of the optic grooves, forming bilateral outpocketings called optic vesicles (Barber, 1955, Mann, 1964) (Figure 1-4, C, D). The optic vesicles extend from the forebrain towards the surface ectoderm through the adjacent mesenchyme. The portion of each optic vesicle that interacts with the surface ectoderm induces that area to form a thickening called the lens placode, which is the precursor of the lens. The lens placode and optic vesicle invaginate in coordination, inducing differentiation of the lens vesicle and the optic cup, a double-layered structure with layers initially separated by intraretinal space (Spemann, 1924) (Figure 1-4, E, F). The lens vesicle forms at approximately 5 weeks' gestation, as the surface ectoderm seals over it to form the future corneal epithelium (Fuhrmann, 2010) (Figure 1-4, E, F). Through the adjacent mesenchyme, the ventral retina and optic stalk connect to create the choroidal fissure, thus forming the vascular tunic (Barber, 1955, Mann, 1964). The choroidal fissure is a groove through which blood vessels, including the hyaloid artery, can connect to the inner eye chamber. Fusion of the choroidal fissure by the sixth week encloses the blood vessels in a canal in the optic stalk. This fissure also facilitates growth of RGC axons from the neural retina into the optic stalk during the eighth week, forming the optic nerve (Fuhrmann, 2010; Heavner and Pevny, 2012) (Figure 1-4, G). Proper optic nerve placement is reliant on signals that pattern the optic cup along the dorso-ventral (D-V) axis, mediated by members of the VAX family of homeodomain transcription factors, including VAX1 and VAX2 (Barbieri *et al.*, 1999). The anterior eye segment encompasses all structures between the cornea and the façade of the vitreous (Idrees *et al.*, 2006). At around week 7, the inner layer of the optic cup forms the neural retina and the outer layer forms the RPE, with the

two layers separated by the intraretinal space (Fuhrmann, 2010). The inner layer is patterned along its D-V and nasal-temporal (N-T) axes. The point where the neural retina and RPE connect will eventually give rise to the ciliary body and iris. Anterior segment dysgenesis (ASD) is a spectrum of rare abnormalities arising from maldevelopment of the cornea, iris and lens (Idrees *et al.*, 2006). It should be noted that Axenfeld anomaly (MIM 109120), a form of ASD where defects are limited to the peripheral anterior segment, is a variable feature of patients with FHONDA syndrome, as studied in this thesis (Pal *et al.*, 2004; Al-Araimi *et al.*, 2013; Poulter *et al.*, 2013).

The hyaloid vascular system expands from the optic nerve head and radiates outwards towards the peripheral retina. It is around the second month that the hyaloid vessels branch to form the capillary network posterior to the lens (Fruttiger, 2007). During the fourth month, macrophage-mediated regression of the hyaloid vasculature takes place, and is replaced by the primary vascular plexus (Fruttiger, 2007). There are two modes of vascularisation that are considered in the developing retina; vasculogenesis and angiogenesis (Hughes, Yang and Chan-Ling, 2000; Fruttiger, 2007). The formation of the primary vascular plexus is believed to occur by vasculogenesis, which is the *de novo* formation of vessels from vascular endothelial precursor cells that enter the retina from the optic nerve head (Hughes, Yang and Chan-Ling, 2000). The sprouting of vessels from the vasculature that is already present, known as angiogenesis, is responsible for the vascularisation of the peripheral retina and development of the intraretinal vascular plexuses (Hughes, Yang and Chan-Ling, 2000).

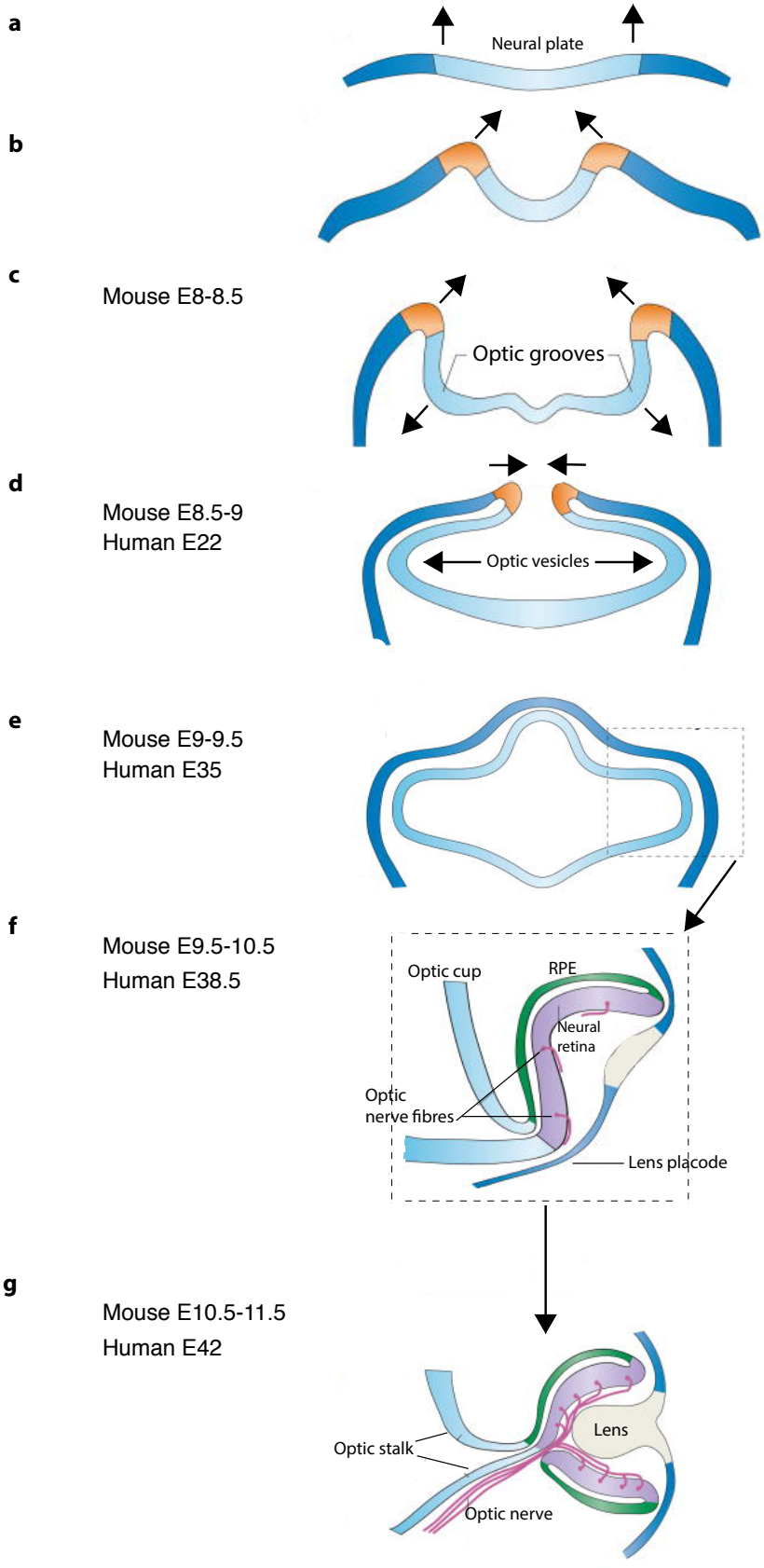


Figure 1-4 Key developmental stages of the vertebrate eye.

Schematic diagram representing the development of the vertebrate eye during embryogenesis. A The development of the vertebrate eye cup begins with the neural plate (light blue) which arises from the ectoderm (dark blue). B-C The neural plate transitions as the neural plate borders (orange) with the precursors of the neural crest cells (orange) dorsally fold upwards and inwards, representing the start of neurulation. Depressions, or optic grooves evaginate on either side of the rostral neural plate as the eye field begins to separate. D The neural plate borders (orange) finally meet, closing off the neural tube. At this stage, the optic grooves expand and become the optic vesicles. E Continuation of the optic vesicle expansion from the diencephalon of the embryonic forebrain results in their contact with the surface ectoderm and induction of the lens placode. The neural tube is separated as the growing forebrain extends upwards. F The dual-layered optic cup is formed, with the neural retina located on the inner layer and the retinal pigment epithelium (RPE) on the outer layer. This takes places as the lens placode and optic vesicle simultaneously invaginate, also forming G the lens vesicle. The lens pit and eventually the lens is formed as the vesicle detaches from the surface ectoderm at the lens placode. Surface ectoderm also differentiates into corneal epithelium (presumptive corneal epithelium, dark blue) after receiving inductive signals from the lens vesicle. The eye cup continues to grow, with the eventual closure of the choroidal fissure, generating a channel through which RGC axons will travel and form the optic nerve. Image adapted with permission from Copyright Clearance Centre on behalf of Springer Nature: Nature Reviews Neuroscience (Lamb, Collin and Pugh, 2007).

The process is mediated by both pro- and anti- angiogenic factors, including vascular endothelial growth factor (VEGF), important in both angiogenesis and vasculogenesis (Hughes, Yang and Chan-Ling, 2000; Provis, 2001; Gerhardt *et al.*, 2003; Fruttiger, 2007; Patel-Hett and D'Amore, 2011). Neural connections of the retina are established by 5 months' gestation and by approximately 7 months' gestation, the eye becomes sensitive to light as the rod and cone photoreceptors mature (Mann, 1969).

1.5 The structure of the posterior eye

The posterior part of the human eye is dominated by the retina, as shown in the example fundus photograph (Figure 1-5). Located temporally to the optic disc is the *macula lutea* which encompasses the entire foveal structure (foveal pit) in an

elliptically shaped area of approximately 5.5mm diameter, including the perifovea, parafovea, foveal slope and foveola (Figure 1-5) (Provis *et al.*, 2010). The foveal pit, which is characterized by its distinctive shape, is the most important region of the human retina because it is responsible for high-acuity vision. In order to read the words of this thesis, the image must be focused on the fovea, which comprises just 0.01% of the retinal area, but accounts for the activity of over 8% of the primary visual cortex of the brain (Talbot and Marshall, 1941; Daniel and Whitteridge, 1961; Cowey, 1964; Azzopardi and Cowey, 1993).

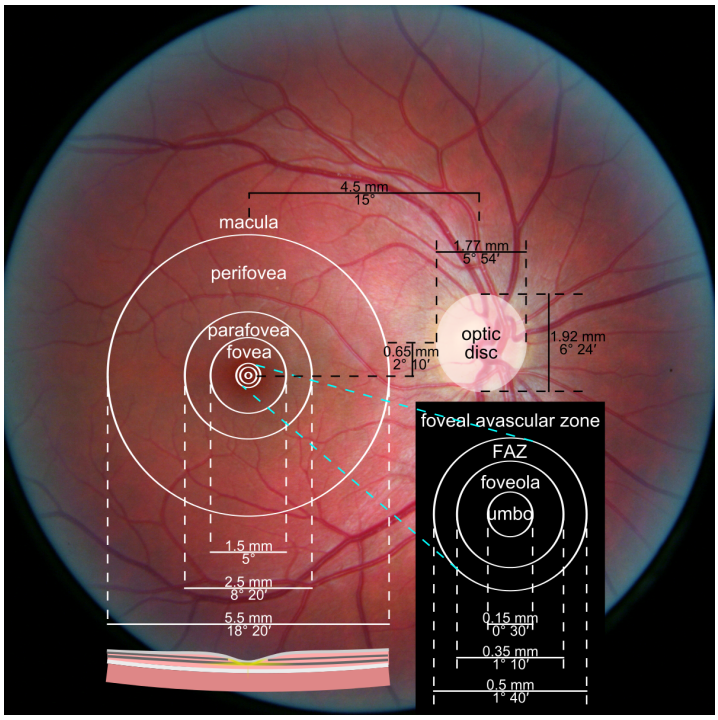


Figure 1-5 Fundus photograph of a normal retina, highlighting the location of the fovea and the optic disc.

Positions and sizes of the fovea, parafovea, perifovea, macula and optic disc are shown. Photograph reproduced with permission: Danny Hope from Brighton & Hove, UK [CC BY 2.0 (http://creativecommons.org/licenses/by/2.0)], via Wikimedia Commons.

The foveola is surrounded by the foveal slope (Figure 1-5), also known as the clivus, which marks the definition between the vascular and avascular retina, a region

known as the foveal avascular zone (FAZ). The fovea is adapted for high-acuity vision, as the FAZ stops the retinal vessels from impeding vision (Hildebrand, 2011). The fovea is entirely devoid of rod photoreceptors and is populated by tightly packed cone photoreceptor cells (Michaelides, Jeffery and Moore, 2012). Cone photoreceptors mediate vision in bright light and colour and are responsible for high visual acuity (whereas rods are specialised for high sensitivity in monochrome in low light conditions) (Yau and Hardie, 2009; Sung and Chuang, 2010). The fovea has the highest visual acuity in comparison with other parts of the retina because each fovea cone is connected to one bipolar cell and one ganglion cell (Yau and Hardie, 2009; Sung and Chuang, 2010). In the peripheral retina, the bipolar and ganglion cells synapse with multiple photoreceptors (Yau and Hardie, 2009; Sung and Chuang, 2010). The RGC axon fibres travel within the NFL to reach the optic nerve disc (Sections 1.6 and 1.7). The development of the fovea and the optic nerve/optic chiasm is a key feature of the disease investigated in patients in this thesis.

1.6 The development and physiology of the optic nerve/optic chiasm and establishment of binocular vision

The region of the optic nerve head underneath the surface NFL is subdivided into three sections: the pre-laminar region (between the lamina cribrosa and the vitreous), the lamina cribrosa itself, which is formed of connective tissue and the post-laminar region (Figure 1-6). The unmyelinated RGC axons surrounded by astrocytes in the pre-laminar region of the optic nerve head become arranged into axon bundles (Anderson, 1969). The nerve fibre bundles pass through the lamina cribrosa, a multi-layered collagen network composed of stacks of connective tissues

including elastin fibres, laminin and collagen types III and IV, synthesised by astrocytes in the optic nerve head (Hernandez, Igoe and Neufeld, 1986; Hernandez *et al.*, 1987; Anderson, 1969). As the nerve fibres pass through the lamina cribrosa, they enter the post-laminar region, which is populated by oligodendrocytes (Figure 1-6). The oligodendrocytes mediate the myelination of the axon fibres (Raine, 1894). After passing through the optic nerve disc the retinal axons travel in the optic nerve to reach the optic chiasm.

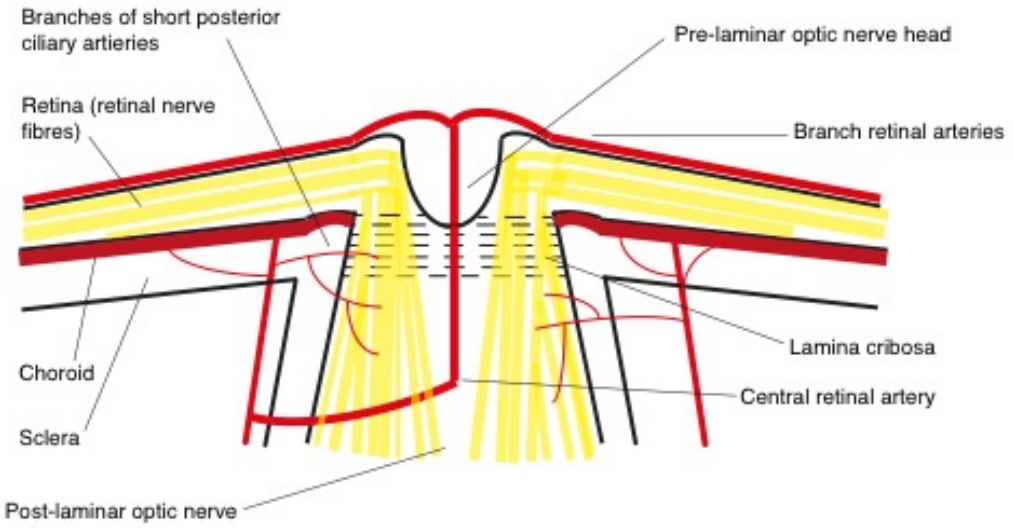


Figure 1-6 Schematic representation of the anatomy of the optic nerve head.

The unmyelinated RGC nerve fibres gather at the pre-laminar region of the optic nerve head. The nerve fibres exit the retina through the lamina cribrosa. Myelination of the nerve fibres takes place to increase impulse conduction along the axons of the optic nerve. Image adapted with permission from Copyright Clearance Centre on behalf of Elsevier: Medical Image Analysis (Chrástek *et al.*, 2005).

The optic chiasm is located at the base of the hypothalamus and is the site of major fibre reordering before the retinal axons separate to form the optic tracts (Jeffery and Erskine, 2005). The nerve fibres either decussate (project contralaterally), or remain

uncrossed (ipsilateral projections), to innervate either the left or right brain hemisphere (Harman and Jeffery, 1992; Jeffery and Harman, 1992) (Figure 1-7). The basis of mammalian binocular vision is formed in a pattern of partial decussation at the chiasm, creating an X-shaped structure (Jeffery and Harman, 1992, Harman and Jeffery, 1992) (Figure 1-7). The level of decussation in different animal species depends on how much of the visual field of each eye overlaps. Therefore, animals with lateral eye positioning, for example many birds and fish, do not have any ipsilateral projections, whereas humans have significantly overlapping visual fields with approximately 50% decussation (Kupfer, Chumbley and Downer, 1967; Horton, 1997) (Figure 1-7). As tadpoles, *Xenopus* have no binocular vision, due to their lateral eye positioning, with solely contralateral projections, but in adult form they develop ipsilaterally-projecting nerve fibres, believed to be due to the *ephrin* gene expression acting as a midline gatekeeper (Nakagawa *et al.*, 2000). Nerve fibres at the ventral midline that project contralaterally across the midline are responsive to chemoattractants, while axons that project ipsilaterally are turned away at the midline by repellents (Petros, Rebsam and Mason, 2008; Dickson, 2002; Evans and Bashaw, 2010). Long-range midline guidance is governed by diffusible molecules known as the netrins. The netrins act locally at the site of the optic discs to direct the RGCs out of the eye and attract all nerve fibre axons to the midline in a range of animal models (Deiner *et al.*, 1997; Flanagan and Van Vactor, 1998; Hopker *et al.*, 1999; Evans and Bashaw, 2010). Studies have shown that mice lacking netrin-1 have optic nerve hypoplasia, due to failure of the RGC axons to exit the eye upon reaching the optic disc (Deiner *et al.*, 1997). Ipsilaterally-projecting neurons are repelled from the midline by the Slit protein that acts via receptors of the roundabout (Robo) family, Robo1 and Robo2 (Thompson *et al.*, 2006; Plachez *et al.*, 2008;

Thompson *et al.*, 2009). On contralaterally-projecting axons, Slit binds to Robo3 receptors, which have no repulsive effect, thus facilitating netrin attraction (Jaworski *et al.*, 2015). It is after axon crossing that Robo3 is downregulated and re-crossing of the axons is prevented (Jaworski *et al.*, 2015).

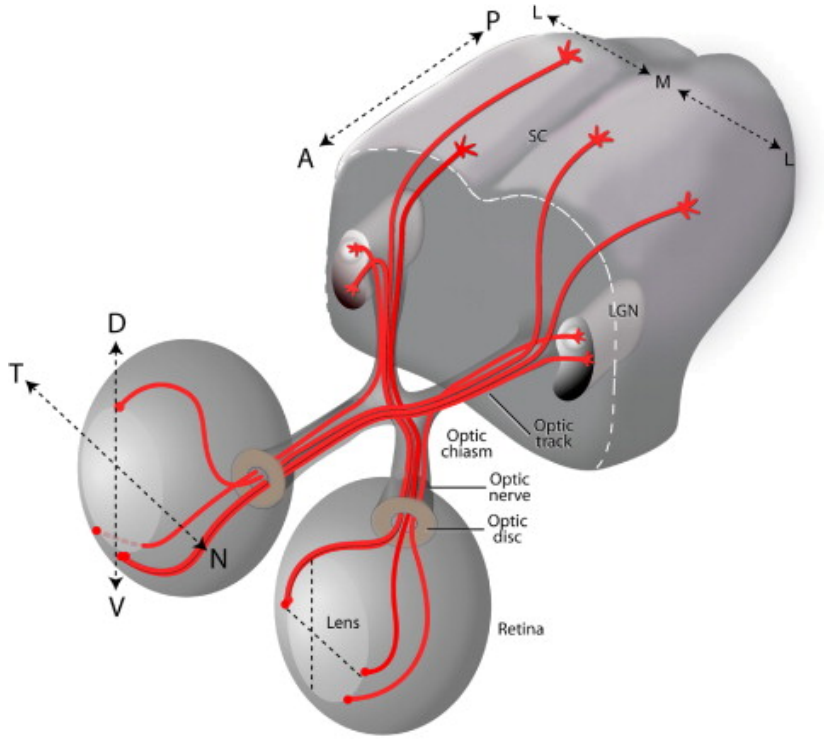


Figure 1-7 Overview of the binocular visual pathway.

The RGC axons from the naso/temporal (N-T) retina exit each eye via the optic disc and extend to the optic chiasm. The axons then either project ipsilaterally or contralaterally to their main targets in the primary visual cortex: the lateral geniculate nucleus (LGN) in the thalamus or the superior colliculus (SC). D dorsal, V ventral, N nasal, T temporal, A anterior, P posterior, L lateral, M medial. Image adapted with permission from Copyright Clearance Centre on behalf of Elsevier: Developmental Biology (Erskine and Herrera, 2007).

Short-range guidance cues include L1, NCAMs, anosmin-1 and the ephrins. L1 and NCAM aid the arrangement of nerve fibres in the optic nerve (Demyanenko and Maness, 2003; Chung *et al.*, 2004; Dai *et al.*, 2013). Mutations in these genes cause decussation defects (Vulliemoz, Raineteau and Jabaudon, 2005). Ephrins, such as

Ephrin-B2 are expressed by radial glial cells residing dorsal to the optic chiasm, and are important in the mapping of the nerve fibres along the retinal D-V axis onto the tectal (the dorsal midbrain) medial-lateral (M-L) axis (Williams *et al.*, 2003). Further to the ephrin-B class, ephrin-As are important in the mapping of the retinal N-T axis on to the tectal A-P topographic visual map (Triplett and Feldheim, 2012). For example, the receptors EphA5 and EphA6 and ephrin-As are expressed in bidirectional gradients, peaking in the temporal region of the central retina (near the presumptive fovea) and tapering toward the N- and T- poles of the peripheral retina (Lambot *et al.*, 2005; Triplett and Feldheim, 2012). Sonic hedgehog (Shh) protein is absent at the optic chiasm prior to RGC axon arrival, with induction instead taking place at the contralateral nerve fibres in the retina (Peng *et al.*, 2018). The morphogen is then transported in an anterograde fashion along the nerve fibres to accumulate at the optic chiasm midline where it functions to repel ipsilateral nerve fibres (Peng *et al.*, 2018). Shh-mediated ipsilateral nerve fibre guidance is therefore a tightly regulated mechanism in the establishment of the binocular pathway and development of neuronal connectivity.

1.7 The development and structure of the fovea

It is not until the FAZ is established at approximately 24-26 weeks that the foveal depression forms (Hendrickson *et al.*, 2006; Provis, Sandercoe and Hendrickson, 2000; Springer, 1999; Springer and Hendrickson, 2005) (Figure 1-8). Retinal vessels enter the developing retina from fetal week 14 and slowly define the foveal rim, ceasing their proliferation in the macula which is avascular (Figure 1-8) (Engerman,

1976; Provis, Sandercoe and Hendrickson, 2000; Sandercoe *et al.*, 2003; Stone, Sandercoe and Provis, 2005).

The slow proliferative process of retinal vascularisation around the periphery of the fovea is governed by VEGF, which is known to be highly expressed in the RGCs in the macula (Sandercoe *et al.*, 2003; Stone, Sandercoe and Provis, 2005). The foveal boundary at which the retinal vessels terminate is reported to be regulated by repellent molecular guidance cues (Kozulin *et al.*, 2009b; Kozulin *et al.*, 2009a). In studies of primate maculae, the axon guidance receptor EphA6 has been found to be expressed in the ganglion cell layer (GCL) of the retina, alongside its ligands ephrin-A1 and –A4 (Kozulin *et al.*, 2009a). This pattern of graded expression to retard the growth of vessels into the central retina has also been detected after birth, even when the retinal vessels are well established, suggesting that this boundary is well-maintained both pre- and post-natally (Kozulin *et al.*, 2009a). It is thought that such repellent signalling molecules first act to drive RGC exit from the RGC layer, before repelling blood vessels to allow generation of the FAZ (Gariano, Iruela-Arispe and Hendrickson, 1994; Hendrickson *et al.*, 2006). In addition to the ephrin family, anti-angiogenic regulators such as the neurotrophin known as pigment epithelium-derived factor (PEDF) and natriuretic peptide precursor B (NPPB) are thought to play a role in the definition of the FAZ (Kozulin *et al.*, 2009b). PEDF has well-known anti-angiogenic effects in the vertebrate retina (Dawson *et al.*, 1999; Karakousis *et al.*, 2001; Huang *et al.*, 2008).

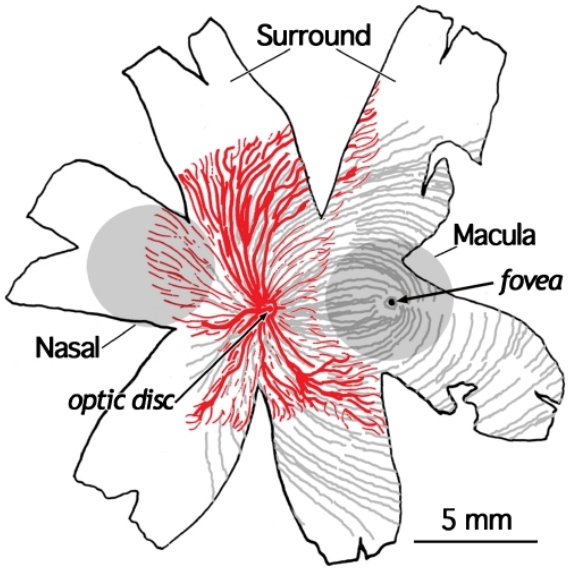


Figure 1-8 Schematic drawing of the human retina aged 20 fetal weeks.

The patterning of retinal axons (grey lines) and of the retinal vessels (red lines) are shown alongside the position of the developing fovea. The macula region is completely devoid of blood supply and is known as the foveal avascular zone (FAZ). The nasal region, however, is supplied by the retinal capillaries. reproduced with permission from (Kozulin and Provis, 2009).

The site of the future fovea is characterised by a thickened GCL stacked with neuronal cells. At 24-26 weeks, the GCL becomes noticeably thinner (3-4 cells deep), as the RGCs migrate out of the FAZ and become displaced onto the foveal rim. The IPL also becomes gradually thinner at this stage as the retinal cell layers move away from the foveal centre. Intraocular pressure, apoptosis and stretching of malleable foveal tissue allowing the passive centrifugal movement of the inner retinal cells, have all been put forward as mechanisms involved in foveal morphogenesis (Kozulin *et al.*, 2009b; Springer, 1999; Georges, Madigan and Provis, 1999). However, because the foveal depression emerges during a period of relatively slow ocular growth, models of retinal stretch are less favoured than the influence of intraocular pressure following definition of the FAZ (Provis *et al.*, 2013). Studies of the developing primate retina have suggested that the region of the fovea instigates

many maturation processes, such as the cessation of mitosis that precedes RGC migration away from the foveal depression (Hendrickson and Yuodelis, 1984).

Concomitant with the RGC and IPL movement away from the fovea, the cone photoreceptors migrate centripetally toward the centre of the fovea. Here the cone photoreceptors become highly concentrated, with their cell bodies layered in oblique columns below the OLM (Hildebrand and Fielder, 2011). It is the efficient packing of the cone photoreceptors in a hexagonal mosaic structure that reinforces the fovea as the retinal region with the greatest visual acuity (Curcio and Sloan, 1992). In the human eye, cone photoreceptors are characterised by long (L), medium (M) and short (S) wavelength-specific opsins, which are sensitive to red, green and blue light, respectively (Xiao and Hendrickson, 2000). The cones of different wavelength sensitivity are the basis of colour perception. The fovea has been reported to initiate their expression, with the early development of S cones at week 10 preceding L/M cones in the fovea at weeks 14-15 (Xiao and Hendrickson, 2000). Cone photoreceptors are approximately 100-fold less sensitive than rods and are able to adjust their photosensitivity to work even in bright light and avoid becoming saturated (Rushton, 1965; Boynton and Whitten, 1970; Baylor, 1987). However, as the retina is highly exposed to damage from bright and UV light, the photoreceptors can still become bleached or more severely damaged or destroyed (Kefalov, 2012). The macula lutea contains yellow screening pigments, xanthophyll carotenoids zeaxanthin and lutein, which act as a short wavelength filters for protection against such light-induced damage (Ahmed, Lott and Marcus, 2005). The fact that the region of the fovea is also avascular also helps to ensure reduced light scatter.

The gradual peripheral displacement of the RGCs along the foveal slope creates the structure known as the parafovea, or foveal rim, which is the thickest portion of the entire retina (Hendrickson, 2016). Outside of the fovea, the cone density falls rapidly to a fairly even density in the peripheral retina, evenly spaced surrounded by rods (Figure 1-9) (Hendrickson, 2016). The optic disc is completely devoid of photoreceptors, giving rise to the blind spot (Figure 1-9).

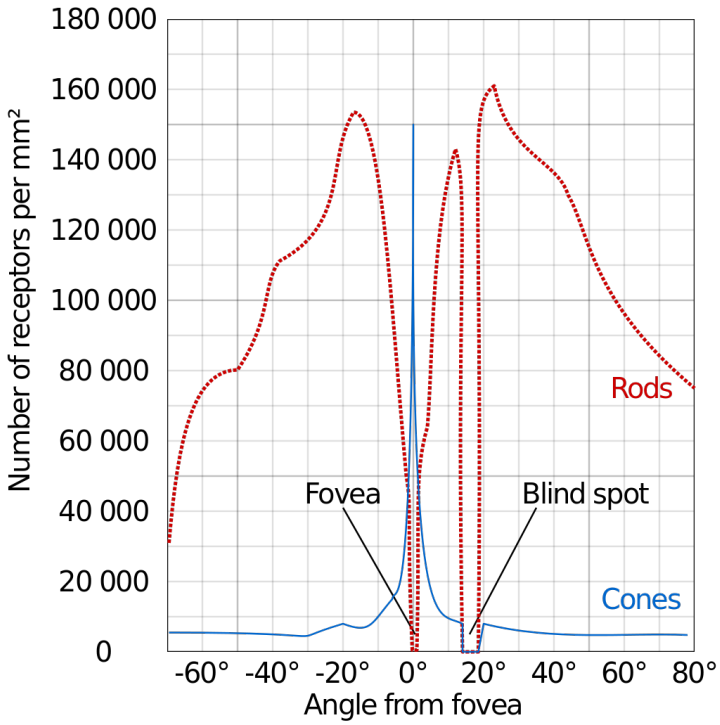


Figure 1-9 The distribution of rod and cone photoreceptors across the human retina.

Density plot of rod and cone photoreceptor cells across the horizontal meridian of the human retina, shown in degrees of visual angle relative to the position of the fovea. The density of rods is higher than cones in the peripheral retina (nasal and temporal locations), with highest density 10-20 degrees peripheral to the fovea. The density of cone photoreceptors peaks at the region of the foveal pit. The optic disc is completely devoid of all photoreceptor cells and is therefore the “blind spot”. Figure reproduced with permission by Cmglee (Own work) [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0) or GFDL (http://www.gnu.org/copyleft/fdl.html)], via Wikimedia Commons.

There is much to learn with regard to foveal pit formation, with the majority of studies being limited to the use of preserved human prenatal eyes or primate whole-mounts. Titrated thymidine labelling in utero in the *Macaca* monkey revealed that retinal cell type differentiation is sequential (La Vail, Rapaport and Rakic, 1991). The first cell types to differentiate in the retina are the ganglion cells, horizontal cells and cone photoreceptor cells, followed by the amacrine cells, bipolar and rod photoreceptors (La Vail, Rapaport and Rakic, 1991). There appears to be no link between the birth sequence of retinal neurons and the lamination of the retinal layers (La Vail, Rapaport and Rakic, 1991; Hendrickson, 2016; Amini, Rocha-Martins and Norden, 2018). Retinal development begins centrally in the region of the macula before extending in a foveal-to-peripheral retina gradient (La Vail, Rapaport and Rakic, 1991). The peripheral retina is fully developed before the posterior pole, therefore explaining how the fovea is immature at the time of birth and continues to mature during childhood (Hendrickson *et al.*, 2012).

1.8 Abnormalities of the fovea and/or optic chiasm

The development of the fovea involves a complex sequence of signalling in a foveal-to-peripheral pattern. The decussation of the RGC axons at the optic chiasm is vital for normal binocular vision, ensuring that visual information is correctly mapped to the higher visual centres (Figure 1-10). The development of the central retina and decussation patterns at the optic chiasm have been linked in literature, as diseases caused by defects to both structures share similar abnormalities.

1.8.1 FHONDA syndrome

The name FHONDA syndrome (foveal hypoplasia, optic nerve decussation defects and ASD in the absence of albinism), was first coined in 2013 (Al-Araimi *et al.*, 2013). This rare disease was notable for its striking phenotypic similarities to albinism (see Section 1.8.5), as all patients had foveal hypoplasia, and (those that were tested) had chiasmal misrouting, although all patients had normal pigmentation (Poulter *et al.*, 2013). The detection of chiasmal misrouting in a patient generally leads to a diagnosis of albinism (Figure 1-10), and so this novel combination of foveal hypoplasia, chiasmal misrouting, but no abnormalities in pigmentation was extremely interesting. The syndrome was first investigated by Maria van Genderen in 2006 in a study of three patients with this unique phenotype; a sib-pair with consanguineous parents and a separate non-consanguineous case who also had Kartageners syndrome (van Genderen *et al.*, 2006).

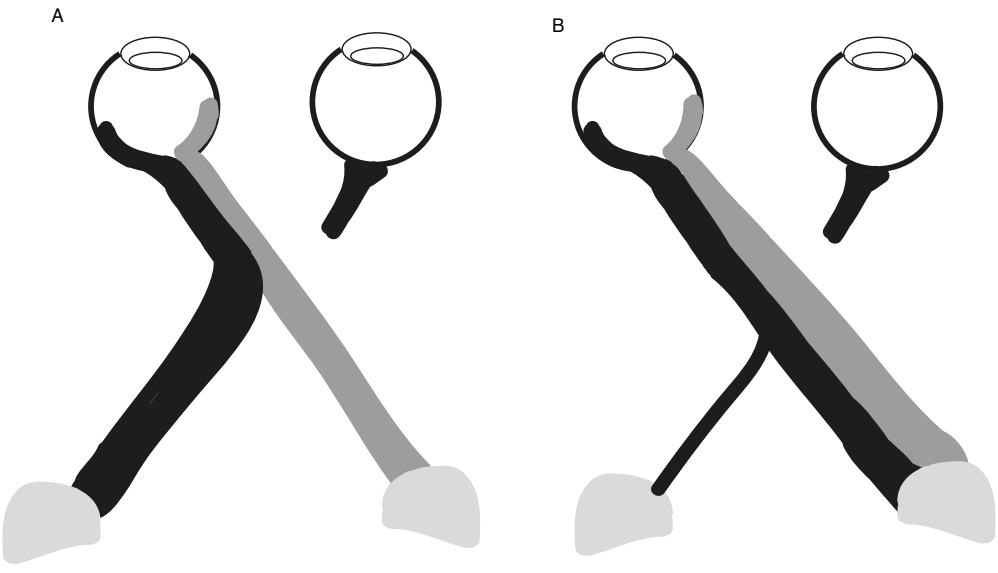


Figure 1-10 Schematic illustration of the axon guidance at the optic chiasm.

Schematic comparing A the normal visual pathway with approximately equal number of axons projecting to the contralateral and ipsilateral hemispheres and B the optic nerve misrouting seen in patients with albinism and FHONDA syndrome, whereby there is a shift in the line of decussation into the temporal retina, resulting in a greater number of axons projecting contralaterally. Image adapted with permission from Copyright Clearance Centre on behalf of Springer Nature: Eye (Neveu and Jeffery, 2007).

FHONDA has since been shown to be caused by recessive mutations in *SLC38A8* (Poulter *et al.*, 2013), to be discussed in more detail in the following section. Due to the fact that foveal hypoplasia and optic nerve misrouting are usually main features of albinism, it has been suggested that melanin pigment is required for the correct development of the fovea and optic chiasm (Creel, Summers and King, 1990). However, most importantly, none of the patients reported in the literature with foveal hypoplasia due to mutations in *SLC38A8* had any signs of reduced pigmentation or albinism (Poulter *et al.*, 2013). To date, as only two of the nine families reported had posterior embryotoxon and Axenfeld anomaly, it should be noted that ASD is a variable feature (Pal *et al.*, 2004; Al-Araimi *et al.*, 2013; Poulter *et al.*, 2013; Perez *et al.*, 2014; Toral *et al.*, 2017). Similarly, one family had the additional phenotype of ocular coloboma and microphthalmia, with the features variable within the family (Poulter *et al.*, 2013). It is currently unclear whether these additional variable features are caused by *SLC38A8* mutations, or if they are unrelated (Poulter *et al.*, 2013). All patients studied in this thesis had a minimum phenotype of foveal hypoplasia and no signs of pigmentation loss, as recorded by the referring clinician.

1.8.2 *SLC38A8*

Prior to the work undertaken in this project, Pal and colleagues had documented a phenotype of recessive foveal hypoplasia and ASD in a large Pakistani

consanguineous family (Pal *et al.*, 2004). In the study by Pal et al, linkage analysis excluded *PAX6*, the only known isolated foveal hypoplasia gene, suggesting the involvement of another defective gene. Whole genome linkage mapped the mutated gene to a 6.5-Mb locus on chromosome 16q23.2-24.2 with a logarithm of the odds (LOD) score of 5.51 (Pal *et al.*, 2004).

Van Genderen and colleagues had also separately reported foveal hypoplasia and optic nerve misrouting in the absence of pigmentation loss in three patients (van Genderen *et al.*, 2006). The two teams collaborated to see if their patients had the same disorder. Their hypothesis turned out to be true, as further clinical evaluation showed that the large Pakistani family investigated by Pal and colleagues had optic nerve decussation defects, ASD and no evidence of albinism (Pal *et al.*, 2004). Similarly, the three cases reported in the van Genderen study all had ASD (posterior embryotoxon, whereby the Schwalbe ring, or outer limit of the corneal epithelium layer, is abnormally thickened and misplaced) (van Genderen *et al.*, 2006). Furthermore, linkage analysis of the sib pair reported in the van Genderen study confirmed that the defective gene mapped to the same chromosome 16q locus mapped previously in the Asian family. This helped to refine the size of the locus to 3.1 Mb (Al-Araimi *et al.*, 2013). As these combined visual defects had previously only been found in association with albinism, this suggested a mechanism independent of the melanin biosynthesis pathway.

Two interesting cases were identified in the study by Al-Araimi et al. One of the patients studied was a Northern European female with non-consanguineous parents

who was previously diagnosed with Kartagener syndrome (MIM 244400) (van Genderen *et al.*, 2006). Her mother was found to be heterozygous for the mutation, and therefore a carrier, but no mutation was identified in the father. Genotyping confirmed paternity and revealed that the father had a large deletion in the region of the mutation, confirming the proband's genotype as hemizygous for the missense mutation in *SLC38A8* (Poulter *et al.*, 2013). This was not the first report of a deletion in this region of the genome. In another patient with Kartagener disease a 640-kb deletion was identified encompassing both *SLC38A8* (chr16: 84,043,272-84,075,762) and the gene *Dynein, axonemal assembly factor 1* (*DNAAF1)* (aka *LRRC50*) (MIM 613190) (chr16: 84178065-84211524) was reported (Loges *et al.*, 2009). Such deletions are likely to be missed by Sanger sequencing. The second interesting case was a patient with consanguineous parents in whom single nucleotide polymorphism (SNP) genotype analysis highlighted a large region of homozygosity spanning *SLC38A8*, although screening of the gene failed to detect the causative mutation, indicating the presence of mutations at the locus not within the coding sequence and splice sites of *SLC38A8*. This patient was subsequently analysed in this study (individual F1310).

To date, there are thirteen published mutations in *SLC38A8* in patients with foveal hypoplasia (Poulter *et al.*, 2013; Perez *et al.*, 2014; Toral *et al.*, 2017). *SLC38A8* is a 10-exon gene located on 16q23.3–24.1 (chr16: 84,043,272-84,075,762; UCSC Genome Browser, RefSeq annotation, GRCh37). *SLC38A8* encodes a 435 aa protein of the SLC38 sodium-coupled neutral amino acid transporter (SNAT) family of proteins (Mackenzie and Erickson, 2004) (Figure 1-11). The use of voltage-clamp experiments in *Xenopus laevis* oocytes has shown that SLC38A8 (SNAT8) is an

Na^+-dependent transporter of L-glutamine, L-alanine, L-arginine, L-histidine and L-aspartate (Hägglund *et al.*, 2015). Although the function of SLC38A8 has not been extensively studied, it is most similar to SLC38A7 (SNAT7), sharing a close evolutionary relationship and 42% sequence identity in the human protein sequences (Sundberg *et al.*, 2008; Schiöth *et al.*, 2013; Hägglund *et al.*, 2015). This has been further supported by proximity ligation assays, showing that SNAT8 and SNAT7 are expressed in the same cell and subcellular regions and are potentially interacting partners (Hägglund *et al.*, 2015). Reports have noted difficulty in the structural modelling of SLC38A8 as the channel has <30% sequence homology to other structural templates in the protein database (Toral *et al.*, 2017). There are no alternative isoforms of SLC38A8 mapped to date.

The transporter is highly expressed in neuronal tissues, especially in the thalamus, hypothalamus, amygdala and pons. It is expressed in both excitatory and inhibitory neurons, as investigated in the mouse using *in situ* hybridisation and immunochemistry (Hägglund *et al.*, 2015). SLC38A8 localises to the cell body, possibly the membrane and the axons of both GABAergic and glutamatergic neurons in the brain and spinal cord, and is also located throughout the retina, with strong expression in the photoreceptor layer (Poulter *et al.*, 2013; Hägglund *et al.*, 2015). SLC38A8 is known to preferentially transport glutamine and it has been suggested that this serves as a carbon energy source for the retinal tissue (Voaden *et al.*, 1978), where it also functions as an intermediate metabolite for the formation of glutamate and gamma-amino butyric acid (GABA). Glutamate is known to be the primary excitatory neurotransmitter of the visual pathway, with glutamatergic neurons

in the retina including the photoreceptors, bipolar cells and RGCs. In contrast, the GABAergic neurons include the horizontal and amacrine cells (Massey, 1990).

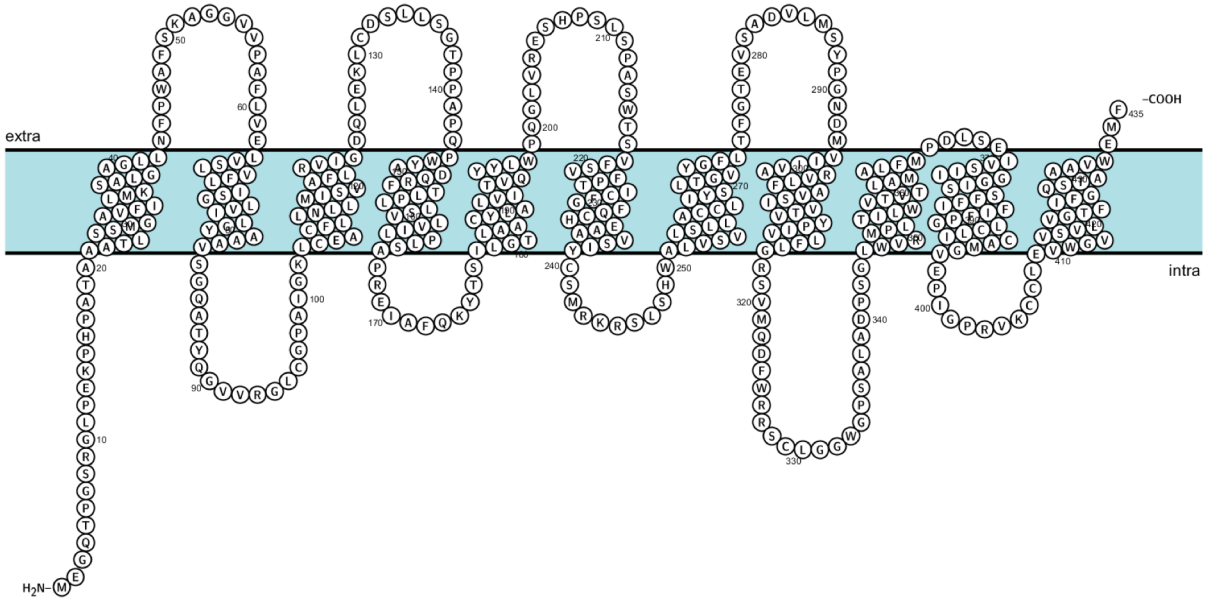


Figure 1-11 Protein schematic of SLC38A8.

SLC38A8 is a 10-exon gene encoding a 435-amino acid transmembrane protein with 11 transmembrane domains. The schematic was generated using the SMART (http://smart.embl-heidelberg.de) and Protter protein tools (http://wlab.ethz.ch/protter/start/) and the consensus is shown as generated by Protter (Section 2.4.10.2).

1.8.3 Isolated foveal hypoplasia

Foveal hypoplasia (MIM 136520) is a congenital disorder characterised by very poor visual acuity caused by a poorly defined foveomacular region and blunted foveal reflex (Curran and Robb, 1976; Oliver *et al.*, 1987). When the normal eye is examined using ocular coherence tomography (OCT), the foveal centre, known as the central light reflex, can be identified as the deepest point of the foveal pit, as it reflects the light from the ophthalmoscope. The OCT assessment can either be

recorded as positive, if the foveal pit is present, or absent (blunted), as seen in cases with foveal hypoplasia. A structural grading system (Grade 1-4) for foveal hypoplasia has been developed, based at the stage at which foveal development ceased, with Grade 1 ranking as highest visual acuity (Thomas *et al.*, 2011). Foveal hypoplasia has been reported as an isolated disorder in a few rare cases (Oliver *et al.*, 1987; Curran and Robb, 1976; Azuma *et al.*, 1996; Al-Saleh, Hellani and Abu-Amero, 2011; Saffra *et al.*, 2012; Karaca *et al.*, 2014; Giocanti-Auregan *et al.*, 2014).

1.8.4 *PAX6*

PAX6 is a transcription factor with a multitude of roles in the developing embryo. It is known as the master regulator in eye development and is highly conserved across species, known as *eyeless* (*ey*) in Drosophila (Quiring *et al.*, 1994), *small eye* (*sey*) in the mouse (Hill *et al.*, 1991) and *PAX6* in humans (Ton *et al.*, 1991). Using a Gal4 system to induce ectopic eyes on the antennae, wings and legs of the fly (Halder, Callaerts and Gehring, 1995), the authors were able to show that *PAX6* represents a master switch in the development of the eye.

The PAX6 protein sequence has recognizable domains; the paired domain (128 aa) and the homeodomain (61 aa) separated by a linker region (78 aa) and a proline, serine, threonine-rich C-terminal region (PST) (152 aa) (Ton *et al.*, 1991; Glaser, Walton and Maas, 1992). *In situ* hybridization with a *PAX6* (*sey*) probe in murine eyes has shown expression in the developing CNS and also corneal epithelium, lens and retina, all ocular tissues originating from ectoderm and neuroectoderm (Walther and Gruss, 1991). Heterozygous mutations in *PAX6* cause aniridia (Ton *et al.*, 1991;

Jordan *et al.*, 1992), which is a congenital disorder characterized by the partial or complete absence of the iris. The phenotypic manifestations usually present within 6 weeks of birth and include corneal opacity, glaucoma, cataract and both foveal and optic nerve hypoplasia (Shaw, Falls and Neel, 1960). More rarely, mutations in *PAX6* are associated with microphthalmia, corneal cataracts, macular and foveal hypoplasia and also Peters anomaly, a form of ASD (Glaser, Walton and Maas, 1992; Glaser *et al.*, 1994; Hever, KA and V., 2006; Azuma *et al.*, 1996; Hanson *et al.*, 1999). Anophthalmia can be caused by homozygous loss of PAX6 function in humans and mice (Hill *et al.*, 1991; Glaser *et al.*, 1994).

A number of genes have been implicated in diseases involving foveal hypoplasia and *PAX6* (MIM 607108) mutations in particular are known to underlie both isolated foveal hypoplasia (Azuma *et al.*, 1996) and foveal hypoplasia with anterior segment anomalies (Hanson *et al.*, 1999), in a pattern of dominant inheritance (MIM 136520). *PAX6* acts on target genes through its DNA-binding paired domain. *PAX6* is alternatively spliced into 2 isoforms, PAX6 and PAX6(5a) which differ only in the structure of their paired domain (Epstein *et al.*, 1994b). The alternatively spliced exon 5a has a 14-amino-acid insertion in the N-terminal DNA-binding paired domain, which serves to modify target specificity by changing the DNA-binding ability of PAX6 (Kozmik, Czerny and Busslinger, 1997). Investigation of foveal formation has suggested a role of the alternatively spliced PAX6(5a) in the developmental cascade leading to high acuity vision (Azuma *et al.*, 2005). In the study by Azuma and colleagues, the overexpression of PAX6(5a) induced ectopic differentiation of well-formed retinal layers in the chick eye, compared to the effect of *PAX6* overexpression inducing less mature retinal structures. This, alongside studies of

mutations identified in the splice-variant region, including p.(Val54Asp) and p.(Arg128Cys), has led to the hypothesis that the *PAX6* isoform containing exon 5a acts as a molecular switch (Azuma *et al.*, 1996; Azuma *et al.*, 1999).

1.8.4.1 The Leiden Open Variant Database (LOVD) of *PAX6* mutations

The LOVD *PAX6* database (http://lsdb.hgu.mrc.ac.uk/home.php?select_db=PAX6) currently contains 472 unique DNA variants, with approximately 65% of mutations in the coding region accounted for by substitutions (last accessed 08/08/17). The majority of reported mutations lie in the paired domain and the most common mutations causing an aniridia phenotype are premature termination mutations, while non-aniridia phenotypes are more commonly caused by missense mutations (Hingorani *et al.*, 2009; Tzoulaki, White and Hanson, 2005).

1.8.5 Albinism

Albinism is the term used to describe a group of rare genetic disorders causing abnormalities in melanin biosynthesis and transport (Kinnear, Jay and Witkop, 1985). Reduced visual acuity (usually 20/60 to 20/400), colour vision impairment, various degrees of congenital nystagmus, iris translucency, photophobia, chiasmal misrouting and foveal hypoplasia are all clinical ophthalmic manifestations seen in albinism patients (Apkarian, 1996; Käsmann-Kellner and Seitz, 2007).

The synthesis and distribution of melanin gives rise to the visible pigmentation in mammals in the eyes, hair bulb and skin. In the eye, the synthesis of melanin pigment molecules takes place in the RPE, more specifically in the melanosomes

(membrane-bound organelles) of melanocyte pigment cells (Bonaventure, Domingues and Large, 2013). The general pathway of melanin biosynthesis begins with tyrosinase, which is fundamental as it catalyses the first two steps in the pathway; hydroxylation of tyrosine to DOPA, and the oxidation of DOPA to DOPAquinone, before further steps resulting in the production of melanin (del Marmol and Beermann, 1996). This step is the same for the production of the two forms of melanin; eumelanins (the black and brown forms) and phaeomelanins (the yellow to red forms). Pheomelanin is generated after DOPAquinone and the amino acid cysteine combine to produce cysteinylDOPA. Many proteins are involved in the generation of eumelanin, including the tyrosinase-related protein 1 and 2 (TRP-1 and TRP-2), OCA2 (also known as the P-gene or P-protein), the membrane-associated transporter protein (MATP) and the ocular albinism protein 1 (OA1). It is mutations in these genes that result in the different forms of albinism.

1.8.5.1 Non-syndromic albinism phenotypes

Albinism is characterised by hypopigmentation caused by an abnormality in melanin synthesis/transportation and there is a clear lack of pigmentation in the fundus of individuals with albinism, as shown in Figure 1-12. Albinism can be classified as oculocutaneous albinism (OCA), whereby individuals have complete absence of pigment in their skin, hair and eyes or ocular albinism (OA), whereby there is only a lack of pigmentation in the eyes (Figure 1-12). Both OCA and OA have proven genetic heterogeneity (Carden *et al.*, 1998; Biswas and Lloyd, 1999; Oetting and King, 1999). The prevalence of all forms of OCA worldwide is estimated to be 1/17,000 newborns (range 1:10,000-20,000), affecting all ethnicities (Gargiulo *et al.*, 2011; Grønskov, Ek and Brondum-Nielsen, 2007; Grønskov *et al.*, 2009; Hutton and

Spritz, 2008a; Hutton and Spritz, 2008b; Oetting and King, 1999; Oetting, 2006; Rundshagen *et al.*, 2004; Rooryck *et al.*, 2008; Zuhlke, Stell and Kasmann-Kellner, 2007). There is much phenotypic overlap between the forms of OCA. OCA1A is the most severe type, with those affected completely lacking pigmentation throughout their lifetime. Other milder forms of OCA do show gradual pigment accumulation, including OCA1B, OCA2, OCA3 and OCA4 (Grønskov, Ek and Brondum-Nielsen, 2007). It is important to note that subtle pigmentation defects can make the diagnosis of albinism problematic in some populations. This means that a clinical diagnosis of albinism is often based solely on the presence of chiasmal misrouting defects using visually evoked potential analysis (Dorey *et al.*, 2003). This misrouting is a key characteristic of albinism, leading to strabismus and reduced stereoscopic vision (Grønskov, Ek and Brondum-Nielsen, 2007).

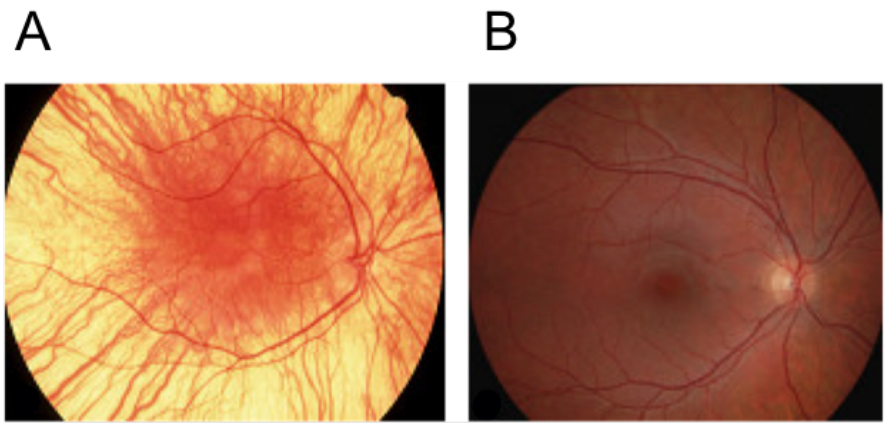


Figure 1-12 The lack of pigmentation in an albino eye.

A Fundus photograph of an individual diagnosed with OCA. B Fundus image of a healthy individual. Figure reproduced with permission from (Grønskov, Ek and Brondum-Nielsen, 2007).

Individuals with OA typically present with all the ocular and visual manifestations of albinism, similar to patients with OCA, but have normal hair and skin (Oetting,

Summers and King, 1994). Ocular albinism 1 (OA1) (MIM 300500) is the most common form, affecting approximately 1/60,000 males, and is characterised by nystagmus, impaired visual acuity, iris transillumination, fundus hypopigmentation and macular hypoplasia. Åland Island eye disease (AIED) (MIM 300600) has been put forward as a form of albinism (OA2) characterised by hypopigmentation and decreased visual acuity. However patients do not present with the classical albinism phenotype of optic nerve misrouting (Forsius and Eriksson, 1964).

1.8.5.2 Syndromic albinism phenotypes

There are a number of syndromes which share many of the distinct clinical phenotypic features of individuals with albinism. The most similar is Hermansky-Pudlak syndrome (HPS), an autosomal recessive disorder of lysosome-associated organelles, characterised by OCA, bleeding tendency and ceroid deposition (Izquierdo, Townsend and Hussels, 1995; Wei and Li, 2013). Griscelli syndrome has three phenotypic subtypes that are determined based on the gene mutated, including pigment deficiency, neurological impairments and recurrent infections (Griscelli *et al.*, 1978; Pastural *et al.*, 1997; Ménasché *et al.*, 2000; Kurugol *et al.*, 2001; Westbroek *et al.*, 2012). Another related rare autosomal recessive disorder is Chediak-Higashi syndrome (CHS), which causes enlargement of intracellular lysosomes and is characterised by hypopigmentation, immunodeficiency, bleeding tendency and peripheral neuropathy (Barbosa *et al.*, 1997; Kaplan, De Domenico and Ward, 2008; Lozano *et al.*, 2014).

1.8.5.3 Genetics of OCA

A brief overview of the clinical features and genes involved in the different forms of OCA are presented in Table 1-1. Complete loss of or reduction in tyrosinase (TYR) activity causes OCA1A and OCA1B respectively (Grønskov, Ek and Brondum-Nielsen, 2007). OCA1A is the most severe form of OCA, with patients never developing any melanin pigmentation, most commonly caused by compound heterozygous mutations in *TYR*. In contrast, patients with OCA1B have some residual enzyme activity (Grønskov, Ek and Brondum-Nielsen, 2007). OCA2, caused by mutations in the *OCA2* gene and OCA4, caused by mutations in *SLC45A2 (MATP),* have similar phenotype severity (Table 1-1). Mutations in *SLC45A2* and OCA2 cause misrouting of tyrosinase in the melanin biosynthesis pathway (Costin *et al.*, 2003; Cullinane *et al.*, 2011). OCA3 is caused by mutations in the *TYRP1* gene. This encodes a protein which is responsible for melanosome structure and is involved in the conversion of tyrosine to DOPA, meaning that it is an essential co-factor for tyrosinase activity (Sarangarajan and Boissy, 2001; Kamaraj and Purohit, 2014). Less is known about the OCA5-7 forms. Two genes have been implicated, but it is clear there are more to find (Grønskov *et al.*, 2013; Kausar *et al.*, 2013; Montoliu *et al.*, 2013; Wei *et al.*, 2013).

Type of OCA (MIM phenotype reference)	Prevalence	Severity of phenotype	Gene/locus (MIM)	Reference
OCA1A (203100)	1:40,000 (African-Americans)	Most severe, Full iris transillumination, VA 1/10 or less	*TYR* (606933)	(Grønskov, Ek and Brondum-Nielsen, 2007)
OCA1B (temperature sensitive/yellow albinism) (606952)	As OCA1A	Moderate-severe, Full/partial iris transillumination, VA 2/10	*TYR* (606933)	(Grønskov, Ek and Brondum-Nielsen, 2007)
OCA2 (203200)	1:36,000 (white Europeans) 1:3,900-10,000 (Africans)	Mild-moderate, Mostly partial iris transillumination, VA 3/10	*OCA2* (*P gene*) (203200)	(Grønskov, Ek and Brondum-Nielsen, 2007)
OCA3 (203290)	Rare (white Europeans, Asians) 1:8,500 (Africans)	Mild-normal, African individuals have red hair and red/brown skin, Mild/absent nystagmus, VA not always detectable	*TYRP1* (115501)	(Grønskov, Ek and Brondum-Nielsen, 2007)
OCA4 (606574)	Rare (white Europeans) 1:85,000 (Japanese)	As OCA2, Increased optic nerve head hypoplasia	*MATP* (*SLC45A2*) (606202)	(Grønskov, Ek and Brondum-Nielsen, 2007) (Hayashi and Suzuki, 2005)
OCA5 (615574)	Unknown	Nystagmus, impaired VA.	Mapped to 4q24	(Montoliu *et al.*, 2013; Kausar *et al.*, 2013)
OCA6 (113570)	Unknown	Full/partial iris transillumination, nystagmus. No defects in platelet dense granules. Impaired VA.	*SLC24A5* (609802)	(Wei *et al.*, 2013)
OCA7 (615179)	Unknown	Full/partial iris transillumination, nystagmus, reduced VA.	*c10orf11* (614537)	(Grønskov *et al.*, 2013)

Table 1-1 Overview of the different forms of OCA.

Brief overview of the prevalence, gene/locus involved and clinical features including the phenotype severity.

1.8.5.4 OA phenotype

Ocular albinism, as described above and as the name suggests, affects only the visual system. X-linked mutations in the gene *GPR143* (MIM 300808) are known to cause OA1 (Bassi *et al.*, 1995). Mutations in the gene *CACNAF1* (MIM 300110) have

been identified as a cause of AIED, a disease with great phenotypic similarity to albinism (Wutz *et al.*, 2002).

1.8.5.5 Genetics of syndromic albinism

To date, nine different forms of HPS and one form of CHS are currently known. HPS is caused by mutations in the following genes: *HPS1* (HPS1) (MIM 604982) (Oh *et al.*, 1996), adapter-related protein complex 3 beta 1 (*AP3B1)* (HPS2) (MIM 603401) (Dell'Angelica *et al.*, 1999), *HPS3* (HPS3) (MIM 606118) (Anikster *et al.*, 2001), *HPS4* (HPS4) (MIM 606682) (Suzuki *et al.*, 2002), *HPS5* (HPS5) (MIM 607521) (Zhang *et al.*, 2003), *HPS6* (HPS6) (MIM 607522) (Zhang *et al.*, 2003), *dystrobrevin binding protein 1* (*DTNBP1)* (HPS7) (MIM 607145) (Li *et al.*, 2003), *biogenesis of lysosomal organelles complex 1 subunit 3* (*BLOC1S3)* (HPS8) (MIM 609762) (Starcevic and Dell'Angelica, 2004) and *biogenesis of lysosomal organelles complex 1 subunit 3* (*BLOC1S6)* (HPS9) (MIM 604310) (Huang, Kuo and Gitschier, 1999). CHS is associated with mutations in the gene *lysosomal trafficking regulator* (*LYST)* (MIM 606897) (Perou *et al.*, 1996).

1.8.6 Congenital nystagmus

Congenital nystagmus (MIM 310700) is an idiopathic condition characterised by increased amplitude of rhythmic bilateral, conjugate, uniplanar and horizontal eye oscillations (Cogan, 1967; Reinecke, 1997). Individuals often present with a head turn or tilt, mildly decreased visual acuity, strabismus and astigmatism. Congenital nystagmus has an estimated frequency of 1:1,000-1,500 (Mellott *et al.*, 1999; Tarpey

et al., 2006) and is commonly a feature of other ocular diseases, such as OCA (Grønskov, Ek and Brondum-Nielsen, 2007; Hutton and Spritz, 2008a).

1.8.6.1 *FRMD7*

Idiopathic nystagmus is caused by mutations in the X-linked *FERM domain containing 7* gene (*FRMD7)* (MIM 300628) (Tarpey *et al.*, 2006). In males with pathogenic *FRMD7* mutations the disease is fully penetrant, but in females with heterozygous mutations, the penetrance is reported to be around 53% (Thomas *et al.*, 2008). *FRMD7* is expressed in the developing neural retina and also in developing ocular motor structures including the cerebellum and vestibulo-optokinetic system. FRMD7 plays a role in the control of eye movement and gaze stability (Tarpey *et al.*, 2006). Mutations in *FRMD7* have also been identified in cases of foveal hypoplasia (Thomas *et al.*, 2014). Notably, the foveal hypoplasia recorded in cases with *FRMD7* mutations is a milder phenotype (Grade 1), compared to patients with albinism where the majority of patients have a higher grade of foveal hypoplasia (Grades 3/4) (Thomas *et al.*, 2011).

1.9 An introduction to next-generation sequencing technologies

Molecular investigation of patients with genetic disorders has been vastly improved as new sequencing technologies continue to emerge. This is evidenced from the sequencing of the mitochondrial genome in 1981 (Anderson *et al.*, 1981), to the completion of the entire human nuclear genome sequence 20 years later (Lander *et al.*, 2001; Sachidanandam *et al.*, 2001; McPherson *et al.*, 2001). The cost of the first

human genome sequence totalled approximately 3 billion dollars and it took over thirteen years to complete. However in the present day a single genome can be sequenced for under $1000, taking only a few days (Hayden, 2014). Next-generation sequencing (NGS), the massively parallel, simultaneous sequencing of millions of randomly sheared DNA fragments, has many benefits, including greatly speeding up diagnosis time and also providing a cost-effective approach to investigating rare disease.

NGS encompasses both whole exome sequencing (WES) (Section 1.9.1) and whole genome sequencing (WGS) (Section 1.9.2). Illumina short-read sequencing is one of the most popular massively paralleled sequencing methods used in the present day. However, long-read technologies, such as the PacBio (Pacific Biosystems, CA, USA) sequencing system, which does not employ optical signals, are becoming more widely available and open up many new applications. The PacBio platform utilises single-molecule real-time sequencing (SMRT), whereby DNA polymerase molecules bound to a single molecule of DNA are attached to the base of nanometer-sized wells termed zero-mode waveguides (ZMWs) (Eid *et al.*, 2009). The light-propagated excitement of the fluorescently-labelled nucleotides allows real-time continuous DNA synthesis over thousands of bases, as the ZMW is confined to observe only a single nucleotide of DNA (Eid *et al.*, 2009). Ion torrent and Ion proton sequencing semiconductor technology (Life Technologies) relies on pH reading to determine the number of bases added to the final sequencing reads, as the addition of each deoxynucleotide (dNTP) to a DNA polymer releases a single proton.

There are now multiple research projects employing the power of NGS, such as the Born in Bradford cohort study (recruitment from 2007-2010) (https://borninbradford.nhs.uk) and the SPEED study (https://bioresource.nihr.ac.uk/rare-diseases/study-specialist-pathology-evaluating-exomes-in-diagnostics/), which is based at the Cambridge Biomedical centre and serves to investigate inherited retinal dystrophies. The 100,000 Genomes Project is another example of a large NGS project which began in 2013 and aims to develop a diagnostic genomic medical service for the NHS with a long-term aim of providing more effective treatments (https://www.genomicsengland.co.uk/the-100000-genomes-project/).

1.9.1 Whole exome sequencing (WES)

WES is the application of NGS targeting only the coding regions, or exons, of known genes, which covers only ~2% of the human genome. This popular application of NGS provides coverage of >95% of the exons, and it is estimated that 85% of disease-causing mutations lie in these coding regions (Majewski *et al.*, 2011; Botstein and Risch, 2003). The principle of WES is to simultaneously read the sequence of all of the exons in a given human genome, and to read the exonic sequences of multiple individuals using a single flow cell. As an example, the Illumina sequencing experimental workflow is shown in Figure 1-13. Exome sequencing is a capture-based method developed to identify variants in the coding regions of genes. DNA libraries are prepared by randomly shearing target DNA into uniform fragments. Adapters are tagged at each end of the DNA fragments (known as paired-end sequencing) and ligated products amplified using PCR. Exon libraries are pooled for application to the flow cell where complementary oligonucleotides

recognise the adapters. PCR amplification takes place as the adapter-tagged DNA ends bind to the flow cell forming a bridge-like structure. There are multiple cycles of annealing, extension and denaturation resulting in clusters, also called “spots” of DNA. Cleavage within one adapter sequence linearizes the DNA by denaturation and paired-end sequencing takes place by attachment to adjacent oligonucleotides. Sequencing incorporates four fluorescently-labelled reversible dye terminators that are added base-by-base at each round of synthesis, ultimately generating a sequence termed a “read”. As each round of extension proceeds, the nucleotides are incorporated in a step-wise manner and through the use of 3’-modified nucleotides the reaction is chemically terminated. The reaction continues for further rounds and the image of the fluorescing flow cell is recorded, showing which base has been incorporated in each spot. The resulting “reads” can then be later manipulated and aligned to the human reference sequence and to each other for interpretation of variants.

As NGS generates huge quantities of data, the main issues are narrowing down the generated lists of mutations to find the best causal candidate mutation, and also storage costs (Cooper and Shendure, 2011). There are many strategies that can be applied in the analysis of WES datasets after sequencing, such as filtering across multiple affected individuals in the hope of finding novel variants in the same gene/s, filtering among multiple related affected individuals from within a pedigree to look for shared variants, or filtering parent-child trios in the hope of identifying *de novo* changes (Bamshad *et al.*, 2011).

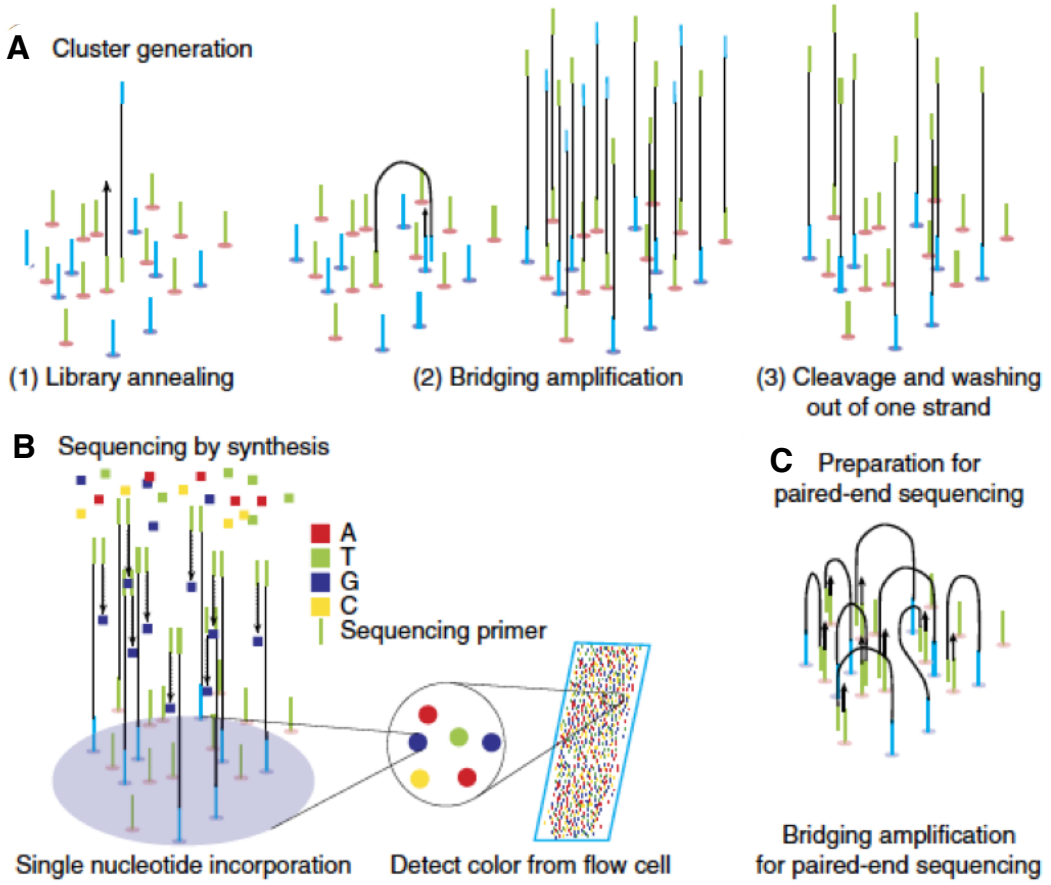


Figure 1-13 The Illumina workflow.

A Cluster generation on the surface of the flow cell. (1) denatured libraries are annealed to the surface of the flow cell at low density. (2) Bridging amplification generates clusters. (3) One strand from the double stranded DNA library is cleaved and washed out for unidirectional sequencing. B Primers for inserts are annealed for the sequencing of the insert DNA. In each sequencing cycle, protected and fluorescently labelled A, T, C and G bases are applied. After the addition of each nucleotide, the sequencing reaction is stopped and the image is taken. Because the newly added nucleotides within each cluster are identical, the signal is high enough to be detected by a light sensor. After the image is taken the protection group and the fluorescent molecules are removed. C When the first-strand sequencing reaction is finished, the synthesised strand is removed and the process is repeated for the opposite strand. Image reproduced by permission from Copyright Clearance Centre on behalf of Springer Nature: Nature Neuroscience (Shin, Ming and Song, 2014).

1.9.2 Whole genome sequencing (WGS)

WGS is the sequencing of the entire DNA sequence of an organism and has been highly successful in the detection of copy-number variation (CNV) and non-coding variation. For example, WGS has detected deletions in *usherin* (*USH2A)* and inverted duplications in *eyes shut homolog* (*EYS)* in patients with RP (Nishiguchi *et al.*, 2013), mutations that would have been overlooked by WES. As the price of sequencing continues to fall and bioinformatics analyses improve, WGS is inevitably becoming the test of choice for many institutions in the present day (Pabinger *et al.*, 2014).

WGS provides an overall view of the human genome, and indeed studies that have compared the outputs from WES/WGS for the same individuals have demonstrated that WGS is also more efficient at detecting mutations in the exome than WES (Belkadi *et al.*, 2015). This is because WGS provides even more coverage as there is no capture step required in the preparation of the sequencing libraries. Interpretation of the variants is more complex for WGS than for WES, with non-coding variants being a particular challenge.

Frequency can be used as a filter, with the gnomAD database now featuring over 15,000 genome sequences to give frequency estimates for intronic and intergenic variants (Lek *et al.*, 2016). In addition, different annotations measure different properties of the mutation under investigation, such as evolutionary conservation, the effect of an amino-acid change on the protein function, or the potential for a non-coding intronic variant to alter splicing,

allowing the user to exclude less likely candidate variants (Ionita-Laza *et al.*, 2016). It is thought that using a combined approach of WGS with RNA studies to investigate variants in non-coding regions will help solve undiagnosed cases (Eisenberger *et al.*, 2013).

1.10 Overview of this study

The aims of this study are as follows:

1. To expand the *SLC38A8* mutation spectrum by screening a novel patient cohort with foveal hypoplasia.
2. Perform WES to discover novel foveal hypoplasia gene(s) in patients with no mutations in the known foveal hypoplasia genes.
3. Use WGS to investigate potential non-coding variation, copy-number variation (CNVs) or structural rearrangements in *SLC38A8*-mutation negative cases.

Chapter 2 Materials and Methods

Room temperature (RT) is in the range of 18-24°C.

All reagents used were purchased from Sigma-Aldrich (Gillingham, Dorset, UK) or Thermo Fisher Scientific (Wilmington, DE, USA), unless indicated otherwise. All primers were ordered from Sigma-Aldrich.

2.1 Materials

2.1.1 1X Tris-ethylenediaminetetraacetic acid (EDTA) (TE) buffer

10 mM tris HCl (pH 8.0)

1 mM EDTA (pH 8.0)

2.1.2 50 X Tris-acetate-EDTA (TAE) buffer for electrophoresis

2 M tris base

0.97 M glacial acetic acid

50 mM EDTA (pH 8.0)

2.1.3 6 X gel-loading dye

10 mM tris-HCl (pH 7.6)

60% weight/volume (w/v) glycerol

60 mM EDTA

0.15% (w/v) Orange G

0.03% (w/v) xylene cyanol FF

2.1.4 Luria-Bertani (LB) broth

1% (w/v) tryptone

0.5% (w/v) yeast extract

1% (w/v) NaCl

For plates, 2% agar was added to LB broth.

ampicillin (AMP), kanamycin (KAN) and spectinomycin (SP) were added at 50 µg/ml.

2.1.5 Super optimal broth with catabolite repression (S.O.C medium)

2% (w/v) bacto-tryptone

0.5% (w/v) bacto-yeast extract

10 mM NaCl

2.5 mM KCl

10 mM $MgCl_2$

10 mM $MgSO_4$

20 mM glucose

2.2 Patients

Informed consent was obtained from all subjects tested. Ethical approval was provided by the Leeds East Teaching Hospitals NHS Trust Research Ethics Committee (Project numbers 03/362 and 17/YH/003), adhering to the Declaration of Helsinki. Local, national and international ophthalmologists and clinical geneticists clinically evaluated the patients and their families, recruited them to the study and obtained DNA via blood or saliva sample. When patients were available for sampling, variants were verified by taking fresh blood samples from the patient and extracting the DNA. Patients were included in the study if they had a minimum phenotype of foveal hypoplasia without albinism (no iris transillumination or loss of pigmentation in the hair, skin or eyes), as reported by the referring ophthalmologist or clinical geneticist. Not all patients had undergone visual evoked potential assessment to resolve if they had optic nerve decussation defects, as visual evoked potential examination is rarely performed in patients. ASD was not seen in all patients, as it is a variable feature of FHONDA.

2.2.1 Patient DNA extraction

DNA was extracted from peripheral blood lymphocytes by the Leeds Genetics Laboratory (St James's University Hospital, Leeds, UK) using a standard salt precipitation protocol. DNA obtained from saliva samples was collected using Oragene DNA collection kits (DNA Genotek Inc., Ottawa, Canada). DNA was extracted according to manufacturer's instructions and pellets were redissolved in 1 X TE buffer (2.1.1).

2.3 Methods

2.3.1 Primer design criteria

Oligonucleotide primers were either designed manually by eye or by using online design tools (2.3.1.1). All primers were purified by desalting unless stated otherwise.

In all instances, the following key criteria were deemed essential for optimal PCR. An annealing temperature (T_A) of 55-70°C, primer length between 17-27 bp and a GC content of approximately 50%. Runs of 4 or more mononucleotides and complementary 5' and 3' ends were avoided to prevent secondary structure formation. Using the BLAT tool through UCSC Genome Browser (2.4.2), primer sequences were checked for specific binding to the region of interest and that no common or known single-nucleotide polymorphisms (SNPs) were present in the sequence.

When screening genes, primers were designed to amplify the coding regions plus a minimum of 50 bp flanking 5' and 3' exonic regions. When possible, primers were designed to amplify products of approximately 500 bp. Larger exons (greater than 750 bp) were split into multiple smaller amplicons with overlaps of a minimum of 50 bp.

For manual calculation of $T_{A,}$ the following equation was applied, where (A+T) represents the total number of adenine and thymine residues in the primer and (G+C) represents the number of guanine and cytosine residues:

$$T_A = 2(A+T) + 4(G+C)$$

For the PCR cycling program, the T_A applied was usually -5°C from the output of the above equation.

2.3.1.1 Primer design software tools

In some instances, primers were designed using online tools including ExonPrimer software (http://ihg.gsf.de/ihg/ExonPrimer.html) available through UCSC Genome Browser (2.4.2). Alternatively, Primer3 was used (http://primer3.ut.ee); however these primers required additional checks using UCSC Genome Browser to ensure no known common SNPs were present in the sequence. Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was also used in some instances.

2.3.2 DNA quantification

DNA was quantified using two methods; The NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific) or the Qubit 1.0 fluorometer (Thermo Fisher Scientific).

The NanoDrop determined the concentration of 1 μl of DNA by measuring absorbance at 260 nm. The NanoDrop was always normalised against TE

buffer or de-ionised water (dH_2O) prior to taking readings, depending on how the sample had been prepared. Alternatively, DNA was quantified using the Qubit dsDNA broad range or high-sensitivity kit, according to the manufacturer's instructions. Qubit reagent was added to Qubit buffer (1:200), with 190 μl of the working solution and 10 μl Qubit standards added to thin-walled, transparent Qubit assay tubes. For test samples, 199 μl working solution was added to Qubit assay tubes with 1 μl test DNA. Tubes were mixed by vortexing and left to stand at RT for 2 minutes. The machine was calibrated using the two standards and the DNA subsequently quantified.

2.3.3 Polymerase chain reaction (PCR)

2.3.3.1 Standard PCR

Typically, standard PCR reactions were performed in a final volume of 25 μl with 25 ng of genomic DNA. In a standard reaction, the following reagents were used at the noted concentrations: 10 X PCR Buffer (Invitrogen) and 50 mM KCl, 10 pmol forward and reverse primers, 20 μM of each dNTP (dATP, dTTP, dCTP, dGTP), 1 unit of *Taq* DNA polymerase (Invitrogen), 1 mM or 1.5 mM magnesium chloride ($MgCl_2$) and dH_2O.

Standard reactions proceeded as follows: denaturation was performed at 95°C for 2 minutes, followed by 35 cycles of 94°C for 30 seconds, an annealing step of 55-65°C (T_A-dependent) for 30 seconds, and an extension step of 72°C for 30 seconds. Final extension was performed at 72°C for 10 minutes.

Some reactions required touchdown-PCR (TD-PCR) conditions, for example, for TD-PCR at 60.5°C, denaturation was performed at 95°C for 2 minutes, followed by 95°C for 30 seconds and an initial annealing step of 68°C for 30 seconds in stage 1 (decreasing by 0.5°C for each of the 13 cycles) before an annealing step at 60.5°C for 30 seconds which continued in stage 2 for the remaining 25 cycles. Final extension was performed at 72°C for 10 minutes. All PCR products were visualised by agarose gel electrophoresis (2.3.4).

2.3.3.2 HotShot MasterMix (HSMM) PCR

Reactions were carried out in a final volume of 10 μl with 25 ng of genomic DNA using the following reagents at the noted concentrations: 1X HotShot Diamond PCR Master Mix (Clent Life Science, Stourbridge, UK), 250 nM forward and reverse primers and dH_2O.

Following manufacturer's instructions, DNA was amplified using a hot start cycle. Initial denaturation took place at 95°C for 10 minutes, followed by 34 additional cycles of 94°C for 30 seconds, an annealing step of 55-65°C for 30 seconds, and an extension step of 72°C for 30 seconds. Final extension was performed at 72°C for 10 minutes. PCR products were visualised by agarose gel electrophoresis (2.3.4).

2.3.3.3 GC-rich or non-standard PCR

In some instances, 50 μM 7-deaza-dGTP was substituted for 25% dGTP. Dimethyl sulphoxide (DMSO) at 3-10%, betaine at 1-1.3 M final reaction

concentration and kit enhancers (as recommended by manufacturer's instructions), were added as PCR enhancing agents in some instances, either in combination or separately.

A range of *Taq* DNA polymerases were used throughout the project, including: Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen), Platinum *Pfx* DNA Polymerase (Invitrogen), Q5 GC-rich Polymerase (NEB, Ipswich, MA), Phusion High Fidelity Polymerase (NEB, Ipswich, MA) and KAPA HotStart Polymerase (KAPA Biosystems, Wilmington, US), all according to the manufacturers' instructions.

2.3.4 Size fractionation by agarose gel electrophoresis

DNA was fractionated by size using agarose gel electrophoresis. 0.8- 3% agarose gels were prepared using molecular biology grade agarose powder (Bioline, London, UK), dissolved in 1 X TAE buffer (2.1.2). 4μl Midori Green Advance (Nippon Genetics, Germany) was added per 100 ml agarose and TAE Buffer solution. Samples were loaded alongside the molecular size standards EasyLadder I (Bioline, London, UK), 1 Kb Ladder (NEB, Ipswich, MA) or 1 Kb Ladder (Promega, Madison, US). Electrophoresis was performed in a horizontal electrophoresis tank at a constant voltage in the range of 100-120 V/cm until the required separation of fragments was obtained. Gels were visualized on the Bio-Rad gel UV transilluminator system and analysed with Image Lab (v4.0) software (Bio-Rad, Hemel Hempstead, UK).

2.3.5 DNA extraction from agarose gel

DNA bands were excised from agarose gels and DNA was extracted using the QIAquick gel extraction kit (Qiagen), according to the manufacturer's instructions. Briefly, DNA was run on a <2% agarose/TAE/Midori Green Direct gel alongside the molecular size marker EasyLadder I (Bioline, London, UK). DNA bands were visualised using a blue/green LED transilluminator (Nippon Genetics, Germany) and cut from the gel using a sterile scalpel. The gel was weighed in a 1.5 ml Eppendorf tube and was subsequently dissolved in solubilisation buffer at 50°C for 10 minutes. The sample was then mixed with an equal volume of isopropanol to precipitate the DNA. A binding column was used to bind the DNA after passing the sample through via centrifugation at 10,000 x g. All waste was discarded and the DNA bound to the column was washed with ethanol wash buffer and eluted from the column in dH_2O.

2.3.6 Whole genome amplification (WGA) of DNA samples

WGA was performed using the GE Healthcare GenomiPhi V2 Amplification kit (GE Healthcare, Waukesha, USA) and was carried out according to manufacturer's instructions. WGA samples were diluted 1/10 or 1/20 in dH_2O before application in PCRs.

2.3.7 Sanger sequencing

2.3.8 PCR product 'clean up' for sequencing

Prior to Sanger sequencing, PCR template DNA underwent a clean-up step using ExoSap-IT (Affymetrix USB, Santa Clara, USA), containing exonuclease

I and shrimp alkaline phosphatase in a PCR template: ExoSAP-IT ratio of 5:2 for 15 minutes at 37°C, followed by a denaturation step for 15 minutes at 80°C.

2.3.9 Sanger sequencing reaction

PCR products (ExoSAP-IT treated) or plasmid DNA template were prepared according to the manufacturer's instructions (Applied Biosystems) with the modifications of half reaction size and the addition of 1 X BigDye Terminator v3.1 Sequencing Buffer (Applied Biosystems). The sequencing reaction was processed by an initial denaturation step of 96°C for 1 minute, followed by 96°C for 10 seconds, 50°C for 5 seconds and 60°C for 4 minutes for 25 cycles. All temperatures were ramped at 1°C/second.

DNA from the sequencing PCR was precipitated from the reaction mixture prior to sequencing on the ABI3130 Genetic Analyser (Applied Biosystems). 5 μl of 125 mM EDTA and 60μl of 100% ethanol was added to each sample and centrifuged for 30 minutes at 3061 x *g* at RT. The supernatant was removed and pellets were washed with 60 μl of freshly prepared 70% ethanol, and again centrifuged for 15 minutes at 805 x *g* at 4°C. Precipitates were redissolved in 10 μl Hi-Di formamide (Applied Biosystems) and resolved at 60°C using 3730 sequencing buffer and POP7 polymer on the ABI3130 Genetic Analyser (Applied Biosystems) using the default RapidSeq36POP7 module. Sequencing results were analysed on SeqScape (V2.5, Applied Biosystems).

2.3.10 Microsatellite marker genotyping

Genomic DNA was PCR-amplified according to Section 2.3.4.1 using 5' fluorescently labelled forward primers (HEX, FAM or TET). Markers were selected and their genetic locations determined using the UCSC Genome Browser (Section 2.4.2). 1 μl PCR-amplified product was then added to a mixture of 8.5 μl Hi-Di formamide (Applied Biosystems) and 0.5 μl 500 ROX size standard (Applied Biosystems). Fragments were resolved on an ABI3130xl sequencer. Peaks were analysed using GeneMapper software v4.0 (Applied Biosystems) and haplotypes manually called, then checked using Cyrillic 2.1 (www.cyrillicsoftware.com).

2.3.11 Isolation of total cellular RNA from whole human blood

Patient RNA was extracted from EDTA0anticoagulated blood using the QIAamp RNA Blood Mini kit (Qiagen), according to manufacturer's instructions. Whole human blood was mixed with erythrocyte lysis buffer in a 1:5 ratio and incubated for 15 minutes on ice, vortexing twice during the incubation period. The sample was then centrifuged for 10 minutes at 4°C, 400 x *g* before removal of the supernatant. Erythrocyte lysis buffer was added to the pellet 2:1, redissolved and centrifuged for 10 minutes at 4°C, 400 x *g* before removal of the supernatant. Cell lysis buffer was added to pelleted leukocytes and the lysate added to a QIAshredder spin column. The sample was centrifuged for 2 minutes to homogenize and 70% ethanol added to the lysate, mixed by pipetting. Sample was then pipetted into a fresh QIAamp spin column and centrifuged for 15 seconds at 8000 x *g*. The spin column was

then transferred to a fresh collection tube and wash buffer added with centrifugation to membrane-bound RNA. Concentrated wash buffer was subsequently added in two rounds to the QIAamp spin column and centrifuged at 8000 x *g* for 15 seconds, then at 20,000 x *g* for 3 minutes. The final step was removal of filtrate and elution of RNA in RNase-free water.

2.3.12 Complementary DNA (cDNA) synthesis

cDNA was prepared using 0.5 µg of total RNA incubated with 100 ng of random primer (Invitrogen) and DEPC H_2O at 70°C for 10 minutes. Samples were chilled on ice and 5X Moloney Murine Leukemia Virus (MMLV) reverse transcriptase buffer, 0.1 M DTT, 10 mM dNTPs (all Invitrogen) and 40 U RNAsin (Promega, Fitchburg, USA), were added. Reactions were first equilibrated at 37°C for 2 minutes, then 100 U MMLV reverse transcriptase added and incubated for 1 hour at 37°C. Samples were finally heated to 95°C for 5 minutes. All cDNA samples were tested in a control RT-PCR reaction using primers in the gene *p53*, amplifying a product of size 410 bp.

2.3.13 RNA Ligase Mediated Rapid Amplification of cDNA ends (RLM-RACE)

The Ambion FirstChoice RLM-RACE kit (AM1700) was used for the amplification of the 5' UTR of *SLC38A8*. The protocol was followed according to the manufacturer's instructions using 250 ng human brain poly(A)+ RNA (Clontech, Takara Bio).

Only full-length mRNAs were selected for by treating either total or poly(A) RNA with 2 μl calf intestinal phosphatase (CIP) to remove the 5'-phosphate from all molecules which contain free 5'-phosphates (ribosomal RNA, fragmented mRNA, tRNA and any contaminating gDNA). Also included in this reaction mix was 2 μl CIP buffer and nuclease-free water, with incubation at 37°C for 1 hour. Any full-length mRNAs (with cap structure found on intact 5' ends) are unaffected by treatment with CIP. The reaction was terminated using 15 μl ammonium acetate solution, composed of 115 μl nuclease-free water and 150 μl phenol: chloroform. The resulting aqueous phase was transferred to a new tube and 150 μl chloroform added. This was repeated but with the addition of 150 μl isopropanol, before being left on ice for 10 minutes. After centrifugation for 20 minutes the pellet was rinsed with 500 μl 70% ethanol and re-dissolved in 11 μl nuclease-free water.

5 μl of CIP-treated RNA was then mixed with 2 μl tobacco acid pyrophosphatase (TAP) to remove the 5' cap structure from the full-length mRNA, 1 μl TAP buffer and nuclease-free water and incubated for 1 hour at 37°C. A 5'-monophosphate then remained.

T4 RNA ligase-mediated ligation of a synthetic 45bp RNA adaptor to the RNA population was then undertaken. The majority of the decapped, full-length mRNAs will acquire the adapter sequence at the 5' end. 2 μl of CIP/TAP-processed RNA, 1 μl 5' RACE adapter, 1 μl RNA ligase buffer, 2 μl T4 RNA

ligase and nuclease-free water were combined and incubated at 37°C for 1 hour.

To amplify the 5'-end of the transcript, a random-primed reverse transcription reaction and nested PCR were performed. In an RNase-free microcentrifuge, 2 μl TAP-positive adapter-ligated RNA was assembled with 4 μl dNTP mix, 2 μl random decamers, 2 μl MMLV-RT buffer, 1 μl RNase Inhibitor, 1 μl MMLV-RT and nuclease-free water. The samples were mixed and briefly centrifuged at 10,000 x *g*, followed by incubation on a hot block at 42°C for 1 hour.

A nested PCR was then performed using the 5' RLM-RACE products. The kit provided two nested primers that correspond to the 5' RACE adapter sequence, from this point labelled 5' RACE outer primer and 5' RACE inner primer. Two *SLC38A8*-specific nested primers were manually designed for the experiment, from this point labelled 5' *SLC38A8* outer primer and 5' *SLC38A8* inner primer. All primers are listed in Appendix 1.

The outer 5' RLM-RACE PCR reaction mix was assembled in PCR tubes as follows; 5 μl complete PCR Buffer (Ambion SuperTaq Plus kit), 4 μl dNTP mix, 2 μl 5' RACE outer primer, 2 μl 5' *SLC38A8* outer primer, 1.25 U DNA polymerase (Ambion SuperTaq Plus kit) and nuclease-free water. The PCR cycle was performed as follows: denaturation at 94°C for 3 minutes, followed by 35 cycles of 94°C for 30 seconds, the annealing step of 60°C for 30 seconds, and an extension step of 72°C for 3 minutes. Final extension was

performed at 72°C for 7 minutes. The same controls and conditions were used for the inner PCR, using the outer PCR product as the template in the nested PCR. 2 μl of 5' *SLC38A8* inner primer was used for the inner PCR.

Products were analysed by gel electrophoresis on a 3% agarose-TAE gel and visualised (Section 2.3.4).

2.3.14 Molecular cloning

2.3.15 Bacterial cell transformation and culture

For bacterial transformations, the heat-shock technique was applied. Chemically competent *E.Coli* cells were defrosted on ice prior to performing the transformation. Cells were then incubated with DNA on ice for 30 minutes, before undergoing heat-shock at 42 °C for exactly 30 seconds and being returned to the ice for 2 minutes. Cells were then mixed with pre-heated S.O.C medium and incubated at 37 °C for 1 hour, with shaking at 225 rpm. Cells were spread at different volumes on LB-agar plates containing the appropriate antibiotic and incubated at 37 °C overnight.

Single colonies were picked from the agar plates using sterile loops and cultured overnight in 5 ml LB broth with appropriate antibiotic at 37 °C with shaking at 225 rpm. For larger scale plasmid preparation and purification, 1ml of the 5 ml culture was used to inoculate 100 ml LB broth with appropriate antibiotic in a conical flask overnight at 37 °C with shaking at 225 rpm.

2.3.15.1 Site-directed mutagenesis (SDM)

The QuikChange II XL SDM kit (Agilent Technologies) was used to introduce point mutations individually, following the manufacturer's protocol. Primers for the SDM reaction were designed and ordered (Section 2.3.1.1). The mutagenesis reaction was performed by amplifying purified plasmid DNA using the SDM primers, *Pfu* Ultra high-fidelity polymerase, 10 X reaction buffer, dNTP mix, QuikSolution reagent and double-distilled water (ddH_2O) (Agilent Technologies).

Primer pairs designed to introduce point mutations into specific plasmid sequences were designed using the QuikChange Primer Design Programme (http://www.genomics.agilent.com/primerDesignProgram.jsp). It is possible to adapt the online tool to the specific SDM kit to be used in the experiment. The DNA sequence was uploaded to the online tool and the position to be mutated was selected. Primers were designed to be complementary, 25-40 nt in length and were purified by high performance liquid chromatography. When possible, primers were designed to have a GC content of no more than 40%, terminate with a G or C base and have the desired mutation lying central in the primer sequence.

The mutagenesis reaction was cycled as follows: 95 °C for 1 minute, followed by 95 °C for 50 seconds, 60 °C for 50 seconds and 68 °C for 1 minutes/kb of plasmid length, for a total of 18 cycles. The final step was performed at 68 °C for 7 minutes. Following the temperature cycling, reactions were cooled on ice

for 2 minutes. DpnI restriction enzyme was then used to digest each reaction for 1 hour at 37 °C in order to remove remaining undigested parental DNA template, leaving only the mutated plasmid which was then transformed into XL10-Gold Ultra competent cells, according to manufacturer's instructions.

2.3.15.2 Gateway cloning

Gateway recombination was used as the preferred method of cloning in this project, using the Invitrogen Life Technologies kits and following manufacturer's instructions. Gateway entry clones (pENTR) were either obtained from plasmid repositories, such as DNASU (https://dnasu.org/) or Addgene (https://www.addgene.org/) were created by PCR with att-tagged primers, following manufacturer's instructions.

Plasmids from the repositories were sent as glycerol stocks in DH5-alpha bacteria. The cells were streak purified and grown overnight on LB agar plates supplemented with the required antibiotic resistance. Isolated colonies were picked and propagated in 5 ml of LB broth supplemented with the required antibiotic overnight at 37 °C with shaking at 225-250 rpm. pENTR plasmid DNA was purified from 1-5 ml of culture using the QIAprep Miniprep Kit (Qiagen), according to the manufacturer's instructions (Section 2.3.16).

For entry clones that were created by PCR using att-tagged primers, following PCR amplification the PCR products were purified according to manufacturer's recommendations (Invitrogen). Briefly, TE buffer and 30%

PEG 8000/30 mM $MgCl_2$ solution were added to 50 μl attB-PCR product and centrifuged at RT for 15 minutes at 10,000 x *g*. Supernatant was removed and the pellet re-dissolved in TE buffer before quantification using the Nanodrop (2.3.2) and visualization on an agarose gel (2.3.4). The BP reaction facilitated propagation of purified attB-PCR product or plasmid DNA in the donor vector (pDONR201) using BP clonase II (Invitrogen), generating the pENTR clone, following manufacturer's instructions. The pENTR clones were verified through Sanger sequencing (Section 2.3.9).

Destination constructs (pDEST), were generated through the use of LR clonase II in an LR reaction, following the manufacturer's protocol (Invitrogen). The expression constructs were produced by combining 150 ng pENTR clone with 150 ng pDEST destination vector. The destination vectors used in this project were pDEST40 and pCI-NEO-*RHO* (a kind gift from Erwin van Wijk, Radboud University Medical Centre). All constructs were verified by Sanger sequencing (2.3.9).

2.3.16 Plasmid DNA isolation and purification

The QIAprep Miniprep kit (Qiagen) was used for small-scale plasmid DNA isolation and purification, according to the manufacturer's protocol. 4 ml of cells from the 5 ml bacterial cell overnight culture were pelleted by centrifugation at 3000 x *g* and subsequently redissolved in resuspension buffer (P1) containing RNAse A (Qiagen). Resuspended cells were then combined with an equal volume of alkaline lysis buffer (P2) and neutralisation

solution (N3). Cell debris was removed after centrifugation at 3000 x *g* and cell lysates were adsorbed on to a silica membrane after centrifugation of the spin column at 100 x *g*. The flow-through containing RNA, metabolites and protein was discarded. The DNA held in the membrane was washed with ethanol buffer and eluted from the spin column in 50 μl dH_2O. The DNA from the mini-prepped plasmids was quantified (2.3.2) and verified by Sanger sequencing (Section 2.3.9).

Following sequence verification, the EndoFree Plasmid Maxi Kit (Qiagen) was used, according to the manufacturer's protocol. Cells from the 100 ml overnight culture were pelleted by centrifugation at 6000 x *g*, 4 °C for 30 minutes and then re-dissolved in 10 ml Buffer P1 (Qiagen). Cells were subsequently lysed and all cell debris was removed by applying the lysed cells to a QIAfilter. The filtered lysate was then incubated on ice for 30 minutes in endotoxin removal buffer before being passed through a QIAGEN-tip 500 by gravity flow. The tip membrane was washed twice with wash buffer and DNA was eluted with 15 ml elution buffer. Plasmid DNA was precipitated by adding 0.7 volumes RT isopropanol and centrifuged at 15,000 x *g* for 30 minutes at 4 °C. The supernatant was carefully discarded and the resulting DNA pellet was washed with 5 ml 70% ethanol and then centrifuged at 15,000 x *g* for 10 minutes. Supernatant was then discarded and pellets left to air dry. DNA pellets were then dissolved in 150-300 μl filter-sterilized TE. Maxi-prepped plasmids were quantified (Section 2.3.2) and verified by Sanger sequencing (Section 2.3.9).

2.3.17 Preparation of plasmid glycerol stocks

Glycerol stocks were generated for all plasmid cultures in the project by mixing 500 μl of the overnight bacterial cell culture with 500 μl of 50% sterile glycerol in a 1.5 ml cryovial. All glycerol stocks were placed in the -80 °C for long-term storage.

2.3.18 Cell culture

Human embryonic kidney cells, highly transfectable (HEK293T) and African green monkey kidney cells (COS7) were sourced from ATCC. All cell culture reagents were purchased from Sigma Aldrich unless stated otherwise. Cell lines were proliferated in 75 cm^2 (T75) flasks at 37 °C with 5% CO_2 in Sanyo MCO 20AIC cell culture incubators. All cell culture work was performed in NuAire Labgard 437 ES Class II Biosafety Cabinets under sterile conditions. Cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) which was supplemented with 10% fetal calf serum (FCS) and 100 U/ml penicillin with 100 mg/ml streptomycin (pen/strep). This was the medium used in this project.

When reaching 80-90% confluency, cells were passaged by removal of culture medium, washing of cells with Dulbecco's phosphate buffered saline (DPBS) and cell dissociation from the culture flask using trypsin/EDTA. After addition of 1 X trypsin/EDTA, cells were incubated for approximately 5 minutes at 37 °C to allow for cell dissociation. Trypsin was neutralised through addition of fresh culture medium and the cells were collected in a 15 ml

Falcon tube. Fresh culture medium was pre-warmed in T75 flasks before addition of the collected cells at an appropriate subculture ratio.

2.3.18.1 Cell counting

Cells were counted using the Countess Automated Cell Counter (Life Technologies) following the manufacturer's protocol. Briefly, 10 µl of resuspended cells were mixed with 0.4% trypan blue dye 1:1 and inserted into the slide for counting. Viable cells were then calculated and uniformly seeded into the wells of plates.

2.3.18.2 Long term cell storage and recovery

Culture medium (supplemented with FCS and pen/strep) with 10% DMSO was used for the long-term storage of cell lines in the liquid nitrogen. 1 x 10^6 cells were transferred into 1.5 ml cryovials and gradually frozen at -80°C in a Mr Frosty container (Nalgene), before being moved to liquid nitrogen 24 hours later. Cryovials of cells were recovered by thawing in a 37 °C water bath, gradual addition of 10 ml culture media and pelleting at 200 x *g* for 5 minutes at RT, before resuspension in total cell culture medium and transfer to a T75 flask. Cells were checked under the microscope for viability 24 hours later and culture medium replaced before cell sub-culturing into a fresh T75 flask after another 24 hours.

2.3.19 Transient DNA transfections

Prior to transfection, cells were seeded at optimal densities in the wells of either a 12- or 24-well plate. The volumes and concentrations used in each specific experiment are detailed in 2.3.20 and 2.3.21. The general transfection protocol was performed using 3-3.6 μl Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions, in a suspension with 125-187.5 μl Gibco OPTI-MEM medium (Life Technologies). Separately, the DNA to be transfected (400-800 ng) was suspended in Gibco OPTI-MEM. Incubation of both mixes were undertaken at RT for 5 minutes. Following the incubation, the DNA mixture was added to the lipofectamine mix and incubated at RT for 20 minutes. 0.5-1 ml of fresh medium (2.3.18) was added to the wells of the plate (24- or 12-well plate respectively) and the transfection mixture was added drop-wise to each well, ensuring even dispersal. Reactions were incubated overnight at 25°C. Proteinase K solution was then added to terminate each reaction, with an incubation for 10 minutes at 37 °C. For both the midigene assay (2.3.20) and the luciferase reporter assay (2.3.21), cells were left to incubate for 24 hours prior to transfection. Fresh medium (2.3.18) was added prior to transfection and cells were harvested 48-hour post-transfection.

2.3.20 Midigene assay

A midigene *in vitro* splice assay system was used to investigate an intronic *SLC38A8* variant (Chapter 5). To create the midigene, a bacterial artificial chromosome (BAC) was purchased from Source Bioscience, Cambridge, UK. The BAC (RPCIB757F20757Q, RP11-757F20, Accession number AQ466782, AQ511190), contained the entire *SLC38A8* gene and arrived in LB agar containing 20 μg/ml chloramphenicol. The clone was streaked onto the same

agar in order to obtain single colonies and incubated at 37 °C overnight. Single colonies were inoculated in 2 ml LB broth supplemented with 20 μg/μl chloramphenicol and grown overnight at 37°C with shaking at 225-250 rpm. Cultures were centrifuged at 800 x *g* for 10 minutes and the supernatant discarded. The pellets were re-dissolved in 0.3 ml filter-sterilized P1 solution (15 mM Tris pH 8.0, 10 mM EDTA pH 8.0 and 100 μg/ml RNAse A), followed by the addition of 0.2 ml filter-sterilized P2 solution (0.2 M NaOH, 1% SDS). The suspensions were mixed gently and left at RT for 5 minutes until the suspension became translucent. 0.3 ml solution P3 (3 M KOAc pH 5.5) was then added with shaking before placing the tubes on ice for 5 minutes. Tubes were then centrifuged at 8000 x *g* for 10 minutes at 4 °C followed by transfer of the supernatant into 0.8 ml ice-cold isopropanol and a brief spin in the centrifuge. Supernatant was discarded and 0.5 ml fresh 70% ethanol added with inversion to wash the DNA pellets. Again, supernatant was discarded and pellets were left to air dry before re-dissolving in 40 μl TE buffer.

Primers to introduce attB1 and attB2 sites and amplify a specific region of *SLC38A8* spanning exons 5 and 6 were manually designed (Appendix 15). The region of interest was amplified using the primers in the following PCR reaction using KAPA HotStart DNA polymerase. 1 U/μl KAPA DNA polymerase, 10 mM of each dNTP, 5 X KAPA HiFi buffer, 1 M Betaine, 5% DMSO, dH_2O and 10 μM forward and reverse primer. The reaction conditions were as follows: 95 °C for 3 minutes followed by 35 cycles of 98 °C for 20 seconds, 68 °C for 15 seconds, 72 °C for 8 minutes and a final step of 72 °C for 8 minutes.

SLC38A8 was amplified and purified as detailed in 2.3.15.2. The BP reaction generated the pENTR clone through the attB sites introduced into PCR-amplified products. Gateway cloning of the pENTR clone into the pDEST pCI-NEO-RHO (a kind gift from Erwin van Wijk, Radboud University Medical Centre), using LR clonase II was undertaken. SDM was performed on the destination vector to introduce the intronic variant of interest. To assess the effect of the intronic variant of interest, HEK293T and COS7 cells were transfected with wild-type (WT) and variant constructs. Cells were seeded at 1.5 x 10^5 cells/well in 12-well plates and left to incubate for 24 hours. Cells were transfected as described in 2.3.19 with a total DNA concentration of 800 ng/well. Prior to transfection, 1 ml fresh total medium was added to each well. Cells were harvested and quantified 48-hours post-transfection.

2.3.20.1 RNA extraction and cDNA synthesis from cell lines

Following transfection of WT and variant constructs, total RNA was isolated from each cell line using the RNeasy Mini kit (Qiagen), following the manufacturer's recommendations. Cells growing in a monoloayer were harvested using a cell scraper on ice. The cells were counted as described in Section 2.3.18.1. Cos7 cells were used at no more than 3 x 10^6 and HEK293T cells at no higher than 1 x 10^7. Twice the volume of complete media was used to collect the dissociated cells which were then centrifuged for 10 minutes at 300 x *g*. The resulting cell pellet was lysed and homogenised in guanidine-thiocyanate-containing buffer (Buffer RLT) which is highly denaturing and purifies RNA by inactivating RNases. The lysate was applied to a QIAshredder spin column and centrifuged for 2 minutes. 1 volume of 70%

ethanol was added to the homogenised lysate and mixed well by pipetting. The sample was passed through a RNeasy spin column for 15 seconds at 8000 x *g* and washed three times with wash buffer before elution of RNA in 50 μl RNase-free water. cDNA synthesis was performed as described in 2.3.12. *SLC38A8* transcript analysis was carried out by performing reverse-transcription (RT) PCR using *SLC38A8* exonic primers and RHO-specific primers, as outlined in 2.3.12. RT-PCR products were size-fractionated using agarose gel electrophoresis, as described in 2.3.4.

2.3.21 Dual luciferase reporter assay

Cells were seeded at 1 x 10^4 cells/well in a 24-well plate and grown to approximately 80% confluency. Cells were transfected 24-hours later (2.3.18) using 3.6 μl Lipofectamine 2000 in 187.5 μl OPTI-MEM (Invitrogen) with a total of 800 ng/well DNA. The 800 ng included approximately 400 ng of pGL3 construct DNA and approximately 400 ng of *FOXD1*-WT DNA. 1 ng of the transfection control *Renilla* luciferase plasmid (pRL-TK) (Promega) was also added to each transfection.

After the transfection had proceeded for 48-hours, the cell medium was removed and cells were harvested. Cells were then lysed at RT for 15 minutes using 1 X passive lysis buffer from the Promega Dual-Luciferase Reporter Assay System kit (Promega).

The dual luciferase assay measured both firefly luciferase and *Renilla* luciferase. The dual luciferase assay was undertaken in an opaque 96-well plate (Grenier BioOne) and plates were assayed using the Mithras LB 940 luminometer (Berthold Technologies). The luminometer was pre-programmed with the timings shown in Table 2-1. The assay was performed using luciferase assay reagent (LAR II) which measured the firefly luciferase levels. Stop and Glo reagent was prepared before the assay at RT and was used to assay the *Renilla* levels derived from the pRL-TK vector transfection control. For each well, the level of firefly luciferase was normalised to the *Renilla* in a ratio, generating relative luciferase units (RFU). Three technical replicates were performed and for each a total of three luminescence values were recorded. However, only one biological replicate was performed due to the direction of the project at this time point (discussed in Results Chapter 2).

Step	Stage	Time (seconds)
1	LAR II reagent released into wells	3
2	Firefly luciferase activity measured	12
3	Stop and Glo reagent injected into wells	3
4	*Renilla* activity measured	12

Table 2-1 Stages of the dual luciferase assay.

2.3.22 Whole exome sequencing (WES)

2.3.22.1 WES capture and library preparation

WES library capture was undertaken by the next-generation sequencing facility (St James's University Hospital, Leeds) (http://dna.leeds.ac.uk/genomics/) using the SureSelectXT target enrichment kit V5 and V6 (Agilent Technologies, Santa Clara, CA, USA), following the manufacturer's recommended 200 ng protocol. The quality of the samples prepared by the Leeds sequencing facility were assessed using the 2200 TapeStation Instrument (Agilent Technologies, Santa Clara, CA, USA). The reagents used were included in the SureSelectXT Library Prep Kit, ILM, SureSelect Target Enrichment Kit ILM Indexing Hyb Module Box #2, Herculase II Fusion DN Polymerase and SureSelect Target Enrichment Kit Box #1 (Agilent).

Genomic DNA (gDNA) samples considered for WES were quantified using the Qubit dsDNA BR assay kit (Invitrogen) (Section 2.3.2). The 200 ng gDNA samples were diluted with 1 X low TE buffer to a total volume of 50 μl and sheared to form fragments with a peak size of 150-200 bp using a Covaris S229 system. The quality of the samples was then assessed using a high sensitivity DNA chip and reagent kit (Agilent) on a 2100 Bioanalyzer instrument (Agilent). The electropherograms generated by the Agilent 2100 Expert software should show a distribution with a DNA fragment size peak of 150-200 bp.

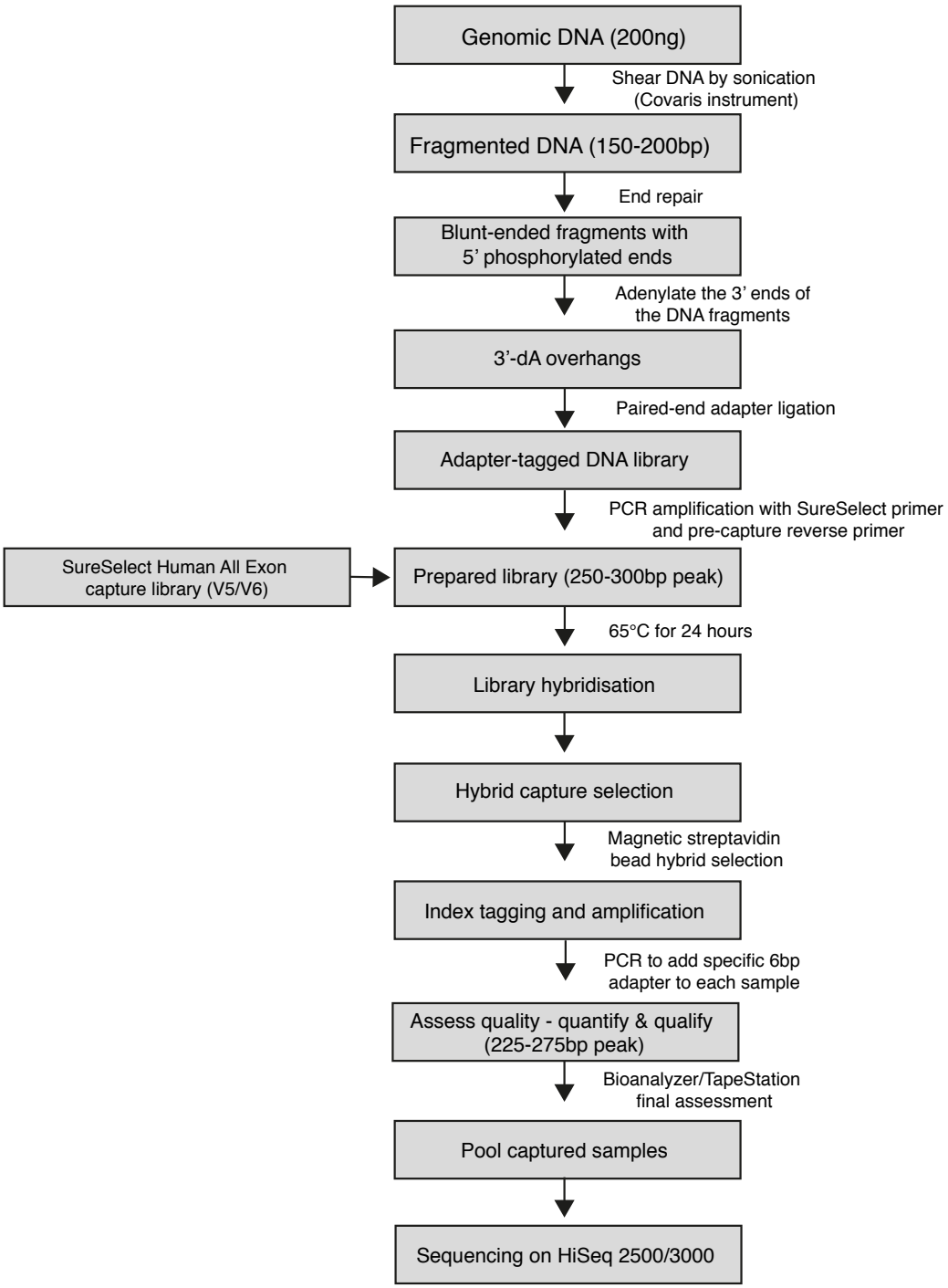


Figure 2-1 Agilent Technologies SureSelectXT library preparation workflow for Illumina paired-end sequencing.

In order to repair the ends of the DNA fragments, the SureSelect XT Library Prep Kit ILM was used to prepare an end repair master mix containing 35.2 μl of nuclease-free water, 10 μl 10 X end repair buffer, 1.6 μl dNTP mix, 1 μl T4 DNA polymerase, 2 μl Klenow DNA polymerase and 2.2 μl of T4 polynucleotide kinase. 48 μl of each sample was added to each well with 52 μl of the mastermix and the mix was incubated in a thermal cycler at 20 °C for 30 minutes. Samples were then purified using homogeneous Agencourt AMPure XP beads (Beckman Coulter Genomics). 180 μl of the AMPure XP beads were added to each sample well and pipetted up and down 10 times, followed by incubation for 5 minutes at RT. The samples were then placed in a magnetic stand in order to separate the PCR product which binds to the beads from contaminants including primers and nucleotides. After removing and discarding the cleared solution, samples were washed twice with 200 μl of freshly-prepared 70% ethanol. 32 μl nuclease-free water was added to each sample well in order to elute the purified PCR product from the beads.

The 3’ end of all DNA fragments were then adenylated by mixing 11 μl nuclease-free water, 5 μl 10 X klenow polymerase buffer, 1 μl dATP and 3 μl Exo(-) Klenow. The mix was incubated at 37 °C for 30 minutes and was followed by purification with AMPure XP beads as described above (using 180 μl beads and 200 μl 70% ethanol for each wash).

To ligate the paired-end adapters, 13 μl of each sample was added to a mix which consisted of 15.5 μl nuclease-free water, 10 μl 5 X T4 DNA Ligase

buffer, 10 μl SureSelect Adapter (AGATCGGAAGAGCAC) Oligo Mix and 1.5 μl T4 DNA ligase. Samples were incubated at 20 °C for 15 minutes. Again, samples were purified as described above using AMPure XP beads using 90 μl beads, 200 μl 70% ethanol for each wash and 32 μl nuclease-free water for the final elution.

The adapter-tagged libraries were then PCR-amplified by mixing 30 μl of each sample with 6 μl nuclease-free water, 1.25 μl SureSelect Primer, 1.25 μl SureSelect ILM Indexing Pre-Capture PCR reverse primer, 10μl 5X Herculase II Reaction buffer, 0.5 μl 100 mM dNTP Mix and 1 μl Herculase II Fusion DNA Polymerase. The PCR program was run as shown in Table 2-2.

Segment	No. of Cycles	Temperature	Time
1	1	98 °C	2 minutes
2	10	98 °C	30 seconds
		65 °C	30 seconds
		72 °C	1 minute
3	1	72 °C	10 minutes
4	1	4 °C	Hold

Table 2-2 Pre-Capture PCR Thermal Cycler Program.

Finally, the amplified, adapter-tagged libraries were purified using the AMPure XP beads as described above (using 90 μl beads, 200 μl 70% ethanol for each wash and 30 μl nuclease-free water for the final elution) and assessed on a 2100 Bioanalyzer using the DNA 1000 assay with an expected DNA fragment size peak of 225-275 bp.

The prepared libraries were hybridized to the capture library of biotinylated RNA oligo probes by mixing 3.4 μl of 221 mg/μl library, 5.6 μl of SureSelect Block Mix (prepared by mixing SureSelect Indexing Block #1-3), 5 μl of SureSelectXT Human All Exon V5/V6 (Agilent), 2μl of RNase Block Dilution (1:3 RNase Block: water) and 40 μl of hybridisation buffer (prepared by mixing SureSelect Hyb #1-4). The mix was incubated at 95 °C for 5 minutes and then at 65°C for 24 hours. Selection of the captured library was then performed using streptavidin-coated magnetic beads (Dynabeads MyOne Streptavidin T1) (Invitrogen) that had been prepared and washed. The beads were separated from the buffer using a magnetic separator. The supernatant was removed and the beads were re-dissolved in 200 μl of SureSelect Wash 1. The beads were then washed 3 times using SureSelect Wash 2, recovered with the magnetic separator and finally the library was dissolved in 30 μl of water.

The DNA libraries were then PCR-amplified with indexing primers. The post-capture PCR reaction mix included 18.5 μl nuclease-free water, 10 μl 5 X Herculase II Reaction buffer, 1 μl Herculase II Fusion DNA Polymerase, 0.5 μl

100 mM dNTP Mix and 1 μl SureSelect ILM Indexing Post-Capture Forward PCR Primer. 5 μl of the appropriate indexing primer was then added along with 14 μl of each library and the PCR program was run as shown in Table 2-3.

Segment	No. of Cycles	Temperature	Time
1	1	98 °C	2 minutes
2	12	98 °C	30 seconds
		57 °C	30 seconds
		72 °C	1 minute
3	1	72 °C	10 minutes
4	1	4 °C	Hold

Table 2-3 Post-capture PCR Thermal Cycler Program.

The libraries were then purified using AMPure XP beads as describe above (using 90 μl beads, 200 μl 70% ethanol for each wash and 30 μl nuclease-free water for the final elution) and assessed on a 2100 Bioanalyzer using the 2100 Bioanalyzer High Sensitivity DNA Assay where the electropherogram is expected to show an average fragment size of approximately 300 bp.

In the final step, 10 samples per lane were pooled aiming at a final concentration of 10 nM in a final volume of 50 μl by using the following

formula: Volume of library = $(V_f \times C_f)/(\# \times C_i)$, where V_f is the final desired volume of the pool (50 μl), C_f is the final desired DNA concentration (10 nM), # is the number of samples (10) and C_i is the initial concentration of each sample. The pooled libraries were finally subjected to either 100 or 150 bp paired-end sequencing on a single lane of a HiSeq2500 or HiSeq3000 (Illumina) by the Leeds Next Generation Sequencing Facility.

2.3.22.2 WES Bioinformatics

The java-based visual interface FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to interrogate the quality of all raw fastq files before proceeding to further analyses. Sequence quality, GC content, duplicate sequences and adapter contamination were checked by this tool. All patient exomes were assessed for quality metrics at each of the three main stages of NGS analysis: raw data, alignment and variant calling. Quality control (QC) was performed by examining the quality of the FastQ files using FastQC software (as described above), the quality of the BAM file generated from the alignment using Picard and SAMtools, and also the statistics of the VCF file. Further QC metrics were performed on the final BAM files using Picard's CollectMultipleMetrics and samtools flagstats. A final QC check was performed on the VCF file using bcftools and plot-vcfstats. All QC commands are detailed in Appendix 2.

The commands used for the bioinformatics analysis of WES data are listed in Appendix 2. FastQ files were firstly modified in a quality control step using

Trim Galore! (v0.4.0 with Cutadapt v1.16), a package developed by the Babraham Institute (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Default settings were used, although the --illumina command was removed to allow the tool to auto-detect the adapter sequence. As shown in Figure 2-2, processed FastQ files were aligned to the human reference genome GRCh37 using Burrows-Wheeler aligner (BWA) (v0.7.12-r1.39) software (Li and Durbin, 2009). The aligned reads were then processed in SAM format using SAMtools (v1.3) and piped in to BAM.

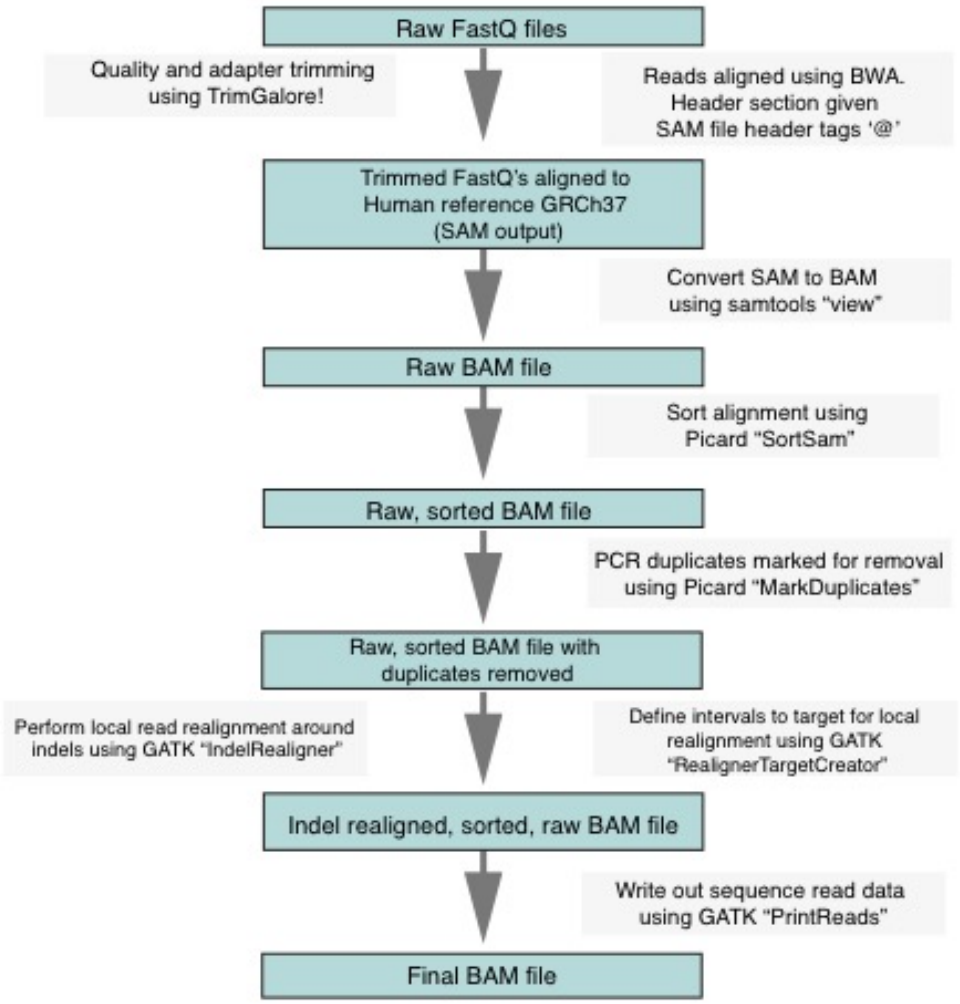


Figure 2-2 WES data alignment pipeline.

Processes applied to WES individuals in the alignment of raw data (FastQ files) to the human reference genome build GRCh37/hg19 using BWA. PCR duplicates were marked and removed using Picard. The final BAM file was prepared using GATK after local indel realignment and finalisation of the read data. In later stages, the final BAM file was subjected to filtering using a strategic pipeline to identify variants in candidate genes of interest.

Picard (v1.141) (http://broadinstitute.github.io/picard/) was used to sort the alignment around indel sites and mark PCR duplicates. Realignment of indels and printing of the final BAM file was performed by GATK (McKenna *et al.*, 2010; DePristo *et al.*, 2011).

Insertions/deletions (indels), single-nucleotide variants (SNVs) and multi-nucleotide variants (MNVs) were called in the genomic variant call format (g.VCF) using GATK's HaplotypeCaller program. GVCF files were then genotyped with two unrelated and unaffected control exomes that had been processed using the same exome library preparation kit, the same alignment and the same variant calling pipeline using GATK's GenotypeGVCFs (GATK v3.5-0). All exomes were filtered against these control individuals in the hope that any sequencing artefacts introduced during library preparation, or common SNPs could be cancelled out of the analysis. SNVs, MNVs and indels were then separately recalibrated and hard filtered, before combining to produce a final VCF file (Danecek *et al.*, 2011).

Examples of WES variant filtering can be found in Appendix 2. The vcfhacks package (https://github.com/gantzgraf/vcfhacks) was used to carry out filtering

and annotation of VCF files. All VCF files generated from alignment and variant calling were annotated as standard using the following databases of known variation:

- The National Centre for Biotechnology Information's (NCBI's) database of single nucleotide polymorphisms and multiple small-scale variations (dbSNP) version 146 (https://www.ncbi.nlm.nih.gov/projects/SNP/) (Sherry *et al.*, 2001).
- The Broad Institute's Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org) (Lek *et al.*, 2016).

The mode of inheritance influenced the filtering strategy applied. Filtering for recessive disease was prioritised due to the predicted mode of inheritance, therefore only biallelic variants were retained. Due to the predicted recessive inheritance pattern and rarity of FHONDA, a cut off of 0.01% (MAF 0.0001) was used to filter the dataset. Patients in the cohort were also filtered for rare dominant variants, therefore in this analysis only heterozygous variants were retained. A cut off of 0.001% (MAF 0.00001) was used in this analysis.

As the majority of the WES cases were sporadic patients, filtering for variants shared within families was not possible. However, in the case of the sib pair (F1372 and F1373), patients were genotyped together using the GATK GenotypeGVCF command and filtered to retain shared variants. In the consanguineous cases, datasets were filtered to retain homozygous or compound heterozygous variants. Filtering to retain shared genes in the entire patient cohort and filtering to investigate different variants in the same gene was also performed. As mentioned above, the combined cohort was filtered to

remove any variants present in two unrelated control exomes. Figure 2-3 details possible strategies for variant filtering.

Synonymous, non-coding, intronic and intergenic variants, other than those potentially affecting splice donor or acceptor sites (within 2 bp of a splicing junction) were removed from the datasets, in order to retain only 'functional' variants. Although coding sequence variants were prioritised, variants referred to as 'functional' in Chapter 4 were also those affecting known transcription factor binding sites or regulatory regions, and so were retained in the analysis. The pathogenicity prediction tool CADD (Section 2.4.4.7) was used to rank variants, and those with a CADD score >15 were retained. Variants with the higher CADD scores were prioritised in the final lists. PolyPhen-2 and SIFT (Section 2.4.4.2 and 2.4.4.5) were also used to assign pathogenicity prediction scores to the variants with a CADD >15. Variants predicted to be pathogenic by at least one of the tools, PolyPhen-2 or SIFT, and with a CADD >15, were retained in the analysis and prioritised in the final lists. Such 'prioritised' variants were then investigated further by manually checking the dbSNP, ExAC and gnomAD databases, to confirm variants were not common. Further investigation of prioritised variants also included literature searches for gene function and expression in relevant tissue types, for example (Figure 2 3).

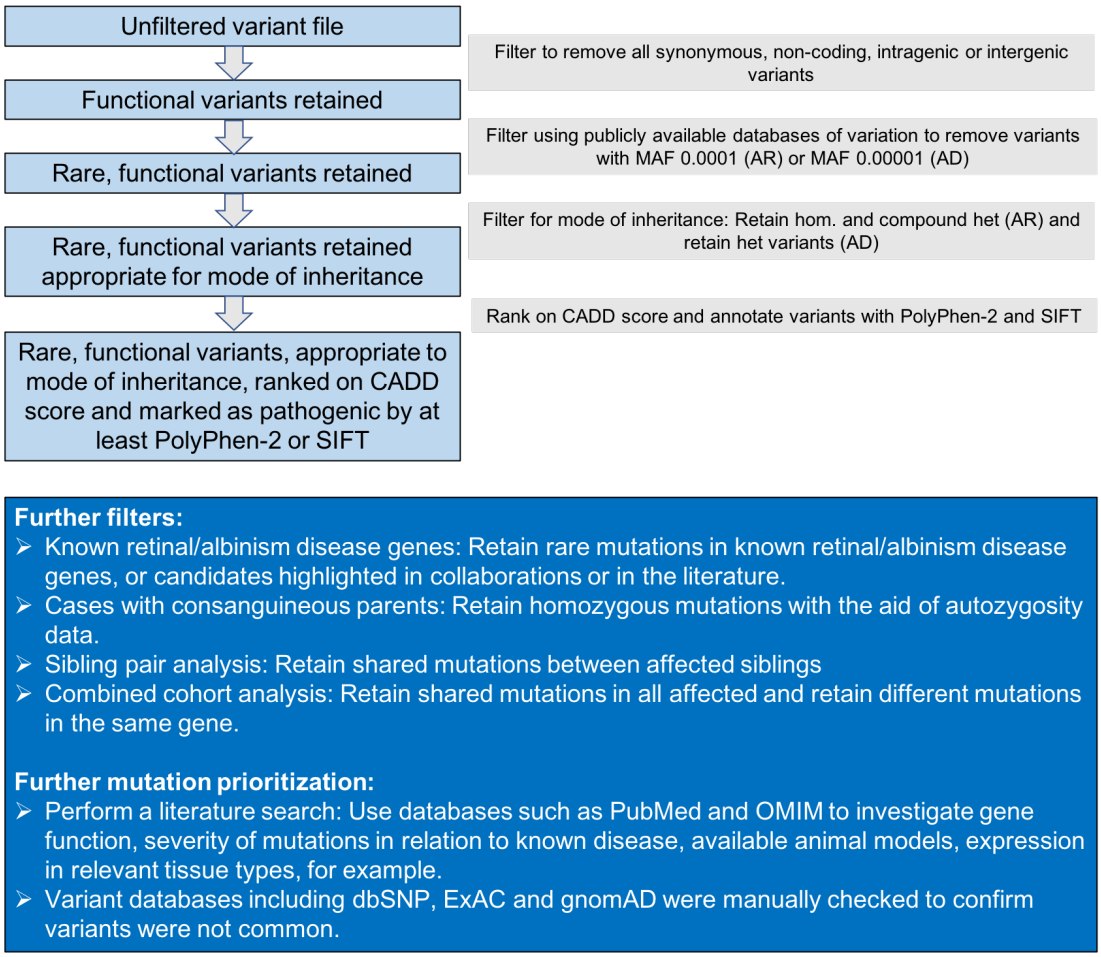


Figure 2-3 WES variant filtration and prioritisation strategies.

The pipeline details the basic strategy applied in all variant filtration. Additional filters are detailed in the lower panel alongside additional variant prioritisation strategies. Prioritisation begins with the predicted pathogenicity of the rare variants.

2.3.22.3 Whole exome sequencing annotation of variants

Examples of simplified variant annotation commands can be found in Appendix 2. Preceeding variant database filtering, VCF files were annotated using Ensembl's Variant Effect Predictor version 89 (VEP) (http://www.ensembl.org/info/docs/tools/vep/index.html) (McLaren *et al.*, 2010). Annotated VCF files were assigned Combined Annotation Dependent Depletion version 1.3 (CADD) scores and ranked on their C-score (http://cadd.gs.washington.edu) (Kircher *et al.*, 2014).

Final variant lists were annotated using vcfhacks GeneAnnotator.pl which assigned Gene Reference into Function (RIFS) (https://www.ncbi.nlm.nih.gov/gene/about-generif), Gene Ontology (http://www.geneontology.org), NCBI summaries, BioGrid (https://thebiogrid.org), Online Mendelian Inheritance in Man (OMIM) (https://www.omim.org) and Mouse Genome Informatics (MGI) phenotypes (http://www.informatics.jax.org/phenotypes.shtml) to each variant's INFO field using the variant Entrez gene ID allocated by VEP. vcfhacks' annovcfToSimple.pl outputs the final variant lists in Excel's .xlsx format. These "raw" Excel outputs were then interrogated by in-Excel filtering and variant prioritisation, as described in the previous section.

2.3.22.4 Whole exome sequencing depth of coverage

An example of the command to attain depth of coverage statistics for WES samples can be found in Appendix 2. The final BAM files (aligned to reference genome, sorted, duplicates marked and base recalibrated), were used to calculate the depth

of coverage using GATK's Depth of Coverage tool. Using the appropriate SureSelect Human All Exon captured regions bed file (V5/V6) (Agilent) as the input file, the coverage was calculated for the BAM files. The mean coverage per exome and also the percentage of bases covered at minimum depths of 5, 10, 15, 20, 25 and 30 x were calculated.

2.3.23 Analysis of copy number variation (CNVs)

In order to identify CNVs too large to be detected by the variant calling protocol described above (Section 2.3.25.3), ExomeDepth (Plagnol *et al.*, 2012), BreakDancerMax (Chen *et al.*, 2009) and DELLY2 (Rausch *et al.*, 2012) were employed in the analysis of WES data, with details as follows.

2.3.23.1 ExomeDepth

ExomeDepth functions in R script and outputs duplications and deletions by comparing the read depth of the test sample to unrelated reference samples. It is compulsory that the reference samples were prepared using the same capture library and processed on the same sequencing run. The software recommends a minimum of 5-10 reference samples. All samples, test and reference, should be aligned in BAM format and indexed. The .csv output is annotated with common CNVs (Conrad *et al.*, 2010) and those marked as uncommon are prioritised alongside CNVs with a high Bayes Factor (BF). The BF is a statistical confidence level, or likelihood ratio of the CNV call against the normal copy number call. ExomeDepth outputs both a table of CNVs and also a graphical plot of the observed: expected read ratio at a specific chromosome location. The table documents the

read ratio of CNVs as follows; heterozygous deletion (0.5), homozygous deletion (0), heterozygous duplication (1.5), homozygous duplication (2). The graph also marks a 95% confidence interval of the read ratio. The code used for this analysis can be found in Appendix 4.

2.3.23.2 BreakDancer

BreakDancer/BreakDancerMax is a C++ program that identifies discordant read pairs (reads of abnormal size or orientation) in an aligned exome/genome, indicative of structural variation. A paired-end read, if concordantly mapped, will show a specific orientation arrangement, with the first reads aligning in the positive orientation and the second reads aligning in the negative. Concordantly aligned read pairs should conform to a uni-modal distribution, since the fragment size usually does. BreakDancer functions to detect reads with discordant orientations or insert sizes and clusters them to label SV breakpoints. The identified structural variation can be in the form of; deletions, insertions, inversions, intra-chromosomal translocations, inter-chromosomal translocations and unknown. The code for this analysis can be found in Appendix 4.

The input requires an aligned BAM file that must first undergo replacement of the read group headers using Picard's "AddOrReplaceReadGroups" command. The tool requires specific read group fields to be defined in the BAM file. The read group fields were renamed as the read group (RG) ID, platform unit (PU) (also flowcell identifiers), the name of the sample sequenced in the read group (SM), the platform used to produce the reads (PL) and the DNA preparation library identifier (LB). The

perl script “bam2cfg.pl” was then run to generate a configuration file of the statistics for each read group, alongside a series of QC commands. The BreakDancerMax program uses the read statistics to identify clusters of discordant read pairs, resulting in the list of structural variation. The tool does not yet integrate read depth.

As mentioned above, BreakDancerMax is able to analyse structural variation (SV) in the form of deletions (DEL), duplications (DUP), insertions (INS), inversions (INV), intra-chromosomal translocations (ITX), inter-chromosomal translocations (CTX), or unknown (UN) from paired-end read alignments, with the data gathered outputted as a .txt file. The output returns predicted breakpoints or coordinates of the two break-ends. The orientation is described as a string that records the number of reads mapped to the plus or the minus strand in the anchoring regions (Chen *et al.*, 2009).

2.3.23.3 DELLY2

DELLY2 is similar to BreakDancer and identifies discordant reads from paired-end sequencing data. However, DELLY2 combines both short- and long-ranged paired-end mapping and splitting-read analysis to ascertain forms of structural variation. Splitting reads can be described as sequencing reads that span a structural variant breakpoint. The discordant read pair analysis identifies breakpoints of genomic intervals and the splitting reads analysis refines the breakpoint at a single nucleotide resolution using a k-mer-based algorithm of alignment. DELLY2 can pinpoint deletions, tandem duplications, inversions and translocations, but not insertions. The code used for this analysis can be found in Appendix 5. DELLY2 is reported to detect balanced and unbalanced forms of SV of variable sizes, including deletions (DEL),

inversions (INV), tandem duplications (DUP) and translocations (TRA). DELLY2 requires a sorted and indexed BAM file and the matching alignment reference genome (GRCh37/hg19) to identify split-reads. The BCF format (plus index) is the output which is subsequently converted into a VCF format using bcftools.

2.3.24 Whole Genome Sequencing (WGS)

2.3.24.1 WGS capture, library preparation and sequencing by Edinburgh Genomics

WGS was performed by Edinburgh Genomics (https://genomics.ed.ac.uk) which utilises Illumina SeqLab. Sequencing libraries were prepared using the Illumina SeqLab-specific TruSeq Nano High Throughput kit. A minimum of 1.5 µg of gDNA, in a concentration of 50 to 100 ng/µl and with a minimum volume of 30 µl were sent to Edinburgh Genomics.

Samples were evaluated for quality and quantity using an AATI Fragment Analyzer and the DNF-487 Standard Sensitivity Genomic DNA Analysis Kit. The AATI ProSize 2.0 software provided a quantification value and a quality (integrity) score for each gDNA sample. Samples were reported to fail the QC check if the genomic DNA was found to have a quality score <7 and have little or no high molecular weight material. Based on the quantification results, gDNA samples were pre-normalised to fall within the acceptable range of the Illumina SeqLab TruSeq Nano library preparation method using the Hamilton MicroLab STAR. Library QC was also performed to evaluate the libraries for mean peak size and quantity using the Caliper GX Touch with a HT DNA 1k/12K/HI SENS LabChip and HT DNA HI SENS Reagent Kit. The libraries were normalised to 5 nM using the GX data and the actual concentration

was established using a Roche LightCycler 480 and a Kapa Illumina Library Quantification kit and Standards.

Each library was sheared to a mean insert size of 450 bp using a Covaris LE220 focused ultrasonicator. Library preparation also included end repair, size selection, 3'-end adenylation, adapter ligation and enrichment with 8 cycles of PCR amplification. The Caliper GX Touch with a HT DNA 1k/12K/HI SENS LabChip and HT DNA HI SENS Reagent kit was used to quantify and control the peak size of the libraries. Libraries were normalised to 5 nM, with the actual concentration defined using a Roche LightCycler 480 and a Kapa Illumina Library Quantification kit. The libraries were then normalised, denatured and pooled in eights for clustering using a Hamilton MicroLab STAR with Genologics Clarity LIMS X edition. The libraries were clustered onto a HiSeqX flow cell v2.5 on cBot2s and the clustered flow cell transferred for sequencing on the HiSeqX, using the HiSeqX Ten Reagent kit v2.5.

Edinburgh Genomics aligned the reads to the GRCh38 human reference genomes using BWA-MEM (0.7.13). Patients aligned by Edinburgh Genomics are shown in Table 2-4. Duplicated fragments were marked using SAMBLASTER (0.1.22) and GATK (3.4-0-g7e26428) was used for the indel realignment and base recalibration. GATK was also used to calculate genotype likelihoods and the final GVCF file was generated using HaplotypeCaller.

2.3.24.2 Alignment of WGS samples

Upon initial receipt of the WGS data, patients F1377 and F1310 had been aligned to build GRCh38 by Edinburgh Genomics (Table 2-4). In order to use the vcfhacks tools (Section 2.3.22.2) for the ensuing analysis, it was necessary that the alignment was to the previous human reference genome, build GRCh37. Therefore, the two genomes were realigned using a similar approach to the pipeline for WES (Section 2.3.22.2), including alignment using BWA, sorting and removal of PCR duplicates using Picard, and variant calling using GATK's HaplotypeCaller. Genome alignments were undertaken on *MARC1* (http://arc.leeds.ac.uk/systems/marc1/), part of the High Performance Computing and Leeds Institute for Data Analytics (LIDA) facilities at the University of Leeds, UK, a large memory node cluster. Genome alignment followed the same principal commands as for WES analysis, with example commands in Appendix 2.

Patient	Alignment reference	Aligned by
F1377	GRCh37, GRCh38	Author, Edinburgh Genomics
F1310	GRCh37, GRCh38	Author, Edinburgh Genomics
F1335	GRCh37	Edinburgh Genomics
F1071	GRCh37	Edinburgh Genomics
F1298	GRCh37	Edinburgh Genomics
F1374	GRCh37	Edinburgh Genomics
F1369	GRCh37	Edinburgh Genomics
F1352	GRCh37	Edinburgh Genomics
F1403	GRCh37	Edinburgh Genomics
F1404	GRCh37	Edinburgh Genomics

Table 2-4 Patients whole-genome sequenced by Edinburgh Genomics and their subsequent alignment and analysis.

2.4 Bioinformatics

2.4.1 Literature searches

Investigation of candidate genes, techniques and bioinformatics, to name but a few topics, was performed using free online tools. The online tool PubMed (https://www.ncbi.nlm.nih.gov/pubmed) provided an invaluable database of journal articles for literature searching. A comprehensive reference of human genes and genetic phenotypes was obtained using OMIM (https://www.omim.org). GeneCards (http://www.genecards.org) database was used to interrogate all annotated and predicted human genes.

2.4.2 Bioinformatics online browser tools

UCSC Genome Browser (http://genome.ucsc.edu/) was used as an essential tool in the bioinformatics search of genomic regions. Additional information, including location of SNPs through links to dbSNP and ExAC, location of microsatellites, and alternate transcripts, was also gained from this database. Ensembl (http://www.ensembl.org/index.html) was also used as a genome browser. The control exome and genome databases, ExAC (http://exac.broadinstitute.org) (Lek *et al.*, 2016) and gnomAD (http://gnomad.broadinstitute.org) were used to interrogate variant allele frequency, alongside dbSNP version 146 (https://www.ncbi.nlm.nih.gov/projects/SNP/) (Sherry *et al.*, 2001).

2.4.3 Integrative Genomics Viewer (IGV)

IGV (http://software.broadinstitute.org/software/igv/home) (Robinson *et al.*, 2011; Thorvaldsdóttir, Robinson and Mesirov, 2013) proved an invaluable high-

performance genome-wide visualization tool in the analysis of NGS data. The tool loads aligned and indexed BAM files for both WES and WGS data and allows exploration of read depth, coverage, structural variation and visualization of single base pair substitutions. Reads were analysed in more depth, using additional features including coloured reads to flag anomalous pair orientations and highlighting paired reads.

2.4.4 Bioinformatics mutation prediction software

2.4.4.1 Polyphen-2

Polymorphism Phenotyping-2, known as Polyphen-2 (http://genetics.bwh.harvard.edu/pph2/) is a tool that annotates nonsynonymous SNPs (Adzhubei *et al.*, 2010). Using the HumVar prediction (for Mendelian disease) of the effect of an amino acid substitution on the 3D structure and function of the resulting protein, it gives a score as follows:

0-0.4	benign
0.4-0.9	possibly damaging
0.9-1.0	probably damaging
No score	unknown

2.4.4.2 MutationTaster2

MutationTaster2 (http://www.mutationtaster.org/) predicts the effect of amino acid substitutions, short indels, intronic and synonymous alterations and variants spanning intron-exon borders (Schwarz *et al.*, 2014). Variants are ranked as follows:

disease causing

disease causing automatic

polymorphism

polymorphism automatic

neutral

The automatic label is given when the variant has been marked as a known disease-causing variant in a public database of known variation, including ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) (Landrum *et al.*, 2016) and HGMD Public (http://www.hgmd.cf.ac.uk/docs/articles.html) (Stenson *et al.*, 2014). MutationTaster2 gives information regarding genotypes found in the 1000 Genomes Project (Abecasis *et al.*, 2012) and phenotypes found in ClinVar. Variants are regarded as neutral when they are found in the homozygous state more than 4 times in 1000 Genomes.

2.4.4.3 Blosum62

Blosum62 scores are generated from a substitution matrix based on aligned proteins from approximately 2000 aligned sequence segments characterizing more than 500 groups of related proteins. Scores range from +3 to -4 and negative scores are more likely to be damaging substitutions (Henikoff and Henikoff, 1993).

2.4.4.4 PROVEAN

Protein variant effect analyzer, or PROVEAN (http://provean.jcvi.org/genome_submit_2.php), predicts the effect of nonsynonymous SNPs, indels and multiple amino acid substitutions on protein

function (Choi *et al.*, 2012). The score is alignment-based and measures the change in sequence similarity between a protein sequence before and after the introduction of an amino acid variation. Scores are as follows:

≤ -2.5 deleterious

≥ -2.5 neutral

2.4.4.5 SIFT

Sorting intolerant from tolerant, known as SIFT (http://provean.jcvi.org/genome_submit_2.php), classifies single amino acid substitutions based on closely related sequences within the relevant protein family (Kumar, Henikoff and Ng, 2009), with scores defined as follows:

≤ 0.05 damaging

≥ 0.05 tolerated

2.4.4.6 MutPred 1.2

MutPred 1.2 (http://mutpred.mutdb.org) is an online tool that predicts whether an amino acid substitution is disease-associated or neutral. There is a general score (g) calculated which is the probability that an amino acid substitution is deleterious or disease-associated, and also a P-value (p) that certain structural and functional properties are unaffected. Scores are classified as follows:

Scores with $g > 0.5$ and $p < 0.05$ are referred to as **actionable hypotheses**.

Scores with $g > 0.75$ and $p < 0.05$ are referred to as **confident hypotheses**.

Scores with $g > 0.75$ and $p < 0.01$ are referred to as **very confident hypotheses**.

2.4.4.7 Combined Annotation-Dependent Depletion (CADD) score

CADD Score (http://cadd.gs.washington.edu/score) (Kircher *et al.*, 2014), or C score, is an algorithm that scores the deleteriousness of SNPs and insertions/deletions variants in the human genome. The score it generates is not a measure of sequence conservation, but is a pre-computed score worked from the differentiation between the annotations of fixed or nearly fixed human-derived alleles and those of simulated variants. The scores rank known deleterious variants within individual genomes, allowing a genome-wide prediction of variant effect. Scores range from 1-99 and a scaled C-score of 10 or more indicates that variants are predicted to be the top 10% most deleterious substitutions that are present in the human genome, a score of 20 or greater indicates the top 1% most deleterious and 30 or higher the top 0.1%.

2.4.4.8 Functional Analysis through Hidden Markov Models (FATHMM)

FATHMM (http://fathmm.biocompute.org.uk) (Shihab *et al.*, 2013) permits scoring of both coding variants and non-coding variants. The scoring can be both weighted/species-specific (a score derived from how frequent functionally neutral or disease-causing amino acid substitutions map on to conserved protein domains), or unweighted/species-independent (automated multiple sequence alignment of homologous sequences which is then analysed by Hidden Markov Models to capture position-specific information). FATHMM is similar to both SIFT and PolyPhen-2 as Hidden Markov Models are applied. Predictions are scored as P-values in a range 0-1. A variant score of >0.5 is considered deleterious, while those listed below 0.5 are predicted to be neutral or benign. P-values that are close to the extremes, be that 0 or 1, are the highest-confidence predictions with highest accuracy, according to the model.

2.4.5 Splicing prediction tools

Human Splice Finder (HSF) (http://www.umd.be/HSF3/) was used to determine whether intronic variants were likely to affect gene splicing. HSF predicts a variant's effect on splicing motifs, such as the acceptor and donor splice sites, and exonic splicing enhancers (ESE) and exonic splicing silencers (ESS) both of which either enhance or suppress splicing respectively. Alamut Batch (http://www.interactive-biosoftware.com/alamut-batch/) was used to annotate variants and assign splice predictions, including HSF, MaxEntScan, NNSPLICE, SpliceSiteFinder and GeneSplicer. Ensembl's VEP was also used to functionally annotate variants and report on regulatory region consequences.

2.4.6 AgileMultiIdeogram

AgileMultiIdeogram is an autozygosity mapping tool available at DNA@Leeds (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/). The software inputs WES VCF files and generates a circular ideogram showing locations of autozygous regions from multiple individuals across Chromosomes 1-22.

2.4.7 HomoloGene

NCBI's HomoloGene (https://www.ncbi.nlm.nih.gov/homologene) was used for the multiple sequence alignment of proteins to examine the level of conservation among other species for the amino acid of interest. Protein homologues, paralogues and orthologues are outputs from this tool.

2.4.8 GENEVESTIGATOR

Genevestigator (https://genevestigator.com/gv/index.jsp) (Hruz *et al.*, 2008) was used as a search engine for curated gene expression and contains publicly available microarray and RNA-seq data.

2.4.9 Expasy translate

Expasy translate (https://web.expasy.org/translate/) was used to translate all nucleotide DNA or RNA sequences into a protein sequence.

2.4.10 Protein analysis tools

2.4.10.1 Simple Modular Architecture Research Tool (SMART)

SMART (http://smart.embl-heidelberg.de) (Schultz *et al.*, 1998) was used to identify and analyse domains in proteins of interest. The SMART output domains are based on phyletic distributions, functional class, tertiary structures and functionally important residues.

2.4.10.2 Protter

Protter (http://wlab.ethz.ch/protter/start/) (Omasits *et al.*, 2014) was used as a tool for the visualization of proteins and the predicted annotation of amino acid sequence features.

2.4.10.3 Prediction of Protein Sorting Signals and Localisation Sites in Amino Acid Sequences (PSORTII)

PSORTII (https://psort.hgc.jp) (Horton and Nakai, 1997) was applied as a prediction of protein localization sites in cells.

2.4.10.4 SignalP

The SignalP server (http://www.cbs.dtu.dk/services/SignalP/) (Nielsen *et al.*, 1997) was used to predict the presence and location of signal peptide cleavage sites in amino acid sequences. Based on a combination of several artificial neural networks, the tool incorporates a prediction of cleavage sites and the presence/absence of a signal peptide.

2.4.10.5 Phobius

Phobius (http://phobius.sbc.su.se) (Käll, Krogh and Sonnhammer, 2004) was used for the prediction of transmembrane topology and signal peptides from the amino acid sequence of a protein.

Chapter 3 The identification of novel mutations in *SLC38A8*

3.1 Introduction

Before this study commenced, recessive mutations in *SLC38A8* had recently been identified and published in patients with foveal hypoplasia and optic nerve misrouting in the absence of albinism (Al-Araimi *et al.*, 2013; Poulter *et al.*, 2013) (Section 1.8.2). Since the gene was identified, further patients with the same or similar phenotype had been recruited to the Leeds study for mutation screening. The aim of this study was to screen this new cohort to expand the mutation spectrum of *SLC38A8* and to clarify the phenotype associated with mutating this gene. Patients that were screened in the original study (F1071, F1287, F1288, F1298, F1310, F1405, F1419 and F1421), but who did not have a mutation in the coding region or splice sites of *SLC38A8*, are also included in this study (investigated in Chapters 4 and 5) and presented in Table 3 1.

3.2 Results

3.2.1 Screening of *SLC38A8* in a novel patient cohort

An international cohort of fourteen new patients (F1335, F1351, F1352, F1368, F1369, F1371, F1372, F1373, F1374, F1377, F1378, F1402, F1403 and F1404), suspected of having *SLC38A8*-related disease (Section 2.2) (Table 3 1), was

amplified by PCR and Sanger sequenced (Section 2.3.4.2, 2.3.9) using primers previously designed to target the 10 exons and splice sites of *SLC38A8* (Poulter *et al.*, 2013) (Appendix 12). All of the new patients were sporadic cases, with the exception of one who had an affected sibling, labelled as F1372 and F1373 in Figure 3-1. A pedigree structure was available for members of three additional families as shown in Figure 3-1; two probands, F1374 and F1335, were from consanguineous partnerships. Full details of the cohort are in Table 3-1 and pedigrees for patients that was known are presented in Figure 3-1.

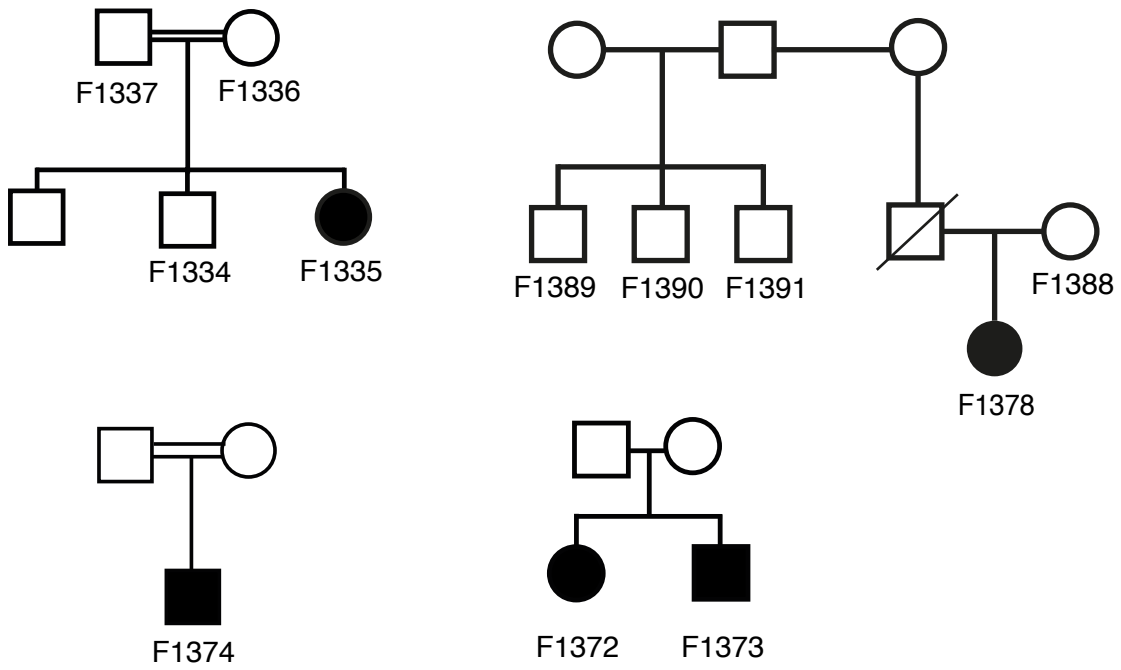


Figure 3-1 Pedigrees of cases with foveal hypoplasia.
Two probands were from consanguineous partnerships (F1335 and F1374). F1372 had an affected sib (F1373).

Patient	Ethnicity	Sex	Parental consanguinuity	Relatives	Foveal hypoplasia	Nystagmus	Optic nerve misrouting	ASD	Iris transillumination
F1335	South Asian	F	Yes	sibling F1334, mother F1336, father F1337. All unaffected.	+	NR	NT	NR	—
F1351	North European	M	No	NR	+	NR	NT	NR	—
F1352	East Asian	M	No	NR	+	NR	NT	NR	—
F1368	USA (Caucasian)	M	No	NR	+	NR	NT	+	—
F1369	USA (Caucasian)	F	No	NR	+	NR	NT	NR	—
F1371	West European	F	No	NR	+	NR	NT	NR	—
F1378	South European	F	No	mother F1388, 3 half-uncles F1398, F1390, F1391. All unaffected.	+	NR	NT	NR	—
F1372	USA (Jewish)	F	No	sibling of affected F1373	+	NR	NT	NR	—
F1373	USA (Jewish)	M	No	affected subling F1372	+	NR	NT	NR	—
F1374	South Asian	M	Yes	DNA not available for unaffected parents	+	NR	+	NR	—
F1377	West European	M	No	NR	+	+	+	NR	—
F1402	West European	F	No	NR	+	+	NT	NR	—
F1403	West European	M	No	NR	+	NR	NT	NR	—
F1404	West European	M	No	NR	+	+	+	NR	—
F1071	South Asian	M	No	NR	+	NR	+	+	—
F1287	South Asian	F	Yes	affected sibling F1288, unaffected sister F1289, unaffected father F1286	+	NR	NT	NR	—
F1288	South Asian	F	Yes	affected sibling F1287, unaffected sister F1289, unaffected father F1286	+	NR	NT	NR	—
F1298	European	F	No	NR	+	+	+	NR	—
F1310	West Asian (Iraqi)	F	Yes	NR	+	+	NT	NR	—
F1405	European	M	No	NR	+	NR	—	NR	—
F1419	NR	NR	No	NR	+	+	NT	NR	—
F1421	NR	NR	No	NR	+	NR	NT	NR	—

Table 3-1 Details of patient cohort screened for *SLC38A8* mutations.

Patients were recruited by clinicians across multiple national and international centres. Phenotypic data was provided by the clinician. Abbreviations are as follows: F female; M male; + affected; - unaffected, NT not tested, NR not reported.

Genomic DNA from 13 probands in the new cohort (F1335, F1351, F1352, F1368, F1369, F1371, F1372, F1374, F1377, F1378, F1402, F1403 and F1404) was amplified by PCR and Sanger sequenced (Table 3 1). This analysis identified variants in *SLC38A8* which looked pathogenic in two members of the cohort; a heterozygous missense change in case F1377 (c.2T>C, p.(Met1?)) and a homozygous missense change in case F1368 (c.848A>C, p.(Asp283Ala)) (Table 3-2).

Patient	Location on Chr16	cDNA change (Zygosity)	Protein change	dbSNP rs ID	Allele count & frequency	No. of reported homozygotes
F1377	84,075,761	c.2T>C (Het)	p.(Met1?)	-	2/240142 8.328e-6	0
F1368	84,050,850	c.848A>C (Hom)	p.(Asp283Ala)	rs139373929	76/277066 2.742e-4	1

Table 3-2 Rare *SLC38A8* variants identified in two patients with foveal hypoplasia.

Two novel mutations were identified by Sanger sequencing in two patients. Their locations on chromosome 16 are shown, (reference sequence NM_001080442.1), using human reference genome build GRCh37. When available, rs number (dbSNP ID), allele count & frequency and the number of homozygotes reported are listed using data sourced from the gnomAD browser (last accessed 03/08/2017).

In individual F1368, a novel homozygous missense change was identified in exon 7 of *SLC38A8,* c.848A>C (NM_001080442.1) (Table 3-2). This transversion results in the substitution of aspartic acid for alanine at amino acid 283, p.(Asp283Ala) (Figure 3-2). F1368 is a Caucasian male child from the USA with reported foveal hypoplasia and ASD. No visual evoked potential examination had been undertaken in clinic, so that chiasmal misrouting defects were not ascertained. Segregation analysis of the variant was not possible, as family DNA was not available. Investigation of the p.(Asp283Ala) variant frequency in publicly available databases (Section 2.4.2),

including dbSNP and gnomAD, revealed that it is present in both (rs139373929), with a reported allele frequency of 76/277066 (0.0002743) in gnomAD (last accessed on 03/08/2017) (Table 3-2). The variant was found in a homozygous state once in gnomAD in a case of Ashkenazi Jewish origin (1/10142) (Table 3-2), and the allele count was highest in this population overall (63/10142, MAF 0.006).

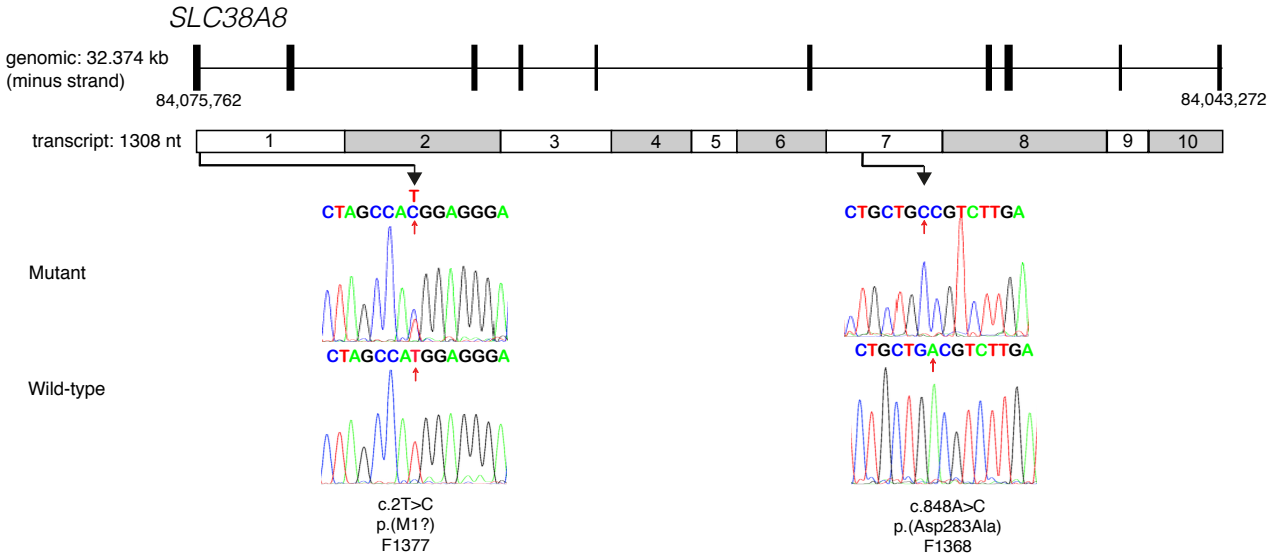


Figure 3-2 Sanger sequencing electropherograms of the two *SLC38A8* variants identified in patients F1368 and F1377.

Schematic representation of the *SLC38A8* genomic structure and transcript of 10 coding exons, showing the location of the two variants identified by Sanger sequencing. Electropherogram sequencing traces of the two variants are mapped at the approximate location on chromosome 16: homozygous missense c.848A>C, p.(Asp283Ala) (Patient F1368) and heterozygous missense c.2T>C, p.(Met1?) (Patient F1377). Variants mapped using reference sequence NM_001080442.1 and human reference genome build GRCh37.

Analysis of this substitution using a variety of different pathogenicity prediction tools (Section 2.4.4), also suggests that this missense change is likely to be pathogenic. PolyPhen2, SIFT, MutationTaster2, Blosum62, MutPred and PROVEAN all predicted this variant to be pathogenic (Table 3-3). CADD outputs are not given a binary

pathogenic/non-pathogenic output, but this variant generated a CADD deleteriousness score of 20.3, indicating that is it among the top 1% of the most deleterious substitutions in the human genome. Only one tool, the FATHMM weighted score, predicted that this variant was tolerated and not pathogenic (Table 3-3). Conservation analysis of the nonsynonymous substitution was undertaken using HomoloGene (Section 2.4.7). The results showed the aspartic acid residue to be fully conserved through evolution down to the zebrafish, indicating the importance of the aspartic acid at this position (Figure 3-3). Investigation of the aspartic acid residue in human SLC38A8 paralogues showed that the residue is only conserved in SLC38A7, SLC38A1 (hATA1) and SLC38A11 (Figure 3-3). The alanine residue is also not present in any of the paralogues at this position (Figure 3-3). Use of the SMART tool (Section 2.4.10.2) predicted the topological structure of SLC38A8 to have 11 transmembrane helices, further supported by the prediction tools Protter (Section 2.4.10.5) and Phobius (Section 2.4.8.5). SMART predicted the variant to lie on the extracellular side of the channel (Figure 3-4).

Position on Chromosome 16	cDNA change	Protein change	PolyPhen-2	MutationTaster2	Blosum62	SIFT	MutPred	PROVEAN	CADD Score	FATHMM (weighted score)
84,075,760	c.2T>C	p.(Met1?)	Possibly damaging (0.546)	Disease causing (probability 0.99)	-1	Damaging	Deleterious mutation (probability 0.968)	Neutral (-1.956)	23.0	Tolerated (2.79)
84,050,850	c.848A>C	p.(Asp283Ala)	Probably damaging (1)	Disease causing (probability 0.99)	-2	Damaging	Deleterious mutation (probability 0.745)	Deleterious (-7.245)	20.3	Tolerated (4.42)

Table 3-3 Bioinformatics analysis of novel *SLC38A8* variants identified in cases F1377 and F1368.

The heterozygous missense variant, c.2T>C was identified in patient F1377 and the homozygous c.848A>C variant was identified in patient F1368. Mutations are annotated using the reference sequence NM_001080442.1 and human reference genome build GRCh37. URLs: PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/ (Adzhubei *et al.*, 2010); MutationTaster2, http://www.mutationtaster.org/ (Schwarz *et al.*, 2010; Schwarz *et al.*, 2014); SIFT, http://sift.jcvi.org/ (Ng and Henikoff, 2003); Blosum62 (Henikoff and Henikoff, 1993); PROVEAN, http://provean.jcvi.org/ (Choi *et al.*, 2012); MutPred 1.2, http://mutpred.mutdb.org/ (Li *et al.*, 2009); CADD Score, http://cadd.gs.washington.edu/score (Kircher *et al.*, 2014); FATHMM score (weighted) http://fathmm.biocompute.org.uk/index.html (Gough *et al.*, 2001). Blosum62 scores range from +3 to -3; negative scores are more likely to be pathogenic substitutions. CADD score of ≥20 means that the variant is in the top 1% most deleterious mutations in the genome.

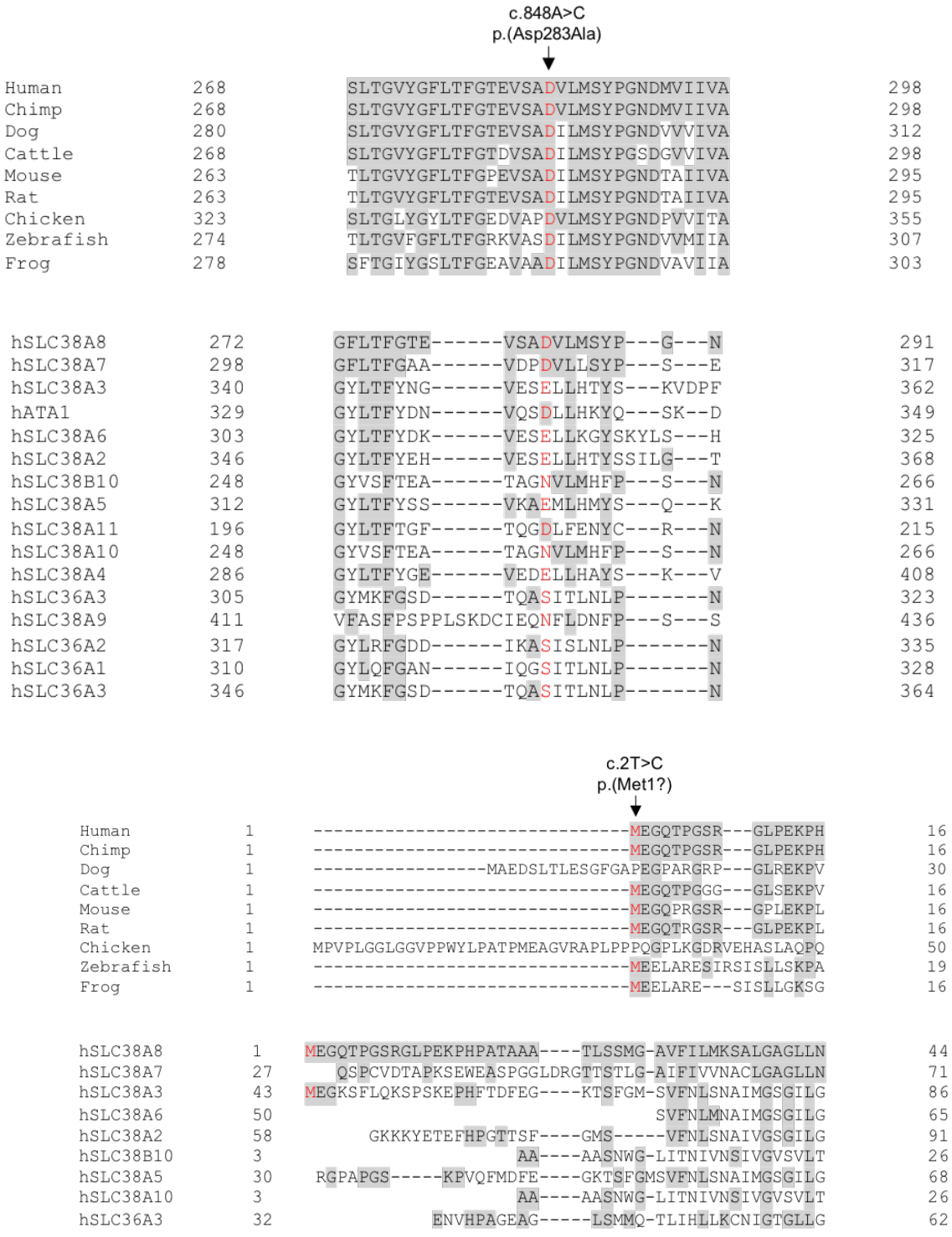
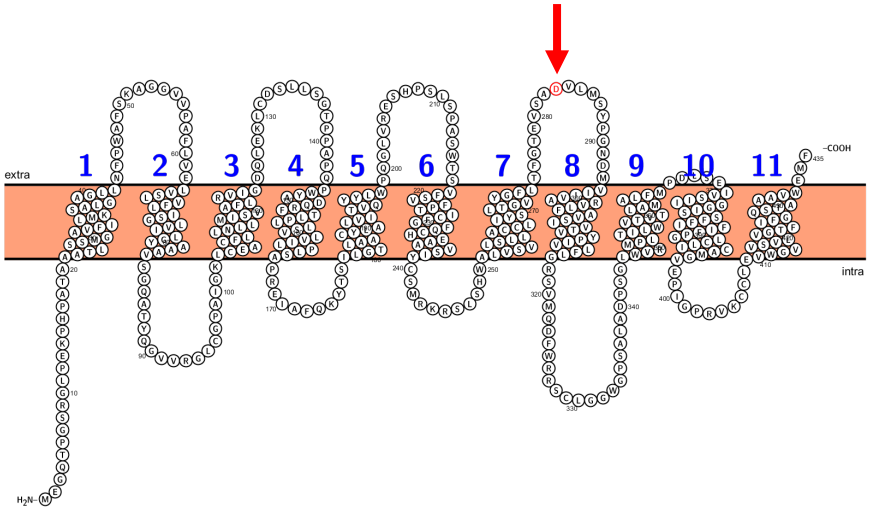


Figure 3-3 Protein sequence alignment of human SLC38A8 with homologues from human and other species for the residues surrounding the two novel variants identified in cases F1368 and F1377.

The shaded amino acid residues indicate those that are shared with the human residue of SLC38A8 at the same position. The residues substituted in the two patients are coloured red. Case F1368 c.848A>C p.(Asp283Ala) and case F1377 c.2T>C p.(Met1?). Sequences of *SLC38A8* used: Human (Homo sapiens) NP_001073911.1; Chimpanzee (*Pan troglodytes*) XP_003315260.1; Dog (*Canis lupus familiaris*) XP_005620941.1; Cattle (*Bos taurus*) XP_015331173.1; Mouse (*Mus musculus*) NP_001009950.1; Rat (*Rattus norvegicus*) NP_001178938.1; Chicken (*Gallus gallus*) XP_003641956.1; Zebrafish (*Danio rerio*) NP_001138287.1; Frog (*Xenopus tropicalis*) NP_001037900. SLC38A8 human (h) paralogue sequences used: SLC38A8 NP_001073911; SLC38A7 NP_060701; SLC38A3 NP_006832; SLC38A1 (hATA1) NP_109599; SLC38A6 NP_722518; SLC38A2 NP_061849; SLC38B10 NP_612637; SLC38A5 NP_277053; SLC38A11 NP_775783; SLC38A10 NP_001033073; SLC38A4 NP_060488; SLC36A3 NP_001138489; SLC38A9 NP_775785; SLC36A2 NP_861441; SLC36A1 NP_510968; SLC36A3 NP_001138489.

A



B

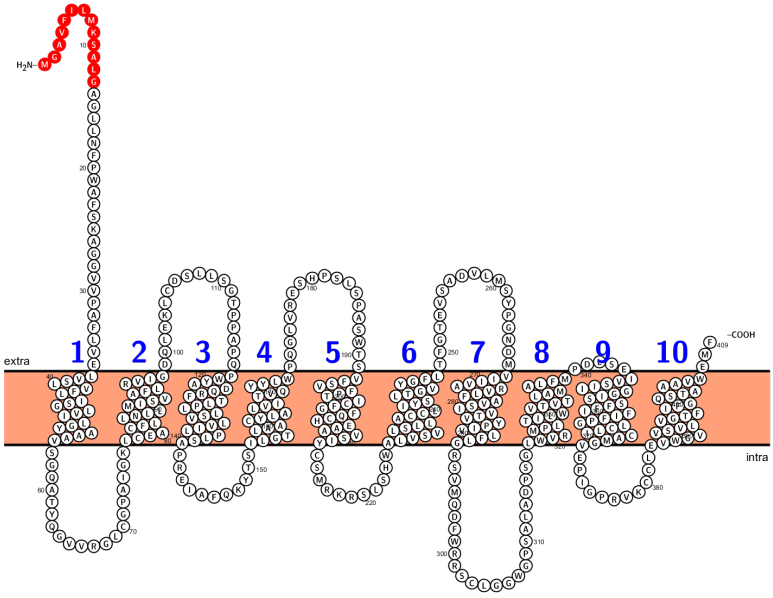


Figure 3-4 Predicted 2D model showing patient SLC38A8 variants.

Membrane topological structure programmes predict SLC38A8 has 11 transmembrane helices. (A) WT SLC38A8. The p.(Asp283Ala) mutation is predicted to lie on the extracellular surface of the channel (highlighted by red arrow). (B) Schematic representation of the predicted SLC38A8 protein introduced by the c.2T>C, p.(Met1?) variant which has an alternative ATG translation initiation site located 78bp away from the original TIC (ΔSLC38A8(26aa)). The transmembrane domains predicted by the tools are numbered and the predicted N-terminal signal peptide sequence is highlighted in red. One of the 11 transmembrane domains is abolished upon protein truncation. A variety of tools were used and the consensus is shown as generated by Protter (Section 2.4.8.2).

In patient F1377, a novel heterozygous *SLC38A8* missense variant in the start codon was identified (c.2T>C; p.(Met1?)) (Table 3-2, Figure 3-2). The affected individual was a Caucasian 2-year old male from the UK diagnosed with congenital nystagmus, foveal hypoplasia, probable optic nerve misrouting, and no signs of reduced pigmentation. Segregation analysis of the variant was not possible as family DNA was not available. Interrogation of publicly available databases showed this variant to be very rare. It is absent from dbSNP, but the variant is present in gnomAD, with the allele reported only in a heterozygous state in 2/240142 alleles (Table 3-2) (last accessed 03/08/2017). PolyPhen2, SIFT, MutationTaster2, Blosum62 and MutPred all predicted this variant to be pathogenic, but PROVEAN predicted the variant as Neutral and the weighted FATHMM score was noted as Tolerable. The variant has a CADD score of 23.0 (Table 3-3). Conservation analysis of the substitution, using HomoloGene, showed the methionine residue to be conserved in all species down to the zebrafish, with the exception of chicken and dog (Figure 3-3). SLC38A3 is the only paralogue that shares the same coding start site as SLC38A8 (Figure 3-3).

No second mutation for case F1377 was found within the coding sequence and flanking 10 bp of intronic sequence of each exon. Therefore, the search was widened to all sequence data generated by Sanger sequencing. Six heterozygous intronic variants were identified in this search (Appendix 13). However, only one of these variants was found to be rare (≤0.01) when interrogating the gnomAD and dbSNP databases. This variant, c.189+28G>A (rs144889482), was not identified in any other patients in the cohort and had a reported allele frequency of 1398/233994 (0.005975) in gnomAD (last accessed 03/08/2017). The variant was interrogated using the Human Splice Finder (HSF) online bioinformatics tool (v3.0) to investigate

potential aberrations of splicing signals. To maximise the detection of auxiliary motifs such as exonic splicing silencers (ESS), HSF employs various prediction matrices and integrates the scores. The variant was predicted to break an ESS site, with five prediction algorithms utilised by HSF to yield this result (Table 3-4). The altered ESS site was predicted to be caused by the loss of four motifs, including a hnRNPA1 motif (Table 3-4). The software interpreted the variant to have 'probably no impact on splicing'; it was therefore not pursued further at this stage.

cDNA position	Motif	Reference sequence	Mutant sequence	HSF prediction	Ensembl annotation
c.189+28G>A	-ESS	cctggggacgggt	cctgggaacgggt	-ESS -IIE -Fas-ESS hexamers -hnRNPA1 +PESS octamer	Modifier

Table 3-4 Human Splice Finder prediction for the rare intronic variant identified in Patient F1377.

HSF prediction for the variant c.189+28G>A identified by Sanger sequencing. The position of the variant is underlined. – indicates the motif was abolished by the mutation; + indicates a new site was created by the mutation. The algorithms and matrices used by HSF to identify the motifs were: exonic splicing silencer (ESS) motifs from (Sironi *et al.*, 2004), intron identity elements (IIEs) (Zhang *et al.*, 2008), Fas-ESS hexamers (Wang *et al.*, 2004), hnRNPA1 motifs from HSF, putative exonic splicing enhancers (PESS) octamers (Zhang and Chasin, 2004). The 'modifier' Ensembl annotation can be defined as non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact.

To investigate the potential for a novel translation start site introduced by the c.2T>C, p.(Met1?) variant in case F1377, the sequence upstream and downstream of the WT primary start codon was interrogated. No alternative upstream ATG codons were identified. The distance between the translation initiation codon (TIC)

and the next downstream ATG codon of *SLC38A8* is 78 bp, or 26 aa (ΔSLC38A8(26aa)). This would result in a truncated SLC38A8 protein size of 409 aa, compared to the WT size of 435 aa. The topology prediction tools Protter (Section 2.4.10.2) and SMART (Section 2.4.10.1), were applied in the analysis of predicted domains, repeats, motifs and features of *SLC38A8* between WT and the potentially truncated protein (ΔSLC38A8(26aa)). Also, PSORTII (Section 2.4.10.3), SignalP (Section 2.4.10.4) and Phobius (Section 2.4.10.5) were applied in the prediction of protein localisation sites and signal peptide identification.

The variety of tools were used to generate a consensus schematic to show the location of the transmembrane domains as well as the topology, for full length SLC38A8 (Figure 3-4). Protter predicted that ΔSLC38A8(26aa) would eliminate the first transmembrane domain and introduce an extracellular N-terminal peptide sequence (Figure 3-4). It is not yet known whether WT SLC38A8 has a signal peptide, as localisation signals for the WT protein have not yet been defined. PSORTII analysis of WT SLC38A8 confirmed the absence of a signal peptide, however, the PSORTII analysis of ΔSLC38A8(26aa) predicted a cleavable signal peptide located at amino acid positions 1 to 46, similar to Protter's prediction. However, the application of SignalP and Phobius failed to detect a signal peptide for both WT and ΔSLC38A8(26aa). SMART protein analysis did not predict the introduction of an N-terminal signal peptide for the truncated protein ΔSLC38A8(26aa). Instead, this tool predicted a shift in the positioning of the transmembrane domains.

To understand the importance of the protein's initial 26aa, HomoloGene was used to analyse the start of SLC38A8 amongst orthologues. As shown in Figure 3-5, there is a relatively high level of conservation through evolution in the first 26aa of SLC38A8. As mentioned previously, dogs and chickens have a different start site to the protein. Human paralogues of SLC38A8 do not appear to be conserved in this region of the protein (Figure 3-5).

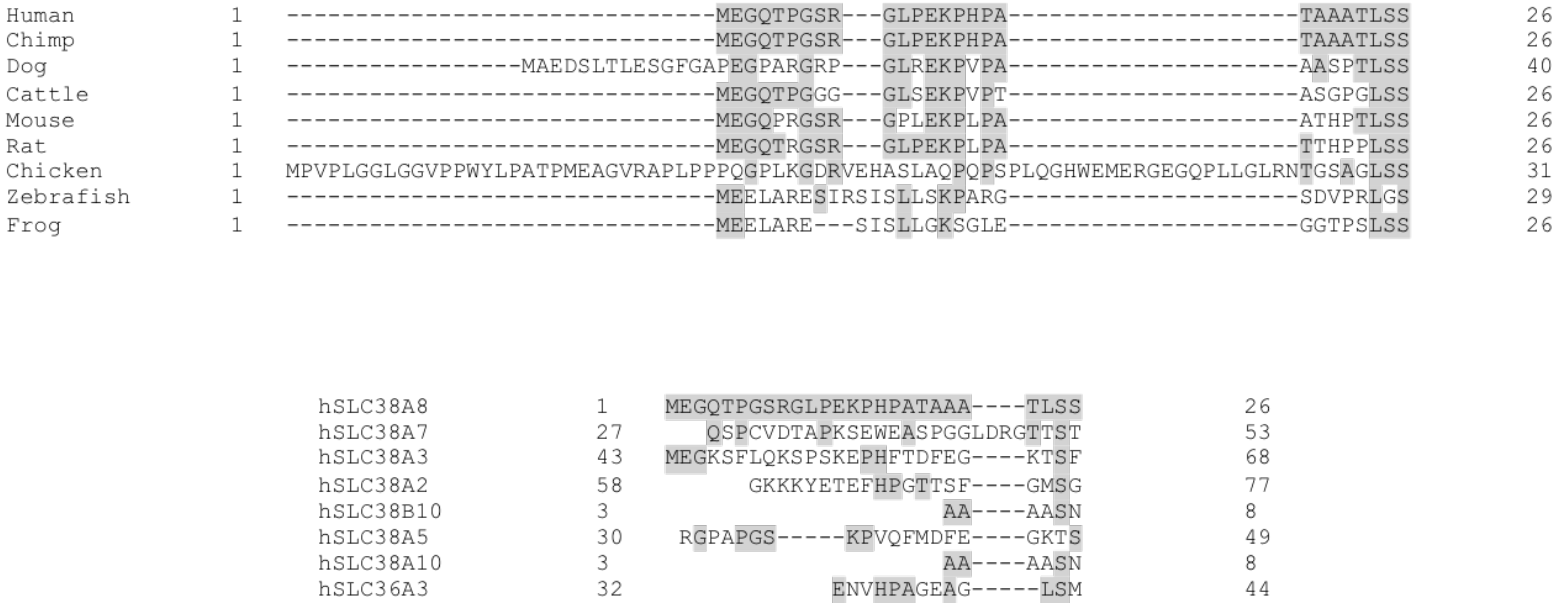


Figure 3-5 HomoloGene multiple sequence alignment of orthologous *SLC38A8* sequences and *SLC38A8* human paralogues at the start of the protein.

The shaded amino acid residues indicate those that are shared with the human *SLC38A8* residue at the same position. (Top pane) Sequences of *SLC38A8* orthologues used: Human (Homo sapiens) NP_001073911.1; Chimpanzee (*Pan troglodytes*) XP_003315260.1; Dog (*Canis lupus familiaris*) XP_005620941.1; Cattle (*Bos taurus*) XP_015331173.1; Mouse (*Mus musculus*) NP_001009950.1; Rat (*Rattus norvegicus*) NP_001178938.1; Chicken (*Gallus gallus*) XP_003641956.1; Zebrafish (*Danio rerio*) NP_001138287.1; Frog (*Xenopus tropicalis*) NP_001037900. (Lower pane) SLC38A8 human (h) paralogue sequences used: SLC38A8 NP_001073911; SLC38A7 NP_060701; SLC38A3 NP_006832; SLC38A6 NP_722518; SLC38A2 NP_061849; SLC38B10 NP_612637; SLC38A5 NP_277053; SLC38A10 NP_001033073; SLC36A3 NP_001138489.

No further potentially pathogenic mutations were identified in any of the remaining 11 patients. Ten variants were identified in the coding sequence and flanking 10bp of *SLC38A8* in the 11 remaining patients (Table 3-5 and Appendix 13), however, the majority of these were too common in publicly available databases to be causative mutations for this rare disorder. Two of the variants (F1403; c.159C>T, p.(Gly53=) and F1372; c.406A>T, p.(Ser136=))) were considered to be rare or unreported in the publicly available databases, as shown in Table 3-5. However, the variants were synonymous and so were not prioritised and carried forward at this stage.

Patient	Location on Chr16	cDNA change	Protein change	dbSNP rs ID	Allele count & frequency (gnomAD)	No. of reported homozygotes
F1403 (Het)	84,075,604	c.159C>T	p.(Gly53=)	rs147597105	462/273576 0.001689	0
F1372 (Het)	84,067,055	c.406A>T	p.(Ser136=)	-	-	-

Table 3-5 Rare heterozygous non-pathogenic variants identified in the coding region of *SLC38A8*.

Rare heterozygous (Het) variants identified in patients from this cohort are shown here with their locations on chromosome 16 (reference sequence NM_001080442.1), using human reference genome build GRCh37. When available, rs number (dbSNP ID), allele count & frequency and the number of homozygotes reported are listed using data resourced from the gnomAD browser (last accessed 03/08/2017).

3.2.2 Haplotype analysis in *SLC38A8* negative patients

Alongside the mutation screening of the new patient cohort, selected patients who had tested negative for mutations in *SLC38A8* in the original gene identification study (Poulter *et al.*, 2013) were investigated using microsatellite haplotype analysis. The full details of patients who had previously tested negative for mutations in

SLC38A8 are documented in **Error! Reference source not found.**. Haplotype analysis was undertaken in a consanguineous sib pair, F1288 and F1287, to check whether they were homozygous or shared haplotypes across the *SLC38A8* locus. Haplotype analysis was also performed in a sporadic case from a consanguineous relationship, F1310. This individual was not from the patient cohort of 13, but had previously been analysed by whole genome SNP microarray analysis which had highlighted a large region of homozygosity spanning *SLC38A8* (Sergouniotis, 2012), although Sanger sequencing had failed to detect the causative mutation (Poulter *et al.*, 2013). Unfortunately, the SNP data was not available for this patient, so haplotype analysis was undertaken to confirm the reported homozygosity.

Patient F1290, from the original gene identification study, was known to harbour a large deletion encompassing *SLC38A8* that was missed by Sanger sequencing and was only detected when a pseudo-homozygous variant identified in F1290 was not present in her father (Poulter *et al.*, 2013) (Section 1.5.4). Similar large deletions encompassing the neighbouring gene, *DNAAF1*, have also been reported in patients with primary ciliary dyskinesia (PCD) (Loges *et al.*, 2009). F1290 was analysed in this experiment to try and establish a PCR-based assay to detect the deletion and aid mapping of the breakpoints, so it could be screened for in additional patients. Finally, haplotype analysis was also undertaken on consanguineous individual F1335 from the present study, who was noted to have homozygous SNPs within *SLC38A8* from Sanger sequencing (Appendix 13). This case was included in the genotyping assay to confirm homozygosity across the region. The Sanger sequencing was still ongoing for the remaining cases of the new cohort. To perform

the haplotype analysis, thirteen microsatellite markers (Table 3-6), spanning nearly 10Mb (30cM) of the *SLC38A8* locus, were amplified and genotyped (Section 2.3.11).

Marker	Size (bp)	Physical Position on Chr16	Genetic Position	Left primer (5'-3') Right Primer (5'-3')	Marker Repeat
D16S3096	173-219	79,044,896-79,045,207	97.64	GATCTGGCTTACGATGATTTCTAAC CCGTGATGATGTCTGCAAC	CA
D16S505	239-261	81,674,263-81,674,567	-	GACTGTGTCTGCCCAA TCTGCCTCCATACGTG	CA
D16S3091	115-129	82,980,450-82,980,774	105.09	GGGAGATAGCCTTAAACTTTCTTAC TGTTGCTAATAACACTAGGCCA	CA
D16S402	161-187	83,293,414-83,293,898	109.59	TTTTGTAACCATGTACCCCC ATTTATAGGGCCATGACCAG	CA
D16S2625	183	84,717,619-84,717,940	116.89	TACGCAAGTCAAAGAGCCTC GGACACATGAGACCCTGTCT	GATA
D16S3061	241-253	84,934,541-84,934,804	117.55	CTACTGGTGAGGCTGAGGTG ATATCTCGGGATTTGTTGCTTTAC	CA
D16S3037	201-217	84,943,849-84,944,196	117.55	GAGCCAAGATGACACCACT GCACTGGGAACCTAAGGA	CA
D16S539	157	86,386,034-86,386,428	-	GATCCCAAGCTCTTCCTCTT ACGTTTGTGTGTGCATCTGT	GATA
D16S520	181-197	86,516,112-86,516,335	122.84	GCTTAGTCATACGAGCGG TCCACAGCCATGTAAACC	CA
D16S3074	193-217	87,084,745-87,085,073	125.07	GGTGTTTTTCTGCAACAGG TGGGTTTCCAAAGTCAGTC	CA
D16S3063	218-260	87,278,632-87,278,923	125.94	AGCTTGTTTTAGCAANAATTTC TGCCTGATTCCACCAA	CA
D16S3123	108-128	87,649,701-87,649,912	-	GCCTGGGCAATAGAACAAGACTCTG TTTAAGCAGAACAGCAACGCAATC	CA
D16S413	131-149	87,893,836-87,894,172	-	ACTCCAGCCCGAGTAA GGTCACAGGTGGGTTC	CA

Table 3-6 *SLC38A8* microsatellite marker specifications.

SLC38A8 location on chromosome 16: 84,043,272-84,075,762, (Reference sequence NM_001080442.1, human reference genome build GRCh37). Where available, the genetic map positions were taken from the deCode map (Kong *et al.*, 2002).

After resolving the fragments on the ABI3130xl, sequencer peaks were analysed using the GeneMapper software v4.0 and genotypes were manually called, then checked using Cyrillic 2.1. The results for each case are represented as haplotypes (Figure 3-6). *SLC38A8* is located between markers D16S402 and D16S2625, highlighted with an arrow (Figure 3-6). A set of representative microsatellite

genotyping electropherograms for marker D16S2625 is presented for the family of affected individual F1335 (Figure 3-7).

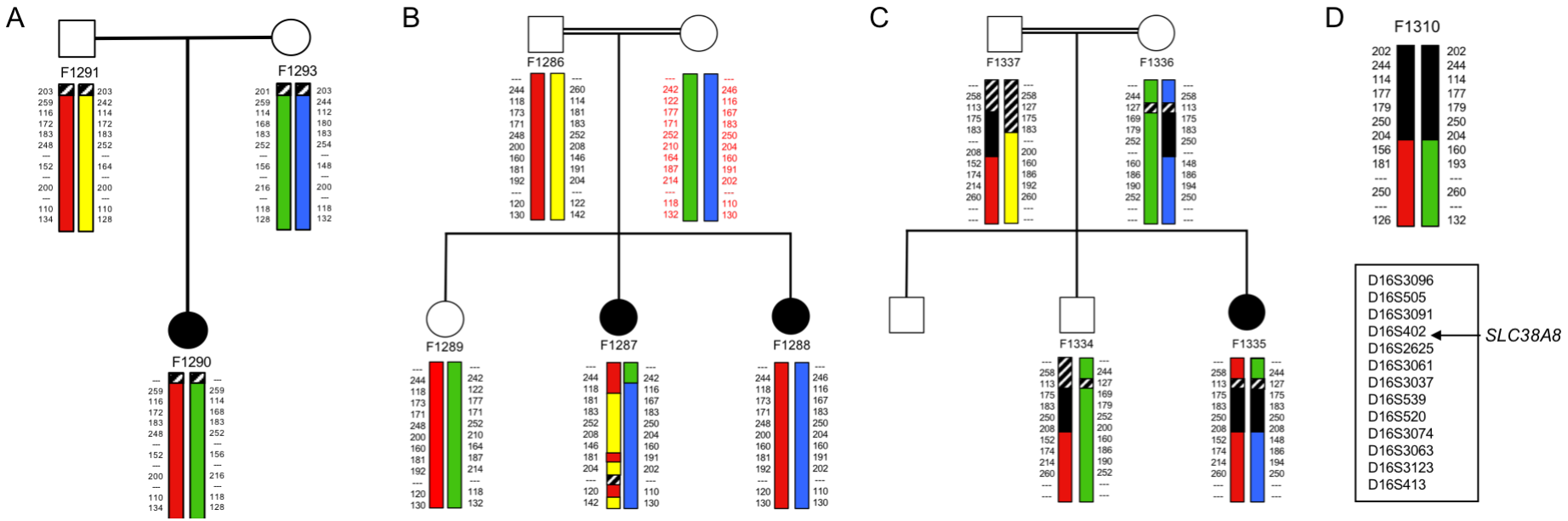


Figure 3-6 Haplotype analysis using microsatellite markers in F1290, affected sibs F1287 & F1288, F1335 and F1310.

Genotyping results are shown for 13 microsatellite markers surrounding *SLC38A8*. Genotypes have been assembled into haplotypes, represented by different coloured bars. Black regions highlight potential autozygous/hemizygous haplotypes and cross-hatched regions are uninformative, whereby parent allele origin could not be assigned. The numbers alongside each haplotype bar correspond to the size of each allele amplified by the microsatellite markers, in base pairs. The numbers coloured red for the mother of F1287 and F1288 (B) were inferred as maternal DNA was not available. The location of *SLC38A8* is shown using the arrow. The markers are displayed in order of their physical location. (A) F1290 (with known heterozygous deletion spanning *SLC38A8*); (B) F1287 and F1288; (C) F1335; (D) F1310 (case previously reported to be homozygous across at the locus).

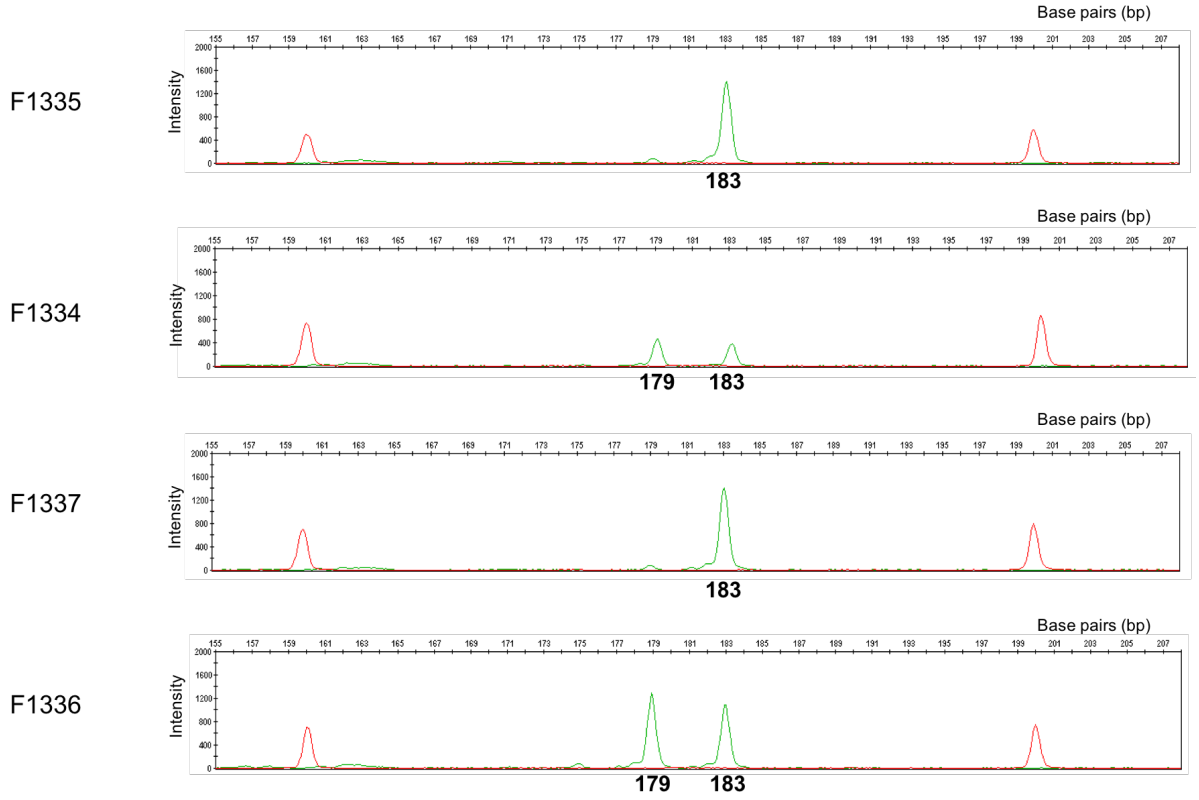


Figure 3-7 Genotype electropherogram for microsatellite marker D16S2625 in case F1335 and family.

Genotype profile of the microsatellite marker D16S2625 generated using GeneMapper 4.0. Affected individual (F1335), unaffected sibling (F1334), father (F1337) and mother (F1336).

The results showed that the deletion in case F1290 was not very large and it could not be detected by the microsatellites due to the uninformative nature of the marker closest to *SLC38A8*, D16S2625, and the limited number of markers in the region of the gene. Therefore, this assay failed to detect the breakpoints of the deletion in case F1290, so a PCR based screening assay could not be developed (Figure 3-6, A).

Haplotype analysis in the consanguineous family with affected siblings F1287 and F1288, showed that the affected siblings did not share a detectable region of

autozygosity across *SLC38A8*. Furthermore, the haplotypes were also inconsistent with compound heterozygous inheritance, as each sib had inherited a different haplotype from the father, indicating that *SLC38A8* is not the cause of the foveal hypoplasia in these siblings (Figure 3-6, B). In patient F1335, the results show a region of homozygosity surrounding the gene highlighted by markers D16S402, D16S2625, D16S3061 and D16S3037 (Figure 3-6, C). This region spans approximately 3.4Mb between markers D16S3091 and D16S539. This data is consistent with *SLC38A8* being the mutated gene in this patient and suggests that the Sanger sequencing methodology used to screen the gene may have missed the mutation. The results also showed a region of homozygosity surrounding *SLC38A8* in patient F1310, with a minimum size of 7.34Mb between markers D16S3096 and D16S539 (Figure 3-6, D).

Once the microsatellite analysis and Sanger sequencing of the new cohort was complete, the intronic sequence generated by Sanger sequencing of *SLC38A8* was analysed in all 18 mutation negative patients (from both the new and original patient cohorts), to look for potential intronic mutations. This analysis identified 2 rare variants in addition to the variant previously described in F1377 (Table 3-2) and 19 common variants (Appendix 13). All rare variants were investigated using Ensembl VEP and Human Splice Finder (Section 2.4.5), as shown in Table 3-7. None were judged to be pathogenic as the splice prediction results described them as neutral, meaning they are unlikely to have any impact on gene splicing.

Patient	Location on Chr16	cDNA change	dbSNP rs ID	Allele count & frequency (gnomAD)	No. of reported homozygotes	Ensembl annotation	HSF prediction
F1421 (Het)	84,067,120	c.389-46A>C	-	-	-	Modifier	1
F1298 (Hom)	84,050,061	c.1162+63C>G	rs190583132	147/30964 0.004747	0	Modifier	2

Table 3-7 Rare heterozygous and homozygous non-pathogenic variants identified outside of the coding regions and splice sites of *SLC38A8.*

Heterozygous (Het) and homozygous (Hom) variants located in intronic regions of *SLC38A8* identified by Sanger sequencing with a MAF ≤0.01, but not thought to be pathogenic and causative of the phenotype are shown here. Their locations on chromosome 16 are shown, (reference sequence NM-001080442.1), using human reference genome build GRCh37. When available, rs number (dbSNP ID), allele count & frequency and the number of homozygotes reported are listed using data resourced from the gnomAD browser (Last accessed 03/08/2017). The 'modifier' Ensembl annotation can be defined as non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact. Human Splice Finder (HSF) (Section 2.4.5) was used to predict splicing consequences and intronic variants were annotated as follows: 1 No impact on splicing; 2 Creation of an intronic ESE site, probably no impact on splicing.

Using the data generated from Sanger sequencing (Table 3-5, Table 3-7 and Appendix 13), the zygosity of SNPs over the region of *SLC38A8* for all mutation negative individuals was then ascertained (Table 3-8). The results show that no other patients apart from F1335 and F1310 appear to be homozygous across *SLC38A8*.

To enable haplotype analysis to be undertaken in the non-consanguineous F1372 and F1373 sib pair, F1373 was screened by Sanger sequencing as eight SNPs had already been identified in F1372. The results show that the two sibs do not share the same alleles across the locus and are unlikely to have *SLC38A8* related disease (Table 3-8 and Figure 3-8).

This analysis, alongside Sanger sequencing, left a cohort of 11 patients from this study (plus 7 from the original study), who appear not to harbour mutations in *SLC38A8* and which can be put forward for gene identification studies.

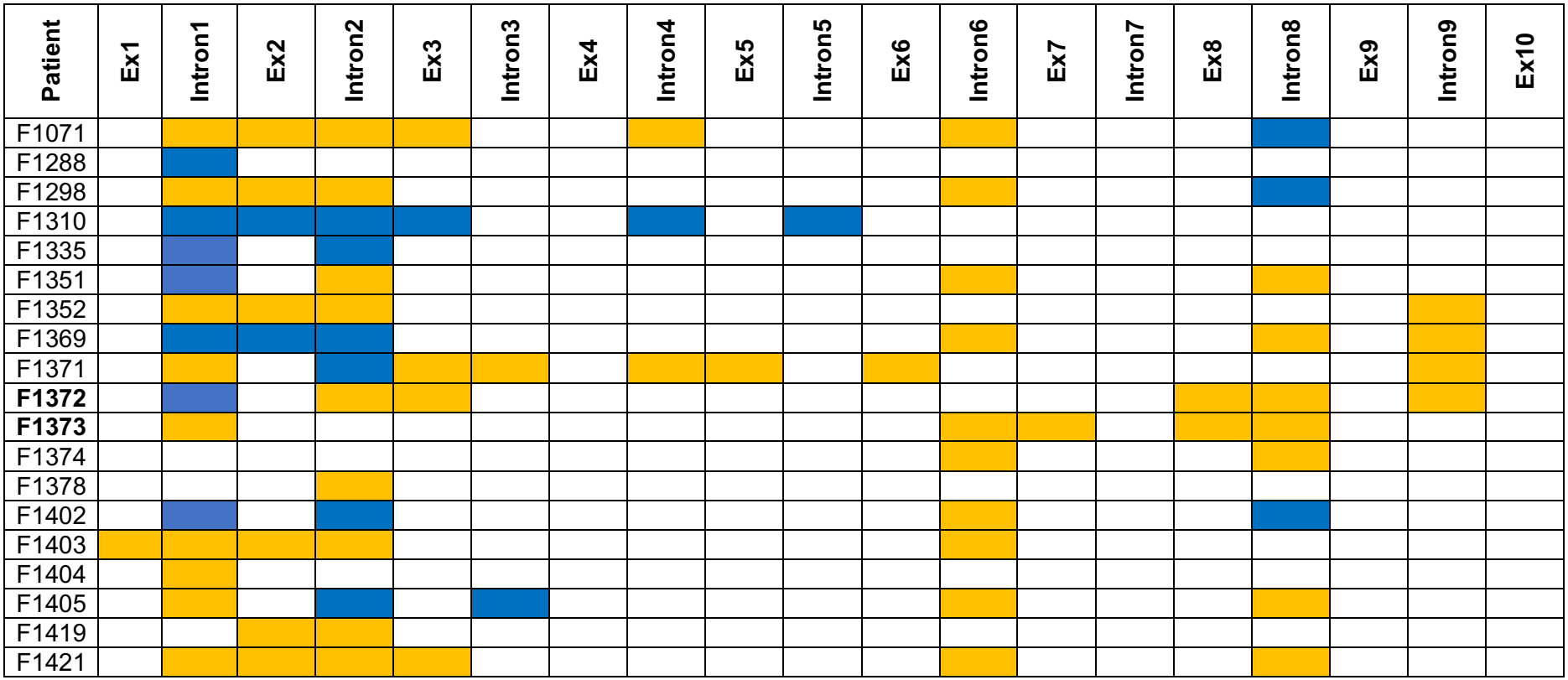


Table 3-8 Zygosity of SNPs over the region of *SLC38A8* for all mutation negative individuals.

All heterozygous (orange) and homozygous (blue) polymorphisms identified by Sanger sequencing ~3,750bp over the region of *SLC38A8*. The siblings with non-consanguineous parents, F1372 and F1373, are highlighted in bold.

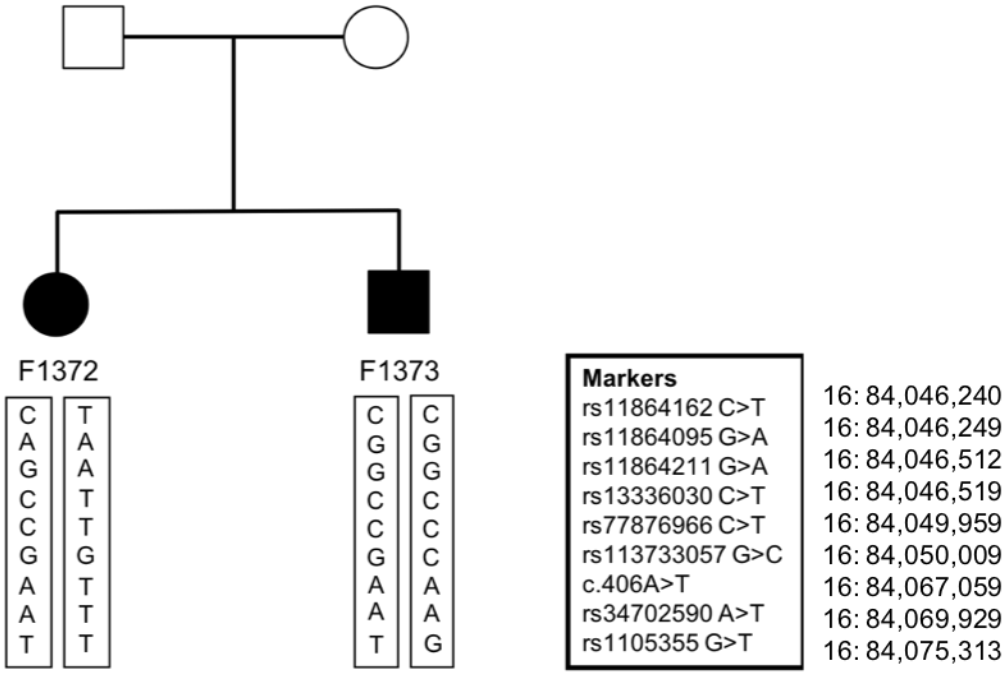


Figure 3-8 Haplotype analysis of SNPs within *SLC38A8* in affected sib pair F1372 and F1373.

Genotyping results are shown for nine SNPs. Haplotypes are presented for both siblings. No parental DNA was available.

During Sanger sequencing of patient F1373, a heterozygous variant in exon 8 of *SLC38A8* (c.1027G>C; p.(Gly343Arg)), was found. The variant is present in the publicly available databases (rs113733057) with a reported allele frequency of 364/276174 (0.003491) in gnomAD (last accessed 03/08/2017). Four out of eight pathogenicity prediction tools predicted this variant to be pathogenic, but the weighted FATHMM score was noted as Tolerable, both MutPred and PROVEAN scored the variant as Neutral and SIFT scored the variant as Tolerated (Table 3-9). The variant has a CADD score of 23.4 (Table 3-9). The variant was not present in his affected sib F1372 and was therefore not investigated further.

Position on Chromosome 16	cDNA change	Protein change	PolyPhen-2	MutationTaster2	Blosum62	SIFT	MutPred	PROVEAN	CADD Score	FATHMM (weighted score)
84,050,259	c.1027C>G	p.(Gly383Arg)	Probably damaging (0.996)	Disease causing (probability 0.98)	-2	Tolerated (0.734)	Neutral (score 0.3)	Neutral (0.883)	23.4	Tolerated (4.44)

Table 3-9 Bioinformatics analysis of the rare variant identified in affected sib F1373.

The heterozygous missense variant, c.1027C>G was identified in patient F1373. The variant is annotated using the reference sequence NM_001080442.1 and human reference genome build GRCh37. URLs: PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/ (Adzhubei *et al.*, 2010); MutationTaster2, http://www.mutationtaster.org/ (Schwarz *et al.*, 2010; Schwarz *et al.*, 2014); SIFT, http://sift.jcvi.org/ (Ng and Henikoff, 2003); Blosum62 (Henikoff and Henikoff, 1993); PROVEAN, http://provean.jcvi.org/ (Choi *et al.*, 2012); MutPred 1.2, http://mutpred.mutdb.org/ (Li *et al.*, 2009); CADD Score, http://cadd.gs.washington.edu/score (Kircher *et al.*, 2014); FATHMM score (weighted) http://fathmm.biocompute.org.uk/index.html (Gough *et al.*, 2001). Blosum62 scores range from +3 to -3; negative scores are more likely to be pathogenic substitutions. CADD score of ≥20 means that the variant is in the top 1% most deleterious mutations in the genome.

3.3 Discussion

The original report of FHONDA identified the first mutations in *SLC38A8* by screening two consanguineous families linked to the locus (Al-Araimi *et al.*, 2013; Poulter *et al.*, 2013). The authors then screened the gene in a further 12 unrelated individuals with similar phenotypes and identified mutations in 5 of these (42% detection rate) (Poulter *et al.*, 2013). In comparison, in this study, a screen of 13 unrelated individuals only led to the identification of likely causal mutations in 2 cases (F1368 and F1377) (15% detection rate). One of these mutations was homozygous, located in exon 7 of *SLC38A8* (F1368), and the second was heterozygous, located in the start codon of the gene (F1377). Further investigation of the locus did not reveal a second mutation in the latter individual. A possible reason for the lower detection rate in this study is that none of the patients had any hints of linkage to the *SLC38A8* locus, whereas in the earlier study, 3/12 cases were consanguineous and had autozygous regions overlapping *SLC38A8* (although no mutation was identified in one of these samples until this study - see below for patient F1310) (Poulter *et al.*, 2013).

3.3.1 Expanding the mutation spectrum of *SLC38A8*

The homozygous change identified in F1368 was a non-conservative substitution and the heterozygous variant p.(Met1?) identified in F1377 was predicted to delete the first 26AA of the protein. For both variants, the disease-causing nature was supported using a combination of *in silico* pathogenicity prediction softwares, conservation analysis and publicly available allele frequencies in the gnomAD database. GnomAD, an extension of ExAC (Lek *et al.*, 2016), contains 123,136

exomes and 15,496 genomes from unrelated individuals, not solely control individuals, but also patients from various disease-specific and population genetic studies. The database is not reported to contain any patients with severe developmental or paediatric disorders, yet it is evident that patients with visual defects are contained within this database. For example, homozygous frameshift alleles in *LTBP2* (MIM 602091) are present and these cause primary congenital glaucoma with Marfan-like extra-ocular features (Ali *et al.*, 2009).

The small number of affected individuals highlights that FHONDA is extremely rare in the general population, meaning that the causal mutations must be of large effect and under strong negative selection. Allele frequency thresholds for such a rare Mendelian disease are discussed in more detail in Chapter 4, with regards to WES variant prioritisation (Whiffin et al., 2017).

3.3.1.1 The identification of a homozygous *SLC38A8* missense variant in case F1368

In gnomAD, the allele frequency of the homozygous missense variant identified in F1368 was notably higher in the Ashkenazi Jewish population (0.006). The homozygous change, p.(Asp283Ala) in gnomAD was detected in a case of Ashkenazi Jewish ancestry. Further familial information is required for F1368, who was only reported as a Caucasian male from the USA, but no details were provided regarding his ethnicity. The patient was recruited directly by an ophthalmologist, who, in the experience of the Leeds Vision Research Group, rarely enquire about a patient's ethnicity, unlike clinical geneticists.

It is always difficult to determine with absolute certainty whether missense mutations are disease-causing without further supportive data. While this higher allele frequency in the Jewish population could suggest that the p.(Asp283Ala) variant is a polymorphism specific to this population, there have since been two additional reports of this homozygous mutation in unrelated FHONDA patients, suggesting that it may represent a founder mutation. The first is another patient with FHONDA syndrome from the USA (Dr Carmel Toomes, personal communication). Secondly, the mutation has been identified in two affected sibs in a nonconsanguineous Ashkenazi Jewish family (Toral *et al.*, 2017). The affected sibs in the study by Toral and colleagues were diagnosed with foveal hypoplasia but had not been tested for optic nerve misrouting. Toral and colleagues performed structural modelling of the mutation based on hydropathy using the membrane topology prediction method, TMHMM (Krogh *et al.*, 2001). Later modelling of the tertiary structure was performed using a threading approach (I-TASSER), as there was a lack of structural templates in the Protein Data Bank. Subsequently, mutations were introduced into the model using FoldX (Zhang, 2008; Roy, Kucukural and Zhang, 2010; Yang and Zhang, 2015; Schymkowitz *et al.*, 2005). This investigation predicted SLC38A8 to have 11 helices, similar to the consensus of analyses performed in this study using tools such as SMART, Protter and Phobius (Section 2.8.4). Using SMART, the p.(Asp283Ala) variant was predicted to lie on the extracellular side of the membrane. Likewise, Toral and colleagues located the mutation extracellularly, proximal to the channel pore (Toral *et al.*, 2017). Using the linearized adaptive Poisson-Boltzmann solver software (Baker *et al.*, 2001) to analyse the channel's electrostatic potential, the change from a negatively-charged polar aspartic acid residue to an uncharged and hydrophobic alanine residue was reported to potentially affect the local sodium

concentration and disrupt glutamine transport (Toral *et al.*, 2017). The altered aspartate residue is well conserved among the orthologues, but the conservation of the residue within the paralogues, however, is less convincing (Figure 3-3).

In the study by Toral and colleagues, all known *SLC38A8* mutations were mapped on to a hydropathy-based structural model of the protein (Toral *et al.*, 2017). The predicted mechanisms of pathogenicity ranged from failed localisation of the SLC38A8 channel to the membrane (p.(Gln200*) and p.(Ala282del)), to affected membrane stability due to hydrogen bonding aberrations (p.(Glu233Lys)) (Toral *et al.*, 2017; Poulter *et al.*, 2013). What the Toral study failed to note, however, is that the p.(Gln200*) mutant protein is unlikely to be synthesised, as the mRNA is likely to be targeted for nonsense-mediated decay (NMD) as the premature termination codon (PTC) is before the last exon. Interestingly, the Toral study revealed that three of the mutations from the gene identification study, (p.(Ser336Ala), p.(Leu344Cys) and p.(Gly412Arg)), were predicted to be protein-stabilizing, with their pathogenic mechanism unknown (Toral *et al.*, 2017).

Seven out of the eight pathogenicity prediction tools employed predicted the p.(Asp283Ala) change to be pathogenic. It is clear that no one algorithm performs best. A range of pathogenicity prediction tools that employ sequence and evolutionary conservation-based methods (SIFT, PROVEAN, Blosum62), protein sequence and structure-based methods (PolyPhen-2) and supervised learning methods (MutPred and MutationTaster), were therefore employed in this study. This

was to reduce the prevalence of either a false negative or false positive result and to help reach a consensus when outputs conflicted.

The variant identified in F1368 was predicted to have a CADD score >20, ranking the variant's deleteriousness above all 99.5% of all possible human SNVs, more than 97% of missense SNVs in a typical exome and more than 56% of Mendelian pathogenic CNVs in ClinVar, according to the creators of CADD (Kircher *et al.*, 2014). The creators of CADD recommend retaining all scores ≥15 as potential pathogenic variants, as it is the median value for all possible canonical splice site changes and nonsynonymous variants (Kircher *et al.*, 2014). CADD scores, however, should be interpreted in terms of "likelihood of deleteriousness", rather than the likelihood of pathogenicity, as there is no calibrated relationship between the two at present (Kircher *et al.*, 2014).

3.3.1.2 The identification of a heterozygous *SLC38A8* missense variant in case F1377

The heterozygous p.(Met1?) variant identified in patient F1377 had a reported allele frequency of <0.01 in the gnomAD database for exomes and genomes in all populations. In the gnomAD database, two additional variants in the start codon of *SLC38A8* are reported. The first is a rare SNP, c.3G>A, present in dbSNP (rs746806089) (1/30926, MAF $3.2e^{-5}$, gnomAD) that abolishes the start codon. The second variant, c.1A>T, does not have an rs number, but it is reported in the gnomAD database with an allele count of 1/238975, MAF $4.19e^{-6}$.

Six out of the eight pathogenicity prediction tools applied in this study predicted the p.(Met1?) variant to be pathogenic. FATHMM and PROVEAN conflicted the other six tools by predicting p.(Met1?) in case F1377 to have a neutral effect. In a study by Walters-Sen and colleagues, PROVEAN was found to give the best performance for detecting credibly benign variants, with 94.7% correctly predicted (Walters-Sen et al., 2014). Their study analysed all prediction tools applied in this thesis, with the exception of Blosum62, and found MutPred to be the best performer, with an accuracy of 82% in a dataset of approximately 120 credibly pathogenic and credibly benign variants in genes involved in the RASopathy family of disorders and limb-girdle muscular dystrophy (Walters-Sen et al., 2014). In the original *SLC38A8* screen, SIFT scored the most variants as benign, with MutationTaster and MutPred scoring the most variants as pathogenic (Poulter *et al.*, 2013). As described in Chapter 2, SIFT and PROVEAN are based on the same algorithm. Dependent on the gene and the mutation in question, it is clear that pathogenicity prediction tools should not be used alone to determine whether or not a mutation is disease-causing. However, at present there aren't any appropriate functional tests that could have been performed in order to test the variants identified in F1368 and F1377.

The c.2T>C transition identified in patient F1377 is predicted to cause the loss of the primary start codon AUG coding for methionine. Start codon mutations have been reported in human disease, including the detection of a *de novo* mutation (c.3G>A) in the phenylalanine hydroxylase gene in a patient with phenylketonuria (Eiken *et al.*, 1992), a novel mutation (c.3G>A) in the initiation codon causing beta-thalassemia (Saba *et al.*, 1992) and a c.3G>A mutation reported in a Dutch family with Pelizaeus-

Merzbacher disease, an X-linked recessive disorder characterised by demyelination of the CNS (Sistermans *et al.*, 1996).

Although there are many upstream ORFs, in general, the first AUG on the mRNA will become the translation initiation codon (TIC) and in combination with the embedded Kozak consensus sequence (specifically there is strong bias for a purine at position -3 and a G at position +4), will maximise translation efficiency (Kozak, 1986; Kozak, 1991; Kozak, 1997; Nakagawa *et al.*, 2008). There is no 5' UTR sequence currently annotated for *SLC38A8* (UCSC Genome Browser, last accessed 03/08/2017). Studies of the *SLC38A8* 5' UTR are undertaken in Chapter 5, whereby RT-PCR of the region upstream of the WT AUG codon revealed a short sequence of what is strongly believed to be 5' UTR.

Reports of changes to both the TIC and the consensus sequence have been shown to be detrimental to the initiation step of translation *in vivo*, with the potential to activate an alternative AUG upstream or downstream (Kozak, 2002; Kochetov *et al.*, 2005; Wethmar, Smink and Leutz, 2010). Such mutations are well reported as causes of human disease, for example a 5' UTR mutation at base -34 in *CDKN2A* giving rise to a novel AUG initiation codon that leads to decreased translation in a number of patients with melanoma (Liu *et al.*, 1999). Additionally, a pathogenic initiation codon mutation in *U2HR*, the inhibitory upstream ORF in the 5' UTR of the gene encoding the human hairless homolog HR, leads to cases of Marie Unna hereditary hypotrichosis (Wen *et al.*, 2009). Such mutations are also seen in

ophthalmic disease, including start codon mutations in *FRMD7* causing idiopathic nystagmus (Choi *et al.*, 2015).

Alternative translation start sites can generate more than one protein isoform from the same gene (Kochetov, 2008), although it has been suggested that this can reduce protein expression by 30-80% (Calvo, Pagliarini and Mootha, 2009). Adhering to the "scanning model" of protein translation initiation, the direction of this model suggests that a TIC mutation might preferentially activate the nearest downstream in-frame ATG codon, which was located 26aa away in SLC38A8 (ΔSLC38A8(26aa)). As an aside, it should be noted, however, that not all genes have an AUG start codon (Kearse and Wilusz, 2017). To determine if the downstream in-frame start codon is used and such a "leaky scanning" mechanism is taking place, functional studies need to be performed. In original experiments exploring this mechanism, the first AUG codon was placed out of frame with respect to a bacterial chloramphenicol acetyltransferase (CAT) reporter gene sequence (Kozak, 1995). Using this method, it is possible to test whether ribosomes can initiate translation at a downstream AUG codon, with CAT protein presence/absence dependent on whether this is possible (Kozak, 1995). More recent studies have employed *in vitro* GFP-reporter assays, for example *MLH1*-GFP fusion constructs generated by fusing 5' sequence preceding the downstream ATG initiation codon to GFP, to assess the consequences of translation initiation disruption (Parsons *et al.*, 2015). In this study, due to the limited expression of SLC38A8 (Section 5.2.4.1), patient protein lysates and RT-PCR experiments would not be possible.

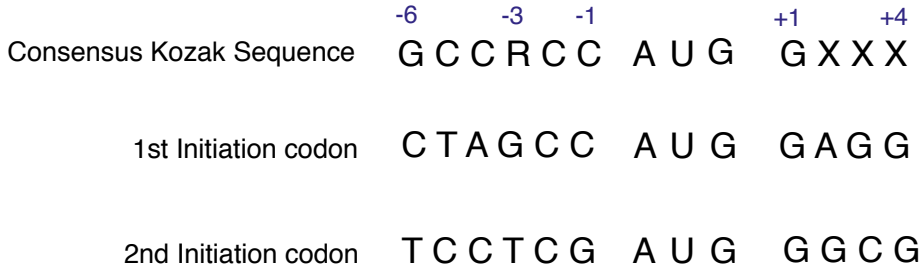


Figure 3-9 Sequence of the Kozak consensus and the translation initiation start site of SLC38A8.

The nucleotide sequence surrounding the first (WT) and second AUG codons. The Kozak consensus sequence is shown at the top.

As the sequence flanking the TIC has also been suggested to influence translation efficiency (Kozak, 1986; Kozak, 1997), the similarity between the Kozak consensus sequence and the sequence surrounding ΔSLC38A8(26aa) (positions -6 to +4) was assessed. The results for the shift start site protein ΔSLC38A8(26aa), compared to the WT SLC38A8 (Figure 3-9). It is interesting to note that the Kozak consensus sequence for WT SLC38A8 is not a conventional consensus sequence (5'-GCCRCCAUGG-3', whereby R represents a purine A/G) (Figure 3-9).

Three similarities were found between the SLC38A8 WT Kozak consensus site and the sequence surrounding ΔSLC38A8(26aa), including the C at position -2, the G at position +1 and the G at position +4. The latter position is evolutionarily conserved in vertebrates and mutations at this position have been shown to exert a negative effect on translation (Kozak, 2002; Nakagawa *et al.*, 2008). Studies of translation initiation in eukaryotes have utilised fluorescence-activated cell sorting and high-throughput DNA sequencing (FACS-seq) to determine the efficiency of start codon recognition for all possible translation initiation sites using AUG start codons (Noderer *et al.*,

2014; Diaz de Arce, Noderer and Wang, 2018). From a genetic reporter library representing all 65,536 possible translation initiation sequences spanning the -6 to +5 positions, the motif RYMRMVAUGGC was found to enhance ribosome recognition and translation efficiency (Noderer *et al.*, 2014). In line with previous findings (Kozak, 1986; Kozak, 1987), positions -3 and +4 were identified as important sites for TIC efficiency (Noderer *et al.*, 2014). The fact that the -3 purine isn't present at the potential shifted start site suggests that even if this downstream AUG was utilised, the mRNA would be translated at much lower efficiency. This further supports the pathogenicity of the variant. Research performed by Wolf and colleagues suggested that single base-pair substitutions reported in TICs outnumber those identified in termination codons by >3 fold (Wolf *et al.*, 2011). There is therefore an important role of TIC-mutations in gene pathology, as approximately 0.7% of all disease-causing amino acid changes occur in the start codon (Wolf *et al.*, 2011). Further to this, studies interrogating Ensembl data have revealed that 11,261 variants affect the initiation codons of 7,205 genes, with 99.5% of these having low or unknown MAF, thought to reflect their pathogenic nature (Abad-Navarro *et al.*, 2018).

Further to a shifted AUG TIC, research has been executed to determine the efficiency of translation by non-AUG codons and this possibility cannot be ignored in the case of F1377. In other studies, all nine non-AUG codons that differ by a single base pair, in combination with every possible TIC sequence from the -4 to +4 position, have been investigated using FACS-seq (Diaz de Arce, Noderer and Wang, 2018). Results show that CUG codons are the most effective of all non-AUG codons, causing approximately 50% as much protein expression as the consensus Kozak

sequence (Diaz de Arce, Noderer and Wang, 2018). Investigation of the *SLC38A8* sequence revealed that the nearest CUG codon is located 23aa away from the WT TIC.

It is highly likely that the second mutation has been missed in case F1377, given the autosomal recessive inheritance pattern and the fact that other mutations are known to be missing in this disease (case F1310, (Sergouniotis, 2012)). A less convincing hypothesis is that the heterozygous variant identified in case F1377 could have a dominant-negative character. If this was the case, the mutation would be *de novo*, as the parents are unaffected. Similar dominant-negative mechanisms have been observed in IRDs, for example, mutations in *NRL* (Kanda *et al.*, 2007) and *CRX* (Hull *et al.*, 2014; Ibrahim *et al.*, 2018), both reported in autosomal dominant disease, but now also reported in autosomal recessive cases. The most likely explanation in case F1377 is that a second mutation has been missed.

Bringing together the results of this chapter with the original findings and subsequent reports, the total number of likely pathogenic mutations identified in *SLC38A8* is eleven (Poulter et al., 2013, Perez et al., 2014, Toral et al., 2017a) (Figure 3-10).

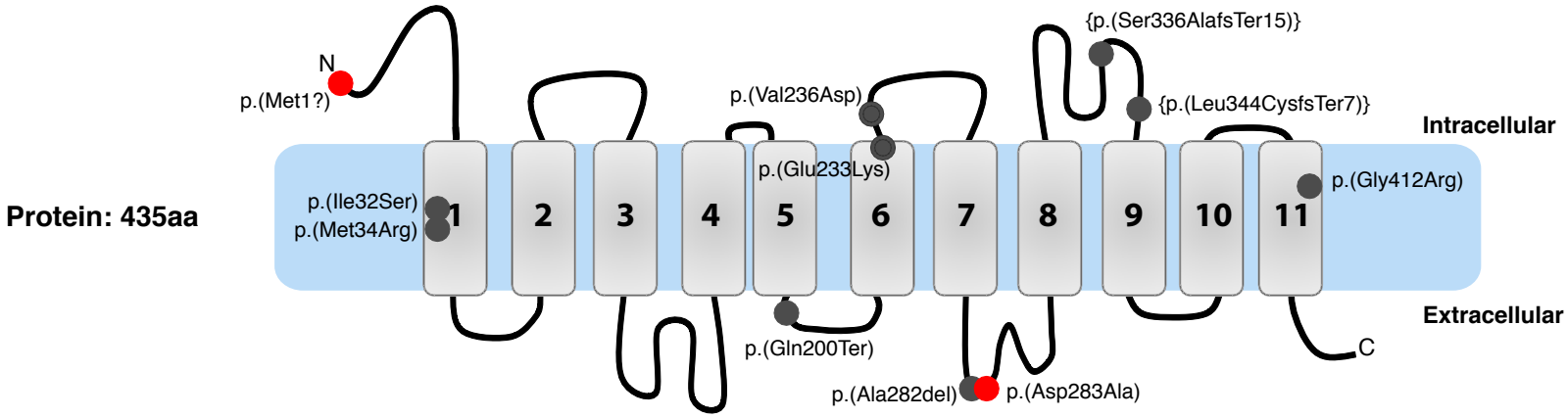


Figure 3-10 Mutation spectrum on the SLC38A8 protein.

A schematic protein structure of SLC38A8 showing the location of novel variants (red) and variants from previous publications (grey) (Poulter *et al.*, 2013; Perez *et al.*, 2014; Toral *et al.*, 2017). The compound heterozygous variants are in parentheses. The structure was generated using the topology prediction tools SMART (Section 2.8.4.1), Protter (Section 2.8.4.2) and Phobius (Section 2.8.4.5).

3.3.2 "Undetectable" mutations in *SLC38A8*

The initial screening strategy applied in this study focused on sequencing the coding regions and splice sites (10 bp of flanking intronic sequence) of *SLC38A8*. This screening method is predicted to detect the majority of mutations, although it will not detect all. The exome accounts for ~2% of the total human genome and 85% of published mutations are harboured in these regions (Majewski *et al.*, 2011). Reporting of mutations in coding regions is likely to be heavily biased, as the majority of researchers, until very recently, only screened these areas of genes. As discussed in the above Section 1.3.1, it is the identification of a single heterozygous mutation in case F1377 and the absence of a mutation in case F1310 (Poulter *et al.*, 2013; Sergouniotis, 2012), that indicates that mutations have been undetected in this recessive disease.

Common locations for mutations undetectable by conventional methods are the promoters, enhancers, 5' or 3' non-coding regions and also intronic fragments. For example, mutations or SNPs in promoters and 5' UTR of genes may affect the binding of *trans*-acting factors, with subsequent consequences for transcription efficiency (Coppieters *et al.*, 2015). Also, 3' UTR SNPs may cause aberrant mRNA localisation and polyadenylation status (Scheper, van der Knaap and Proud, 2007). Epigenetic regulation of the mammalian transcriptome is evident through N^6-methyladenosine (m^6A) methylation and methylation consensus sites are enriched near the stop codons of genes, the 3' and 5' UTR and within long exons (Dominissini *et al.*, 2012; Meyer *et al.*, 2012). mRNA homeostasis, as regulated by m^6A, has been linked to a number of diseases, for example obesity and cancer (Lin *et al.*, 2016; Li *et al.*, 2017).

Additionally, “undetectable” mutations can be located deep in the introns of genes, as is the case for many IRDs. Deep intronic variants can lead to forms of alternative splicing, including exon skipping, alternative splice-site activation, intron retention or mutually exclusive splicing. Examples of non-coding mutations are prevalent in IRDs, including *CEP290* (intronic splicing mutation, novel exon inclusion) (den Hollander *et al.*, 2006), *LCA5* (deletion of promoter) (den Hollander *et al.*, 2007a), *ABCA4* (deep intronic founder variant) (Bauwens *et al.*, 2015), *ABCA4* (intronic splicing mutations) (Zernant, 2014), *NMNAT1* (UTR mutations) (Coppieters *et al.*, 2013; Coppieters *et al.*, 2015) and *ATOH7* (remote enhancer deletion) (Ghiasvand *et al.*, 2011). A search for retina-specific *SLC38A8* transcripts (detailed more in Section 5.2.4.1), as with the alternatively spliced isoform of *RPGR*, *ORF15*, which is solely expressed in the retina (Kirschner *et al.*, 1999), failed to find any.

The ENCODE project has estimated that approximately 80% of the human genome contains elements that play essential biochemical roles in genome functionality (Birney *et al.*, 2012). These functional elements are not solely located in exons, but also specific regional signatures that are enriched for histone modifications or DNaseI hypersensitive sites (DHSs), for example (Birney *et al.*, 2012). PCR-based Sanger sequencing, as employed here, with sole analysis of individual exons will miss mutations present in these regions of gene regulation. In addition, gene rearrangements, such as copy-number variations (CNVs), will be also missed by such screening methods. A large deletion spanning *SLC38A8*, and neighbouring gene *DNAAF1*, has previously been identified in a non-consanguineous FHONDA patient (F1290) with Kartagener’s disease (OMIM 244400) (Poulter *et al.*, 2013; van Genderen *et al.*, 2006) and a large 640kb deletion encompassing both genes has

also been reported in a second patient (Loges *et al.*, 2009). The microsatellite haplotype analysis in this study did not reveal the breakpoints of the deletion CNV in F1290, a type of mutation that would be undetected using the current Sanger sequencing screening strategy in the absence of parental DNA. There are a number of alternative methods to identify structural variation (SV) in the form of CNVs, such as deletions or duplications. Array comparative genomic hybridisation (aCGH) has been used extensively for the detection of CNVs in coding and non-coding regions of IRD genes, with the development of customised arrays, for example arrEYE, which contains probes for 106 known IRD genes, 60 candidate genes and 196 retina-expressed noncoding RNAs (Van Cauwenbergh *et al.*, 2016). The large number of heterozygous SNPs within *SLC38A8* identified in the mutation negative cases (Table 3-8) suggests that they do not harbour large deletions encompassing this gene.

3.3.3 Genetic heterogeneity in FHONDA

Haplotype analysis using microsatellite markers and SNPs revealed no shared haplotypes consistent with recessive inheritance in the consanguineous affected sibs F1288 and F1287 and the non-consanguineous sib pair F1372 and F1373 across *SLC38A8*. This suggests that their phenotype is not caused by mutations in *SLC38A8*. These familial cases along with the high number of sporadic *SLC38A8* mutation-negative patients identified in this study indicates that FHONDA is a genetically heterogeneous disease and that further disease genes remain to be identified.

3.4 Summary

In this study two novel mutations have been identified in two separate cases of FHONDA. Failure to identify the second mutation in case F1377 raises the likely possibility that mutations remain undetected using this screening method and there is a need to apply NGS to identify potential hidden mutations or structural variation. What is more, the high number of sporadic *SLC38A8* mutation-negative cases identified in both this study and the original study (Poulter *et al.*, 2013), and sib pairs not sharing haplotypes across the gene, hints that FHONDA is not a single gene disorder, but genetically heterogeneous, warranting novel gene investigation by WES.

Chapter 4 WES of foveal hypoplasia patients to investigate novel genes

4.1 Introduction

Screening of the gene *SLC38A8* in individuals with foveal hypoplasia and no hypopigmentation defects resulted in the identification of a cohort of 18 *SLC38A8* mutation-negative unrelated cases (11 from the work presented in Chapter 3: F1351, F1352, F1335, F1369, F1371, F1372, F1374, F1378, F1402, F1403 and F1404 and 7 from the initial gene identification study: F1310, F1071, F1287, F1298, F1405, F1419 and F1421 (Poulter *et al.*, 2013)). Among this cohort of 18 were two sibling pairs (F1372 and F1373 and F1287 and F1288) who did not appear to share haplotypes across *SLC38A8*, as would be consistent with recessive inheritance (Section 3.2.2). These results suggested that the rare foveal hypoplasia without albinism phenotype is heterogeneous and that further mutated genes remain to be identified. The aim of the work presented in this chapter was to identify novel gene(s) implicated in this unique disease phenotype using WES.

WES is now widely used to detect new disease genes due to its parallel and unbiased genome-wide search for causal mutations in coding sequence, allowing for a high number of loci to be screened at the same time (Ng *et al.*, 2009). Therefore, this approach was used on 10 unrelated members of the *SLC38A8*-negative cohort and two siblings (F1372 and F1373). Although the data presented in this thesis for

each individual is based on the entire dataset, the findings were not generated at the same time, but over the 3-year period. Cases were selected for WES based on the following selection criteria:

(a) Negative for pathogenic mutations in the coding region and splice sites of *SLC38A8*, excluded using Sanger sequencing;
(b) Sufficient quantity and quality of DNA available;
(c) Patient consent provided.

4.2 Results

4.2.1 Whole exome sequencing strategy in the search for rare alleles

WES was undertaken on genomic DNA from 12 individuals: F1335, F1351, F1352, F1369, F1371, F1374, F1288, F1071, F1298, F1405 and sib pair F1372 and F1373. All patients had a minimum diagnosis of foveal hypoplasia without pigmentation defects (full details in Tables 3.1 and 3.6) and will hereby be referred to as the foveal hypoplasia cohort. WES libraries were prepared by either the author or the Leeds Next Generation Sequencing Facility (http://dna.leeds.ac.uk/genomics/) using the SureSelectXT kit, as described in Section 2.3.25.1.

Patients F1351, F1369, F1374 and the sib pair F1372 and F1373 were analysed in the first batch of exomes (alongside additional patients not studied here), at which time the Agilent SureSelect capture reagent applied was V5. The capture reagent applied for all remaining exomes was V6. A representative TapeStation trace obtained for individual F1351 is shown in Figure 4-1, from the final assessment stage

of the SureSelectXT library preparation, prior to pooling (Section 2.3.22.1). All libraries were pooled in batches of 10 and subjected to either 100 or 150bp paired-end sequencing on a single lane of a HiSeq2500 or a HiSeq3000 (Section 2.3.25.1).

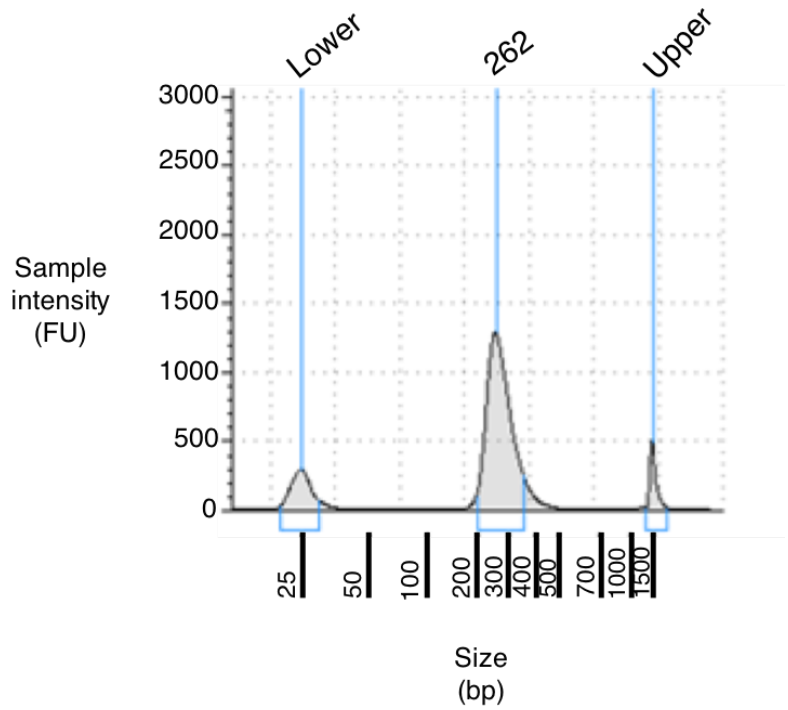


Figure 4-1 Trace from the 2200 TapeStation analysis of the final WES library for patient F1351.

A representative trace from the TapeStation recording taken at the end stage of the SureSelectXT protocol, prior to pooling. The expected fragment size peak is between 225bp-275bp. The upper and lower peaks are the size standards. Abbreviations; bp base-pairs, FU fluorescence units.

The bioinformatics analysis of the WES data was performed as described in Section 2.3.22.2, Figures 2.2 and 2.3. The sequencing data was obtained as raw FastQ files that were initially inspected for quality using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). All the samples passed the QC and an example of the output for patient F1405 is shown in Figure 4-2. The depth of coverage was also analysed (Section 2.3.25.6) and the mean depth of

coverage for each sample is listed in Table 4-1. All samples had a minimum of 88% of bases covered by 20 reads.

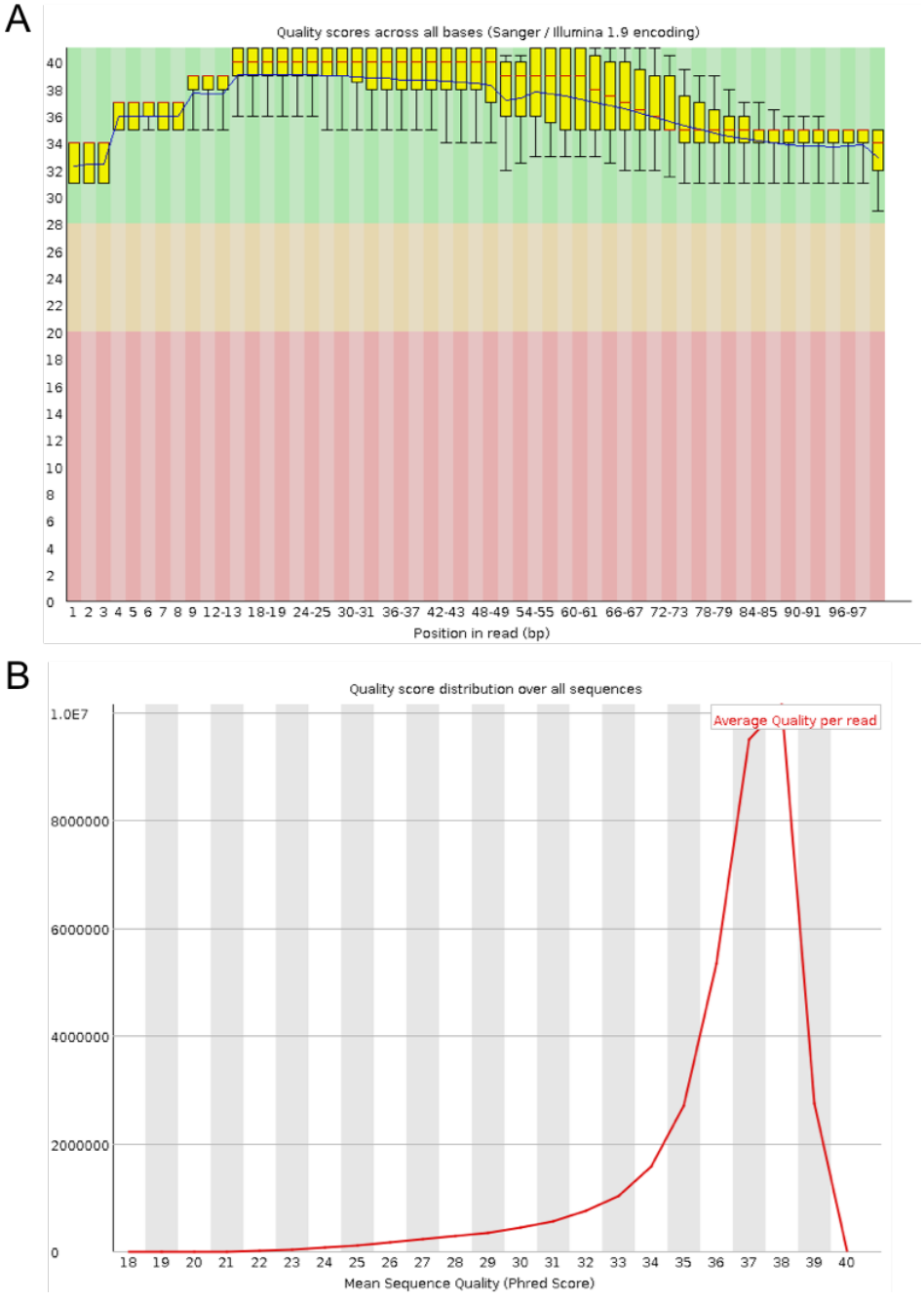


Figure 4-2 Quality control of raw WES data using FastQC for patient F1405.

(A) Box plot of the per-base sequence quality (Phred) scores across all bases at each position. The yellow boxes represent the interquartile range and the whiskers (upper and lower) show the 10% and 90% range respectively. The y-axis represents the Phred quality score, with >20 indicative of 99% base call accuracy and a score >30 indicative of 99.9% accuracy. The red line shows the median and the blue line shows the mean quality score. (B) Quality score distribution over all sequences. The y-axis represents the number of sequence reads. Highlighted in bold font are the sibling pair, F1372 and F1373.

Sample	Mean read depth	% bases covered by >10 reads	% bases covered by >20 reads	% bases covered by >30 reads
F1071	123.23	98.9	97	94.1
F1288	72.05	95.2	88.1	79.2
F1298	114.57	98.3	95.7	92.2
F1335	86.89	99.5	98.8	97.1
F1351	68.11	96.6	90	80.4
F1352	112.77	98.7	96.5	93.2
F1369	105.99	96.4	92.1	87
F1371	97.26	99.2	98	95.6
F1372	72.47	97	91.6	83.5
F1373	82.16	97.8	93.6	87
F1374	73.12	96.1	89.6	80.8
F1405	69.32	98.8	95.3	88.5

Table 4-1 Mean read depth and % of bases covered by >10, 20 and 30 reads for all cases.

All read depth statistics were generated using the GATK 'DepthofCoverage' command.

Variants in the known IRD and albinism genes (taken from RetNet and last accessed 03/08/2017, Appendix 6) were first inspected in the foveal hypoplasia cohort. In this initial analysis, variants with a MAF 0.01 in dbSNP142, ExAC or two locally sequenced exomes (without eye disease) were removed. The remaining variants were annotated using variant effect predictor. Synonymous variants and variants outside exons and their splice site regions were filtered out. The remaining variants were prioritised according to CADD score. CADD scores above a threshold of 15 were considered in the final variants lists. For each variant listed, the prediction tools

PolyPhen-2, SIFT and CADD were used to annotate potential pathogenicity of candidates. Variant lists were created this way for the sib pair, the cohort combined and each patient on a case-by-case basis. For each variant described, its presence in public databases dbSNP, ExAC and gnomAD is shown. It should be noted that the gnomAD database was not released until a later stage of the project, meaning that candidate variant allele frequency and reported loss of function (LoF) variants were only initially investigated using the dbSNP and ExAC databases. The release of gnomAD had an impact on variants that were prioritised in the analysis, to be discussed further in Section 4.2.3. In the sections below, the results of WES of the 12 foveal hypoplasia patients are detailed.

4.2.2 Screening patients for known IRD or albinism genes

As detailed above, the first stage of the WES analysis pipeline for each individual focused on the identification of rare and functional variants in known retinal and albinism disease genes (listed in Appendix 6). The first samples to undergo WES were F1351, F1369, F1374 and the sib pair F1372 and F1373. During this analysis, a heterozygous mutation was identified in exon 6 of *PAX6* in case F1351, NM_000280; c.275T>C; p.(Val78Ala) number of reads WT(wild-type)/Mut(mutant): 22/30 (Table 4-2). Prior to enrolment in this study, this patient had been pre-screened for albinism genes. However, this male patient showed no signs of pigmentation defects. He had only undergone albinism molecular testing as it is often routinely performed in foveal hypoplasia patients in Scandinavia and the Netherlands, as these populations are naturally very pale, making albinism hard to detect.

The SMART and Protter tools (Section 2.4.8.1) were used to predict the domains of PAX6 and were used to generate the consensus 2D schematic shown in Figure 4-3. The encoded PAX6 protein contains two DNA-binding domains (DBD), the paired box of approximately 124 aa and a paired-type homeobox of approximately 62 aa, separated by a linker region (Figure 4-3). The variant identified occurs at a highly conserved amino acid in the *PAX6* paired domain (Figure 4-4). Conservation analysis was undertaken using HomoloGene (Section 2.4.7). The results show the valine residue to be fully conserved through evolution down to the zebrafish, indicating the importance of the valine at this position in the paired domain (Figure 4-4). The variant was located in the coding region of all 11 *PAX6* RefSeq transcripts (Figure 4-5). Investigation of the p.(Val78Ala) variant frequency in publicly available databases (Section 2.4.2), including dbSNP and gnomAD, revealed that the variant is not present (last accessed on 03/08/2017). It has been reported once in the *PAX6* LOVD database (http://lsdb.hgu.mrc.ac.uk/home.php?select_db=PAX6) in a patient with foveal hypoplasia and nystagmus but remains unpublished in the literature. The *PAX6* variant was verified by Sanger sequencing using an independent patient DNA sample (Section 2.2). DNA was not available from family members and so segregation analysis could not be performed. Analysis of this substitution using a variety of different pathogenicity prediction tools (Section 2.4.4), suggested that the missense change is likely to be disease-causing. Five of the six pathogenicity prediction tools applied in this analysis suggested the variant was pathogenic, including CADD (Score of 25.7, v 1.3). Only Blosum62 predicted a neutral effect with a score of 0 (Table 4-2).

To investigate if additional *PAX6* mutations existed in the patient cohort, all ten constitutive coding exons (exons 4-13), the alternatively spliced exon 5a and flanking introns were Sanger sequenced in all *SLC38A8*-negative patients (12 cases) (primers listed in Appendix 7). This sequencing included all patients not already analysed by WES and the remaining cohort not selected for WES. This led to the identification of a second *PAX6* heterozygous missense variant in another individual. The variant, NM_000280; c.278C>A, p.(Ala79Glu), was identified in exon 6 of patient F1378. The female had a phenotype of foveal hypoplasia and no pigmentation defects. Analysis of this substitution using eight different pathogenicity prediction tools, predicted it to be disease-causing (Table 4-2). The alanine residue, located in the paired domain, was conserved through evolution down to the zebrafish (Figure 4-4). The affected amino acid residue is located next to the p.(Val78Glu) variant that was identified in patient F1351.

Interrogation of other *PAX6* transcripts revealed that it is present in all isoforms, as for the previous variant identified in F1351 (Figure 4-5). The A79E variant was absent from publicly available databases dbSNP and gnomAD (last checked 18/09/2017). However, it has been published by a team in Denmark (Grønskov *et al.*, 1999). The phenotype of this published case was aniridia (stromal hypoplasia with eccentric pupil), mild nystagmus and early cataract.

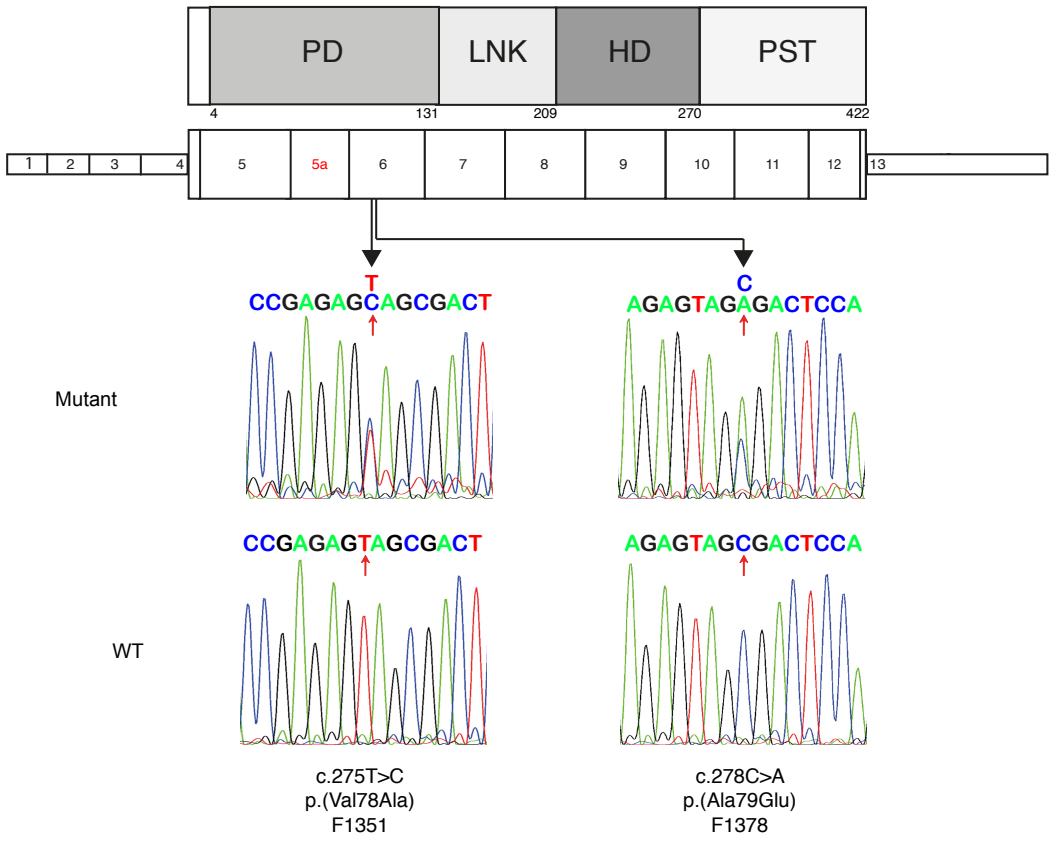


Figure 4-3 Sanger sequencing electropherograms of the two *PAX6* variants identified in patients F1351 and F1378.

Schematic representation of the human *PAX6* gene domains using the SMART and Protter protein tools (Section 2.4.10.1 and 2.4.10.2). The PAX6 protein consists of a Paired Domain (PD) which includes an alternatively spliced exon coloured red (exon 5a) and the Homeodomain (HD) separated from the Proline/Serine/Threonine-rich (PST) transactivation domain by a Linker region (LNK). The first three exons and the majority of exon 4 constitute the 5' UTR. The coding sequence starts in exon 4. The 3' UTR totals approximately 1Kb. Amino acid positions for the PAX6 domains are shown. Electropherogram traces of the two heterozygous variants are mapped at the approximate location on chromosome 11 (NM_000280): c.275T>C, p.(Val78Ala) (Patient F1351) and c.278C>A, p.(Ala79Glu) (Patient F1378). Variants mapped using the Human Reference Genome Build GRCh37/hg19.

Chromosome position	cDNA change	Protein change	PolyPhen-2	MutationTaster2	Blosum 62	SIFT	MutPred	PROVEAN	CADD Score	FATHMM (weighted score)
11:31,823,232	c.275T>C	p.(Val78Ala)	Probably damaging:1.000	Disease causing (probability 1.0)	Score 0	Damaging	Deleterious mutation (probability 0.918)	Deleterious (score -3.368)	25.7	Damaging (-6.27)
11:31,823,229	c.278C>A	p.(Ala79Glu)	Probably damaging:1.000	Disease causing (probability 0.9999)	Score -1	Damaging	Deleterious mutation (probability 0.914)	Deleterious (score -4.260)	28.9	Damaging (-5.81)

Table 4-2 Bioinformatics analysis of novel *PAX6* variants identified in F1351 and F1378.

The heterozygous missense variants, c.275T>C (Patient F1351) and c.278C>A (F1378) were annotated using the reference sequence NM_000280 and Human Reference Genome Build GRCh37/hg19. URLs: PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/ (Adzhubei *et al.*, 2010); MutationTaster2, http://www.mutationtaster.org/ (Schwarz *et al.*, 2010; Schwarz *et al.*, 2014); SIFT, http://sift.jcvi.org/ (Ng and Henikoff, 2003); Blosum62 (Henikoff and Henikoff, 1993); PROVEAN, http://provean.jcvi.org/ (Choi *et al.*, 2012); MutPred 1.2, http://mutpred.mutdb.org/ (Li *et al.*, 2009); CADD Score, http://cadd.gs.washington.edu/score (Kircher *et al.*, 2014); FATHMM score (weighted) http://fathmm.biocompute.org.uk/index.html (Gough *et al.*, 2001). Blosum62 scores range from +3 to -3; negative scores are more likely to be pathogenic substitutions. CADD score of ≥20 means that the variant is in the top 1% most deleterious mutations in the genome.

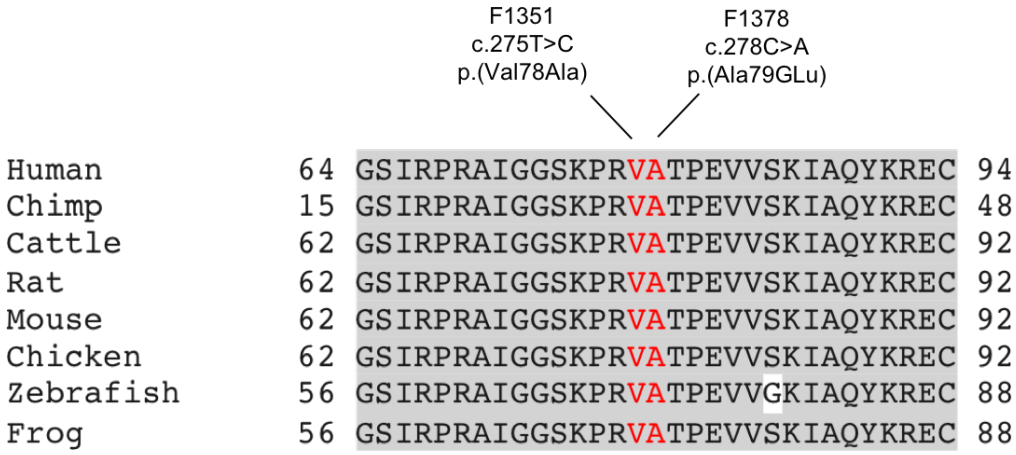


Figure 4-4 HomoloGene multiple sequence alignment of homologous protein sequences for the residues surrounding the two PAX6 variants identified in F1351 and F1378.

The shaded amino acid residues indicate those that are shared with the human residue at the same position. The residues substituted in the two patients are coloured red. Patient F1351 c.275T>C p.(Val78Ala); Patient F1378 c.278T>C p.(Ala79Glu). Sequences of PAX6 used: Human (*Homo sapiens*) NP_000271; Chimpanzee (*Pan troglodytes*) XP_003954413; Cattle (*Bos taurus*) NP_001035735; Rat (*Rattus norvegicus*) NP_037133; Mouse (*Mus musculus*) NP_001231127; Chicken (*Gallus gallus*) NP_990397; Zebrafish (*Danio rerio*) NP_571379; Frog (*Xenopus tropicalis*) NP_001006763.

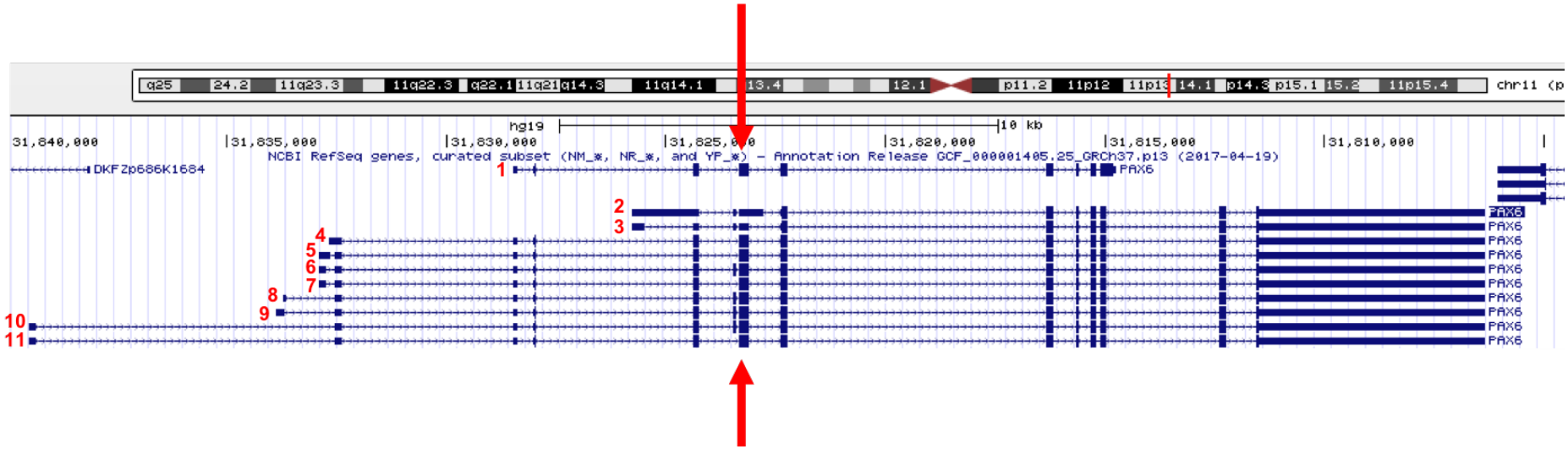


Figure 4-5 UCSC Genome Browser RefSeq PAX6 transcripts 1-11.

UCSC Genome Browser screenshot of all RefSeq *PAX6* transcripts; **1** NM_001310159.1 (size 1393bp 11:31,814,753-31,828,473), **2** NM_001310160.1 (size 8177bp 11:31,806,340-31,825,785), **3** NM_001310161.1 (size 6729bp 11:31,806,340-31,825,785), **4** NM_001258465.1 (size 6854bp 11:31,806,340-31,832,690), **5** NM_000280.4 (size 6966bp 11:31806340-31832901), **6** NM_001604.5 (size 6910bp 11:31,806,340-31,832,901), **7** NM_001258464.1 (size 6868bp 11:31,806,340-31,832,901), **8** NM_001258463.1 (size 6860bp 11:31,806,340-31,833,731), **9** NM_001310158.1 (size 6963bp 11:31,806,340-31,833,890), **10** NM_001258462.1 (size 6922bp 11:31,806,340-31,839,509), **11** NM_001127612.1 (size 6880bp 11:31,806,340-31,839,509). The location of the c.275T>C, p.(Val78Ala) variant (case F1351) and c.278C>A, p.(Ala79Glu) is highlighted by the red arrows. The variant is located in exon 6 of transcript NM_000280 and is present in the coding region of all 11 transcripts.

Family DNA was available for three half uncles and the mother of case F1378. The PAX6 variant was not found in any other family members available, other than the affected proband (Figure 4-6).

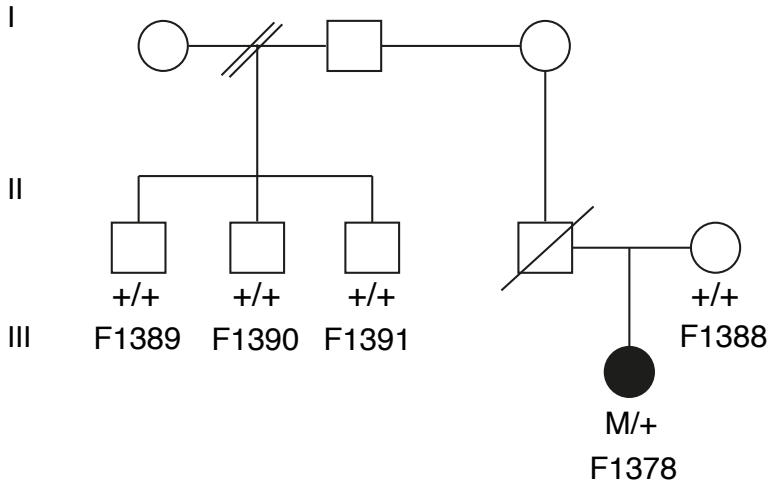


Figure 4-6 Pedigree and mutation status of patient F1378 and family.

Affected individual F1378 is indicated by the shaded symbol. The mutant (M) or wild-type (+) alleles are shown for each individual screened. The father of F1378 was deceased at the time of the study but was not reported to have any visual problems.

No further variants were identified in the coding regions and splice sites of *PAX6* in the entire remaining *SLC38A8*-negative cohort. Further to these findings, the full *SLC38A8* mutation-negative cohort (16 patients) was additionally screened for non-coding variation in the form of *PAX6 cis*-regulatory elements, as these have previously been linked to eye disease (Bhatia *et al.*, 2013). Fifteen eye-related *cis*-regulatory elements were screened by Sanger sequencing using published primers (Bhatia et al., 2013) (Appendix 7). Variants were identified but were not thought to be disease-causing as their allele frequencies were not low enough for a rare dominant disease (Appendix 8). As a result, no further interrogation of these variants was undertaken.

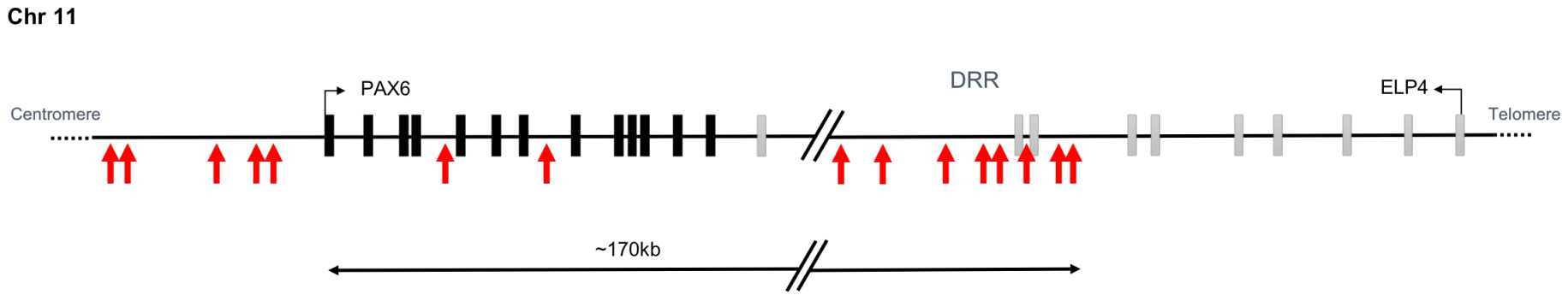


Figure 4-7 Map of the PAX6 locus on human chromosome 11 and location of the 15 cis-regulatory elements screened.

The exons of *PAX6* (black rectangles) and adjacent gene *ELP4* (grey exons) are displayed along with the long-range *PAX6* *cis*-regulatory elements (red arrows). The distal regulatory region (DRR) is shown. There is approximately 170kb between the first exon of *PAX6* and the downstream *cis*-element. Image adapted with permission from Copyright Clearance Centre on behalf of Elsevier: The American Journal of Human Genetics (Bhatia *et al.*, 2013).

Returning to the WES data, further investigation of the known retinal disease genes revealed the following rare variants.

A heterozygous nonsense variant; NM_031935, c.5539C>T; p.(Gln1847Ter), number of reads WT/Mut: 30/25, was found in exon 35/107 of the gene *Hemicentin-1* (*HMCN1)* (MIM 608548) in case F1374. The variant in *HMCN1* was completely unreported in the publicly available databases dbSNP, ExAC and gnomAD (last accessed 12/12/2017). In the gnomAD database overall there are 210 LoF variants in *HMCN1* listed that are common throughout the gene and are predominantly located in the second half of the gene (exons 50-107). The use of CADD predicted the variant to have a score of 38 (Table 4-3). HomoloGene (Section 2.4.7) showed the glutamine residue at position 1847 to be fully conserved through evolution down to zebrafish (*Danio*) (Figure 4-8). Mutations in *HMCN1* have been reported in patients with autosomal dominant macular degeneration (MIM 603075) (Klein *et al.*, 1998; Schultz *et al.*, 2003; Pras *et al.*, 2015). However, *HMCN1's* close proximity to *Complement Factor H* (*CFH)* has questioned whether *HMCN1* is a real AMD gene. No further variants in *HMCN1* (coding region variants or CNVs) were identified in the remaining cohort.

In case F1352, an East Asian individual, a heterozygous missense variant; c.146A>C; p.(Glu49Ala), number of reads WT/Mut: 96/91, was identified in the gene *Fascin Actin-Bundling Protein 2, retinal* (*FSCN2)* (MIM 607643) (NM_012418.3). Interrogation of the variant using publicly available databases revealed the variant was present in dbSNP (rs200600577), ExAC and gnomAD with a reported allele

frequency of 10/220092 (1.544e-5) in gnomAD (last accessed 12/12/2017). The allele was reported in heterozygous state in only East Asian individuals in gnomAD. PolyPhen-2 and SIFT predicted the missense change to be pathogenic, and CADD scored the variant at 24.3 (Table 4-3). Conservation analysis showed the glutamic acid residue to be fully conserved down to the zebrafish (Figure 4-8). The variant identified is similar to mutations which cause AR IRD, making it likely that a carrier has been identified here. Analysis of the remaining cohort did not reveal any further coding region variants or CNVs in *FSCN2*.

In case F1371, rare heterozygous variants in two known retinal disease genes were identified. A heterozygous missense change was found in *Retinitis pigmentosa 1* (*RP1)* (MIM 603937); c.278C>A, p.(Thr93Lys) (NM_006269), number of reads WT/Mut: 80/85. The substitution was listed in the dbSNP (rs139533342), ExAC and gnomAD databases and was rare, with a reported allele frequency of 21/277124 (7.578e-5) in gnomAD (last accessed 25/05/2018). PolyPhen-2, SIFT and CADD all predicted the variant to be pathogenic. HomoloGene showed that the amino acid was fully conserved in all available species (Figure 4-8). Analysis of remaining patient exome data did not reveal any further coding region variants or CNVs in *RP1*.

In the same patient, two heterozygous variants were identified in the gene *Tubulin Tyrosine Ligase-like family member 5* (*TTLL5)* (MIM 612268). Compound heterozygous variants have been reported in 5 patients from 4 unrelated families with autosomal recessive cone-rod dystrophy (CORD19) (MIM 615860) (Sergouniotis *et al.*, 2014). The first variant (NM_015072); c.1310C>T;

p.(Ala437Val); number of reads WT/Mut: 58/37, was present in dbSNP, ExAC and gnomAD (rs112023579). It has been reported in a heterozygous state in 15 African individuals, 4 Latino, 3 South Asian, 2 European (non-Finnish) and 1 East Asian individual in gnomAD (last accessed 25/05/2018). The second variant (NM_015072); c.1409G>A; p.(Arg470Gln), number of reads WT/Mut: 80/78, was also present in dbSNP, ExAC and gnomAD (rs368557430) (Table 4-3). The variant was reported in a heterozygous state in 11 African individuals, 3 Latino, 1 South Asian and 1 European (non-Finnish) individual in gnomAD (last accessed 25/05/2018). The p.(Arg470Gln) variant was predicted to be pathogenic by PolyPhen-2, SIFT and had a CADD score of 26.6. The p.(Ala437Val) variant was scored benign and tolerated by PolyPhen-2 and SIFT, but was reported to be deleterious by CADD, with a score of 23.3 (Table 4-3). The arginine residue was fully conserved at position 470 in all available species (excluding Zebrafish), but the alanine residue at position 437 was not conserved in the chicken (*Gallus*) and frog (*Xenopus*) (Figure 4-8). Analysis of remaining patient exome data did not reveal any further coding region variants or CNVs in *TTLL5*.

No further potentially pathogenic variants in the known retinal or albinism disease genes were identified in the remaining foveal hypoplasia cohort. The variants listed above were not confirmed by Sanger sequencing.

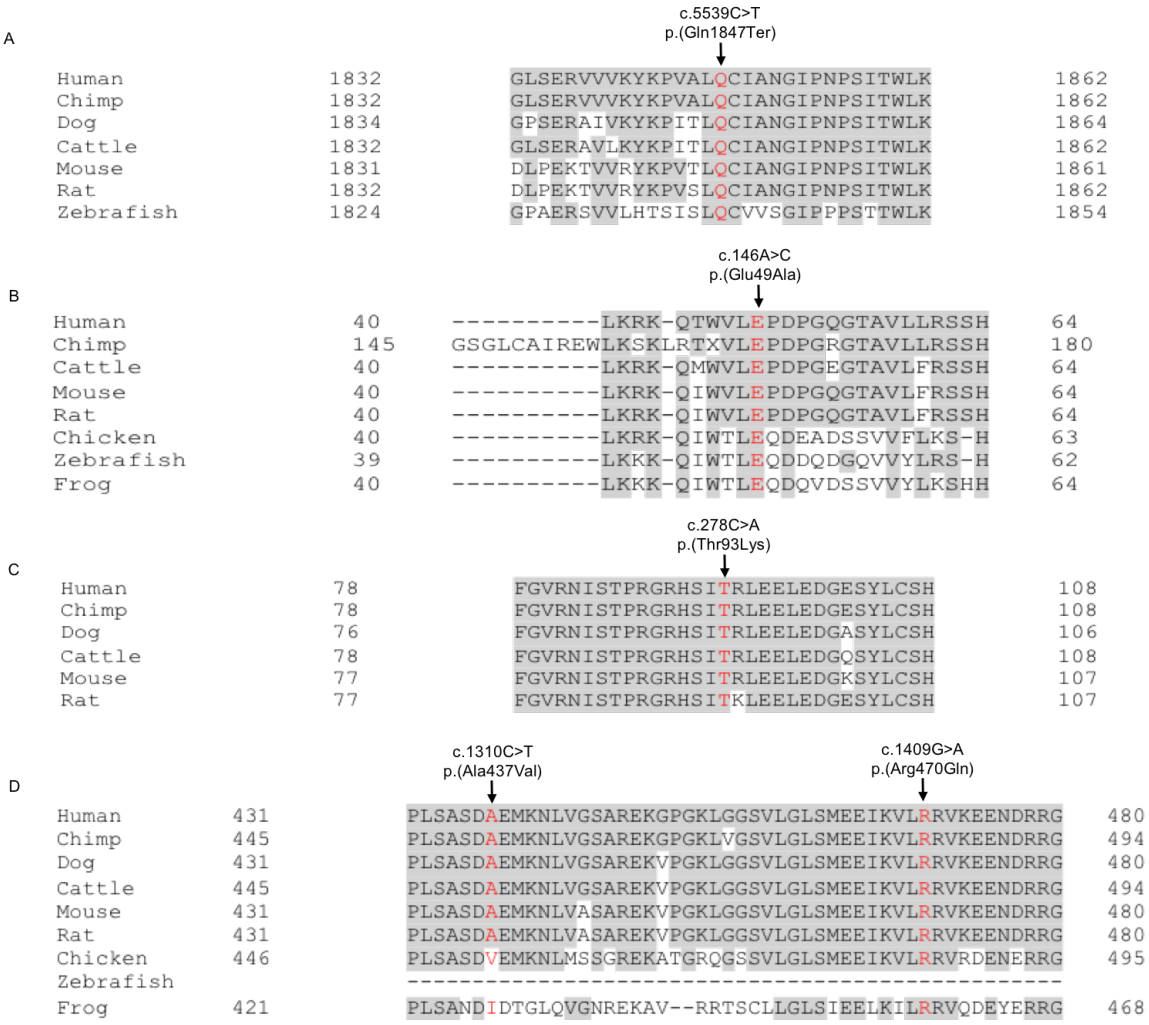


Figure 4-8 Protein sequence alignments of human HMCN1, FSCN2, RP1 and TTLL5 with homologues.

Mutliple sequence alignment was calculated using HomoloGene. The shaded amino acid residues indicate those that are shared with the human residue at the same position. The residues substituted in the patients are coloured red. **A** Patient F1374; *HMCN1* c.5539C>T, p.(Gln1847Ter) - Sequences of HMCN1 used: Human (*Homo sapiens*) NP_114141.2; Chimpanzee (*Pan troglodytes*) XP_514061.2; Dog (*Canis lupus familiaris*) XP_547438.2; Cattle (*Bos taurus*) NP_001179466.1; Mouse (*Mus musculus*) NP_001019891.2; Rat (*Rattus norvegicus*) NP_001258221.1; Zebrafish (*Danio rerio*) NP_001177233.1. **B** Patient F1352; *FSCN2* c.146A>C p.(Glu49Ala) – Sequences of FSCN2 used: Human NP_001070650.1, Chimp XP_003953273.1, Cattle NP_788806.1, Mouse NP_766390.2, Rat NP_001100542.1, Chicken NP_001171209.1, Zebrafish NP_957064.2, Frog NP_001093724.1. **C** Patient F1371; *RP1* c.278C>T p.(Thr93Lys) - Sequences of RP1 used: Human NP_006260.1, Chimp XP_00153192.1, Dog NP_001003040.1, Cattle NP_776383.1, Mouse NP_035413.1, Rat NP_001182605.1. **D** Patient F1371; *TTLL5* c.1310C>T, p.(Ala437Val), c.1409G>A, p.(Arg470Gln) – Sequences of TTLL5 used: Human NP_0055887.3, Chimp XP_001161309.3, Dog XP_005623734.1, Cattle XP_005212294.4, Mouse NP_001074892.2, Rat XP_006240435.1, Chicken NP_001026375.1, Zebrafish XP_003200448.2, Frog NP_001121451.1.

Gene	Patient	Chr position	cDNA & aa change	Variant consequence	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
PAX6	F1351	11:31,823,232	c.275T>C p.(Val78Ala)	Het Missense	-	-	-	-	probably_damaging (1)	damaging	25.7
PAX6	F1378	11:31,823,229	c.278C>A p.(Ala79Glu)	Het Missense	-	-	-	-	probably_damaging (1)	damaging	28.9
HMCN1	F1374	1:185,988,741	c.5539C>T p.(Gln1847Ter)	Het Nonsense	-	-	-	-	-	-	38
FSCN2	F1352	17:79,495,703	c.146A>C p.(Glu49Ala)	Het Missense	rs200600577	4/61280 6.53E-05	10/220092 1.544E-05	0	probably_damaging (0.98)	damaging	24.8
RP1	F1371	8:55,533,804	c.278C>A p.(Thr93Lys)	Het Missense	rs139533342	10/121042 8.26E-05	21/277124 7.578E-05	0	probably_damaging (0.97)	damaging (0.02)	27.1
TTLL5	F1371	14:76,211,846	c.1409G>A p.(Arg470Gln)	Het Missense	rs368557430	5/121364 4.12E-05	16/277012 5.776E-05	0	probably_damaging (0.99)	damaging (0)	26.6
TTLL5	F1371	14:76,211,466	c.1310C>T p.(Ala437Val)	Het Missense	rs112023579	5/120670 4.14E-05	25/276956 9.027E-05	0	benign	tolerated	23.3

Table 4-3 Rare variants identified in known retinal disease genes in the patient cohort.

Filtering WES data using a BED file of known retinal disease genes (Appendix 6). The gene identifier, chromosome (chr) position, amino acid (aa) change, variant consequence, dbSNP rs ID number, allele frequency & count from the gnomAD database (last accessed 12/12/2017), the number of homozygous (hom) alleles present in the ExAC and gnomAD and the pathogenicity prediction scores from PolyPhen-2 (Adzhubei *et al.*, 2010), SIFT (Ng and Henikoff, 2003) and CADD (Kircher *et al.*, 2014) are presented. CADD score of ≥20 means that the variant is in the top 1% most deleterious mutations in the genome.

4.2.3 Investigating newly implicated genes causative of foveal hypoplasia in the patient cohort

Following the analysis of known IRD and albinism genes in the WES cohort, variant lists in unrelated patients were then compared to look for different mutations in the same gene. This combined patient cohort (n=10) included patients from the *SLC38A8*-negative panel (excluding F1351 and F1378 with *PAX6* variants). Only F1288 of the consanguineous sib pair was analysed. This analysis showed that there were no genes in which 2 or more patients had homozygous or compound heterozygous variants in the same gene.

Following the bioinformatics analysis to look for variants in genes shared in the combined patient cohort, the remaining individuals were investigated on a case-by-case basis starting with the consanguineous cases. Regions of homozygosity were investigated in all patients in case there were any unknown consanguineous patients. Homozygosity mapping (Section 2.4.6) was used to refine the candidate variant lists in three known consanguineous cases, F1374, F1335 and F1288 (of the consanguineous sib pair). The final number of potentially pathogenic homozygous variants on a case-by-case basis are presented in Table 4-4.

Patient	All biallelic variants with MAF ≤0.0001	Functional variants	PASS filter variants	Variants with CADD >15	No. of hom candidates
F1071	106	75	46	18	0
F1288*	101	60	42	22	7
F1298	100	70	39	19	0
F1335*	105	98	56	28	1
F1352	105	79	47	16	0
F1369	97	58	31	15	0
F1371	110	78	35	19	0
F1374*	97	58	39	19	7
F1405	101	58	35	15	0

Table 4-4 Final filtered variant statistics for each individual in the biallelic analysis.

Total number of biallelic variants remaining after filtering using the strategic pipeline detailed in Section 2.3.22.2 and Figure 2.3. The raw Excel output of biallelic variants with a MAF ≤0.0001 was manipulated to retain all coding and splice site variants (functional), those variants that passed (PASS) the GATK QC filter and those with a CADD score >15. The final number of homozygous (hom) candidate variants are shown. *Consanguineous patients.

4.2.3.1 Rare homozygous variants identified in consanguineous patient F1374

In patient F1374 homozygosity mapping identified the homozygous regions presented in Table 4-5 and Figure 4-9. Homozygosity mapping using F1374's exome data showed that 7 rare variants lay in these regions of homozygosity (Table 4-6).

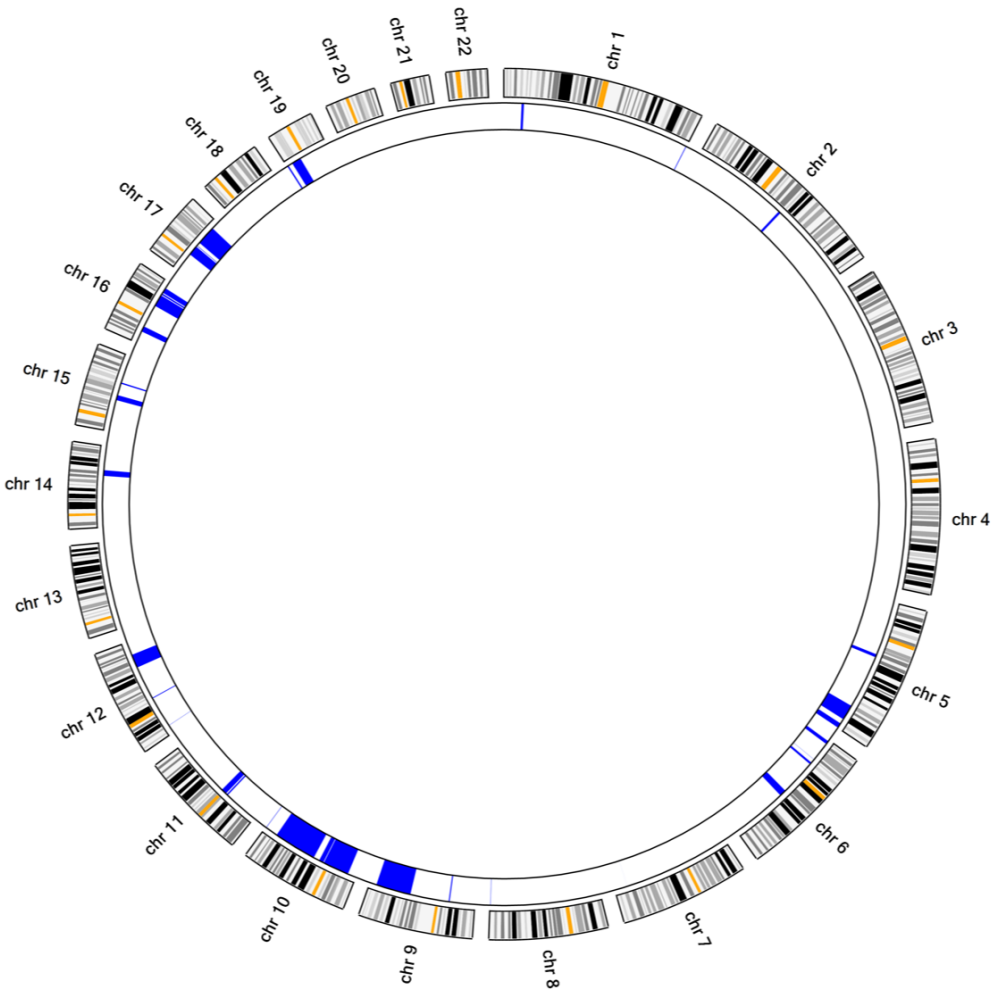


Figure 4-9 Homozygosity mapping in case F1374.

Locations of homozygous regions (blue) identified by AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (Section 2.4.6) from exome sequence data of patient F1374 against a circular ideogram of chromosomes 1-22.

Chromosomal location	Length (Mb)
10:61,122,268-118,404,620	57.28
9:85,987,879-132,402,908	46.42
10:11,299,735-46,242,043	34.94
17: 45,669,428-71,410,891	25.74
5:146,971,253-167,566,600	20.60
16:71,319,539-73,165,038	18.45
12:98,332,704-116,412,922	18.08
16:54,319,571-70,163,691	15.84
17:25,912,828-39,081,713	13.17
19:7,831,628-20,414,254	12.58
14:68,264,867-75,760,855	7.50
16:16,284,272-23,533,008	7.25
11:49,195,090-56,431,216	7.24
16:74,425,548-81,253,917	6.83
15:42,305,852-49,044,538	6.74
5:174,152,100-180,660,918	6.51
10:47,087,608-53,059,406	5.98
6:88,108,048-96,974,386	5.05
6:1,726,623-6,196,142	4.47
5:71,412,373-75,427,935	4.02

Table 4-5 Location (GRCh37/hg19) and size of homozygous regions in case F1374.

The homozygous regions were identified by AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (Section 2.4.6) from exome sequence data and are listed by size (Mb).

Gene	RefSeq	Chr position	cDNA change	aa change	Variant consequence	No. reads WT/Mut	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
KIAA1217	NM_019590.4	10:24,825,761	c.3473G>A	p.(Arg1158Glu)	Missense	0/76	rs145406190	16/121338 1.32E-04	36/276464 1.3E-04	0	possibly_damaging (0.71)	damaging	32
FOXD1	NM_004472 .2	5:72,743,607	c.581C>T	p.(Gly194Asp)	Missense	0/32	.	.	.	.	disease_causing (1)	damaging	25.6
PTGER1	NM_000955.2	19:14,584,580	c.553C>T	p.(Gly185Ser)	Missense	0/13	.	.	.	.	possibly_damaging (0.9)	damaging	25.5
MYOF	NM_013451.3	10:95,132,767	c.2377G>A	p.(Ala793Thr)	Missense	0/140	rs551469563	1/120740 8.28E-06	22/246172 8.94E-05	0	benign (0.31)	damaging	25.3
TCTN1	NM_001082538.2	12:111,070,326	c.674C>G	p.(Ser225Cys)	Missense	0/100	rs766174385	1/120770 8.28E-06	.	0	probably_damaging (0.97)	damaging	24.5
PLCE1	NM_001165979.2	10:96,084,707	c.5855C>G	p.(Thr1952Ser)	Missense	0/33	.	.	.	.	benign (0.35)	tolerated	23.1
CNOT6	NM_001303241.1	5:179,980,427	c.341T>C	p.(Phe114Ser)	Missense	0/120	.	.	1/246154 4.06E-06	0	benign (0.05)	tolerated	21.1

Table 4-6 Rare homozygous variants identified in case F1374.

Summary of frequency and predicted pathogenicity scores for the rare (MAF ≤0.0001) homozygous variants identified in patient F1374. The gene identifier, chromosome (chr) position, amino acid (aa) change, variant consequence, dbSNP rs ID number, allele frequency & count from the gnomAD database (last accessed 12/12/2017), the number of homozygous (hom) alleles present in the ExAC (http://exac.broadinstitute.org) and gnomAD (http://gnomad.broadinstitute.org) databases and the pathogenicity prediction scores from PolyPhen-2, (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al.*, 2010), SIFT (http://sift.jcvi.org/) (Ng and Henikoff, 2003) and CADD (http://cadd.gs.washington.edu/score) (Kircher *et al.*, 2014) are presented.

In a region of approximately 4.02Mb on chromosome 5 (GRCh37, 5:71,412,373-75,427,935) (Figure 4-9), a missense variant in the gene *Forkhead Box D1* (*FOXD1*) (MIM 601091) was prioritised, as *FOXD1* encodes a transcription factor that is required for the specification of the temporal retina in mammals and is expressed in the ventral diencephalon and temporal retina in the CNS. It is one of the earliest cues in the development of the visual system and animal mutant models have aberrant axonal projections (Hatini, Tao and Lai, 1994), making it an excellent candidate gene for the patient phenotype described in this study.

The variant, (NM_004472); c.581G>A; p.(Gly194Asp), number of reads WT/Mut: 0/32 (Table 4-6), was absent from the publicly available databases dbSNP and ExAC (Section 2.4.2). As mentioned previously, gnomAD had not been released at this stage of the project, so the variant was only checked in dbSNP and ExAC. In ExAC, the depth of coverage over the gene was poor, with a mean coverage of 18X but there were also regions with no coverage. The gene was also poorly covered by WES data, as visualised in the IGV (Figure 4-10, A and B).

The variant was predicted to be highly pathogenic by PolyPhen-2 (score of 1) and SIFT and the CADD score for this variant was 25.6 (Table 4-6). There was a single homozygous LoF variant listed in the ExAC database in a European (non-Finnish) individual (frameshift, chr5:72,743,299, p.(Arg297AlafsTer170)). As the gene was a prioritised candidate, more detailed assessment of the variant using a further 5 pathogenicity prediction tools, including MutationTaster and FATHMM was undertaken. All tools scored the variant as disease-causing (Table 4-7). The variant

was confirmed and segregated by Sanger sequencing using primers listed in Appendix 10 (Section 2.2, 2.3.7) (Figure 4-11). The glycine residue was conserved in many species and in other FOXD proteins (Figure 4-12).

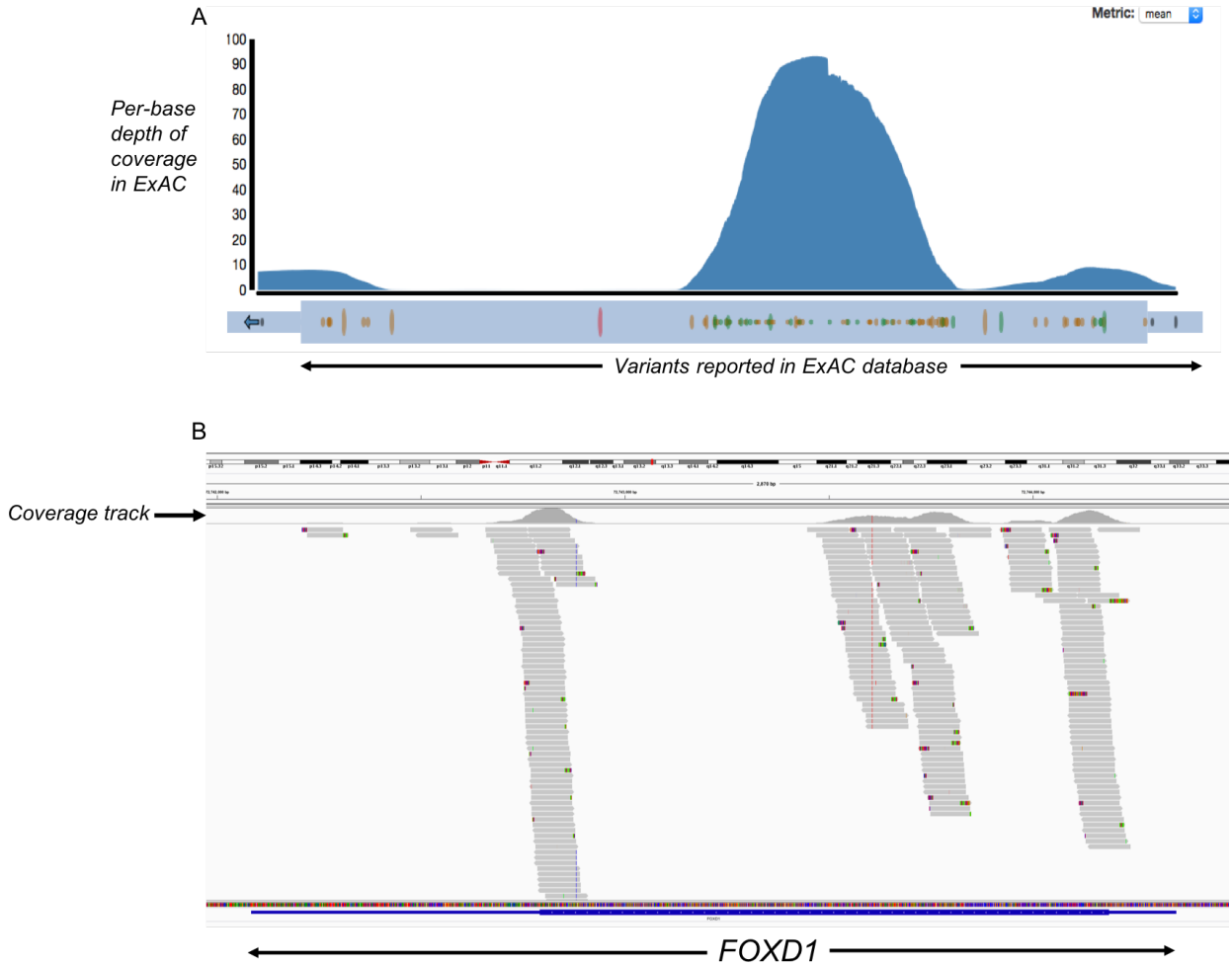


Figure 4-10 Mean depth of coverage of the *FOXD1* gene in ExAC and F1374 exome data.

(A) The poor coverage of *FOXD1* as visualised in the ExAC (http://exac.broadinstitute.org) publicly available database (last accessed 12/12/2017). The mean coverage over the single-exon gene was recorded as 18X in ExAC and there are large regions where there is no coverage at all. (B) Similar low sequence coverage over *FOXD1* in patient F1374, as visualised in the Integrative Genomics Viewer (IGV) (https://software.broadinstitute.org/software/igv/).

Chromosome 5 position	cDNA change	Protein change	PolyPhen-2	MutationTaster2	Blosum 62	SIFT	MutPred	PROVEAN	CADD Score	FATHMM (weighted score)
72,743,607	c.581G>A	p.(Gly194Asp)	Probably damaging (1)	Disease causing (probability 0.99)	-1	Deleterious	Deleterious mutation (probability 0.817)	Deleterious (-6.967)	25.6	Deleterious (-5.01)

Table 4-7 Bioinformatics analysis of the novel *FOXD1* variant identified in F1374.

Bioinformatics analysis of *FOXD1* missense variant identified in individual F1374, c.581G>A, p.(Gly194Asp). The variant was annotated using the reference sequence NM_004472 and Human Reference Build GRCh37/hg19. URLs: PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/ (Adzhubei *et al.*, 2010); MutationTaster2, http://www.mutationtaster.org/ (Schwarz *et al.*, 2010; Schwarz *et al.*, 2014); SIFT, http://sift.jcvi.org/ (Ng and Henikoff, 2003); Blosum62* (Henikoff and Henikoff, 1993); PROVEAN, http://provean.jcvi.org/ (Choi *et al.*, 2012); MutPred 1.2, http://mutpred.mutdb.org/ (Li *et al.*, 2009); CADD Score, http://cadd.gs.washington.edu/score (Kircher *et al.*, 2014); FATHMM score (weighted) http://fathmm.biocompute.org.uk/index.html (Gough *et al.*, 2001). *Blosum62 scores range from +3 to -3; negative scores are more likely to be pathogenic substitutions. CADD score of ≥20 means that the variant is in the top 1% most deleterious mutations in the genome.

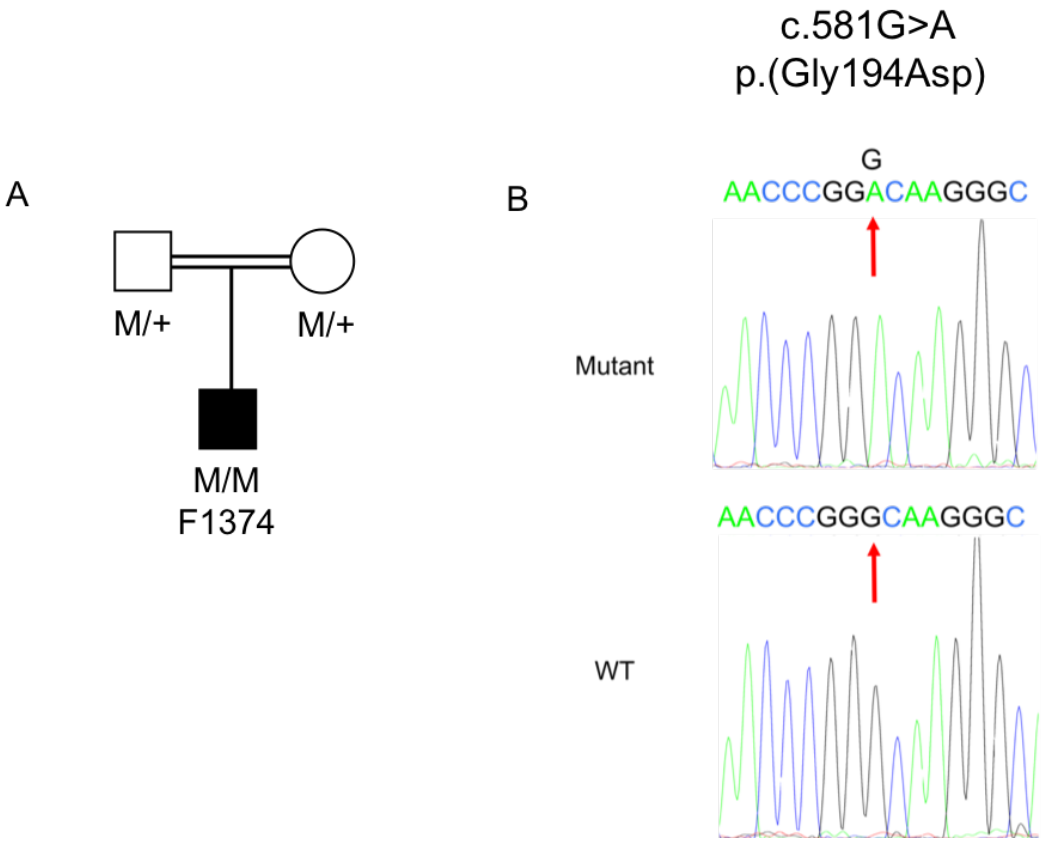


Figure 4-11 Segregation analysis of the *FOXD1* variant in case F1374 and family.

(A) Pedigree and segregation data for the mutation in *FOXD1* (NM_004472.2); c.581G>A; p.(Gly194Asp). The genotypes for all tested family members are shown below each individual, with M representing the mutant allele and + representing the wild-type allele. The affected individual, F1374, is shaded black. (B) Sequence electropherograms highlighting the *FOXD1* mutation in patient F1374 shown against the wild-type sequence from a control individual with no known disease phenotype.

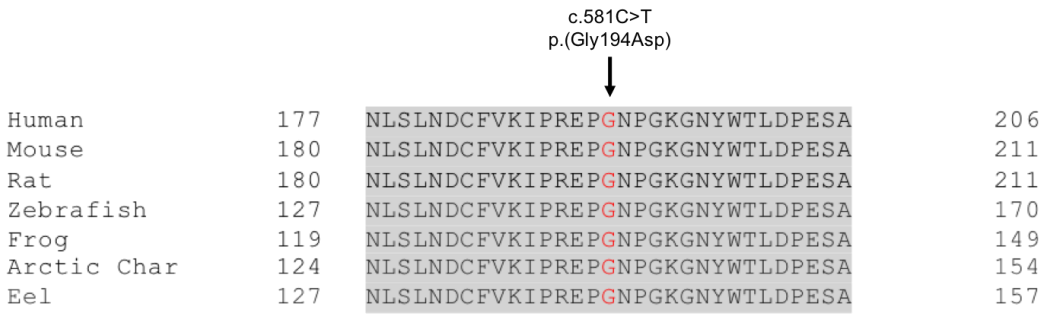


Figure 4-12 Multiple sequence alignment of human FOXD1 and homologues.

The shaded amino acid residues indicate those that are shared with the human residue at the same position. The residue substituted in F1374 is coloured red. Sequences of *FOXD1* available in HomoloGene: Human (*Homo sapiens*) NP_004463.1;; Mouse (*Mus musculus*) NP_032268.2; Rat (*Rattus Norvegicus*) NP_599164.1; Zebrafish (*Danio rerio*) AAH75922.1; Frog (*Xenopus laevis*) NP_001079052.1; Arctic Char (*Salvelinus alpinus*) XP_023843140.1; Eel (*Mastacembelus armatus*) XP_026153509.1.

A schematic of the FOXD1 protein was generated using the online prediction tools Protter and SMART, with the consensus shown as generated by Protter (Section 2.4.10.2) (Figure 4-13). Mapping of the variant on to the Protter schematic predicted the variant to lie in the forkhead DNA-binding domain of the protein.

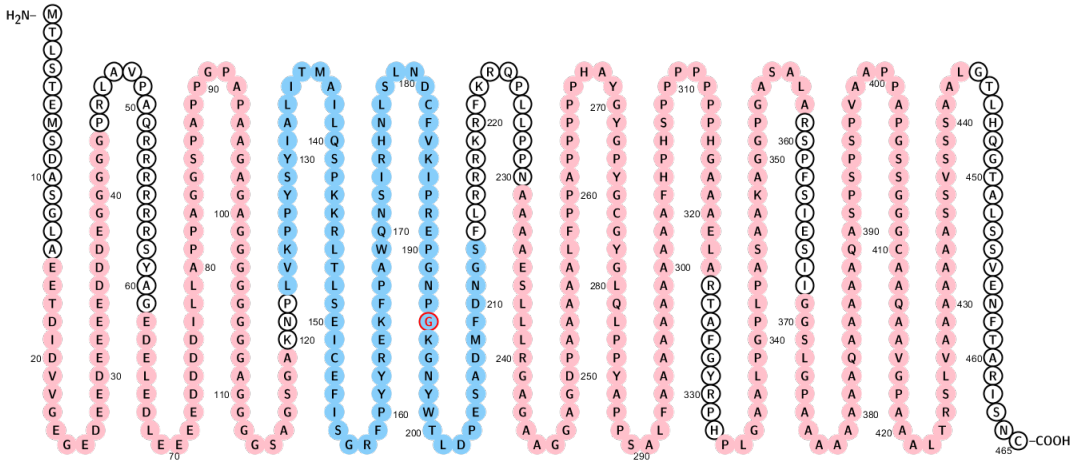


Figure 4-13 The p.(Gly194Asp) *FOXD1* variant mapped on to a protein schematic.

Schematic representation of the 465 amino acid protein structure of FOXD1 (Consensus shown from SMART and Protter tools, Section 2.4.7). Unknown regions are outlined in black, low complexity regions are coloured pink (15-43aa; 62-79aa; 80-119aa; 231-323aa; 333-358aa; 369-443aa) and the forkhead domain (FH) is coloured blue (123-213aa). The p.(Gly194Asp) variant is located in the forkhead DNA binding domain and the affected amino acid is outlined in red.

4.2.3.1.1 Investigating *FOXD1* as a potential novel disease gene

As described earlier in the chapter, the first samples to undergo WES were patients F1351, F1369, F1374 and the non-consanguineous sib pair F1372 and F1373. At this stage of the project, *PAX6* screening in the *SLC38A8*- mutation-negative cohort was underway. This was ongoing alongside the identification of the homozygous missense variant p.(Gly194Asp) in *FOXD1* in patient F1374. The variant in *FOXD1* was prioritised as the main candidate in patient F1374 and the best overall candidate in the foveal hypoplasia cohort.

Examination of the *FOXD1* locus in the IGV ascertained poor coverage over the gene in the first batch of patients (F1351, F1369 and the sib pair F1372 and F1373). There were no obvious structural changes identified in these patients in IGV, however the read depth coverage over the region of this single-exon gene was noted to be extremely poor (Figure 4-10).

To explore *FOXD1* expression in humans, *in silico* analysis was performed using the GENEVESTIGATOR platform (Section 2.4.8). Through this platform, HG-U133 Plus 2.0 GeneChip microarray (Affymetrix, Santa Clara, USA) datasets of the human fetal retina were available for interrogation (Kozulin and Provis, 2009; Kozulin *et al.*, 2009b). The data from this study confirmed high expression of *FOXD1* at 19-20 weeks in the macula and surrounding retina, with moderate expression levels in the nasal retina, further supporting the gene as a strong candidate for the disease phenotype (Figure 4-14).

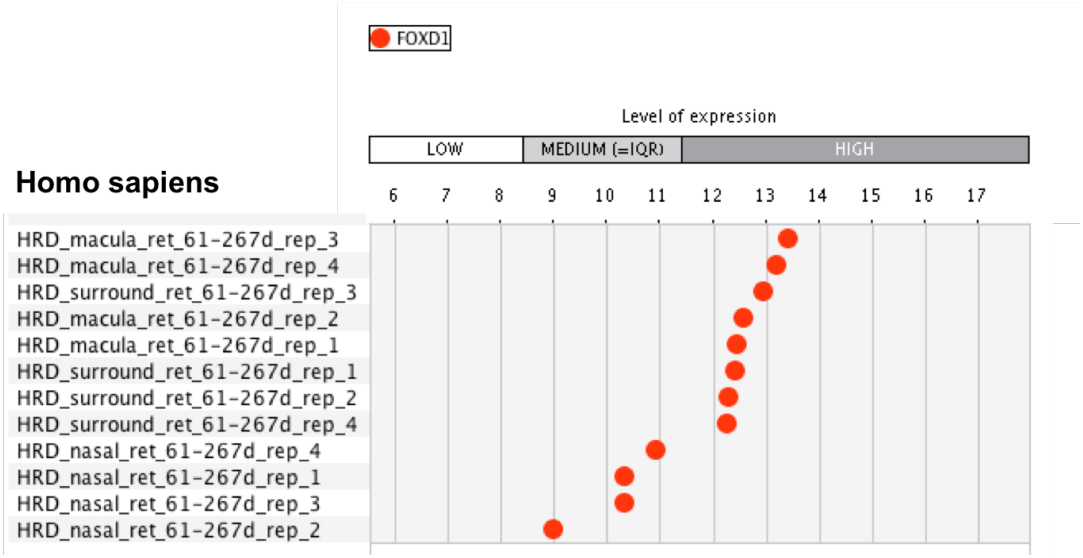


Figure 4-14 Human fetal expression of *FOXD1* in macula, nasal and surrounding retina.

GENEVESTIGATOR (Section 2.4.8) was used to investigate *FOXD1* expression in human microarray datasets (Kozulin and Provis, 2009; Kozulin *et al.*, 2009b). Fetal tissue (with no known disease) used was 19-20 weeks post gestation. *FOXD1* is highly expressed in fetal macula and surrounding retina, with moderate expression in nasal retina. Low (first quartile), Medium (inter-quartile range) and High (fourth quartile) are shown on the plot. The software states that these ranges are determined by looking at all expression values of all genes over all samples.

Sanger sequencing of *FOXD1* was undertaken in the cohort of 17 patients (not including case F1351, as the *PAX6* missense change p.(Val78Ala) had already been identified by WES in this patient). Amplification of the single-exon gene was performed in three amplicons. PCR and Sanger sequencing conditions and primers are described in Appendix 10. Considerable effort went into designing the PCR screen for this gene due to its GC-rich nature. Five different commercial kits, including Invitrogen Taq (Life Technologies), KAPA HiFi HotStart Polymerase (KAPA Biosystems, US), HotShot Polymerase (Clent Life Science, UK), Phusion DNA Polymerase (NEB) and Q5 Polymerase (NEB), were tested before the final method using the KAPA and NEB Q5 kits were optimised (Appendix 10). Various

combinations of PCR additives, including DMSO, betaine and 7-deaza-dGTP were also optimised in the assay. Variants in *FOXD1* that were identified in the cohort of patients, but not thought to be pathogenic are shown in Appendix 18. No further likely pathogenic variants were identified.

4.2.3.1.2 Investigating the potential pathogenicity of the *FOXD1* variant using a dual-luciferase reporter assay

To investigate the pathogenicity of the variant identified in F1374 (c.581G>A, p.(Gly194Asp)), functional analysis of the transcription factor was assessed using a dual-luciferase reporter assay. A recent study by Laissue and colleagues utilised a luciferase reporter assay to show that variants in *FOXD1* altered its ability to interact with and transactivate the promoter of placental growth factor (*PGF*) (Laissue *et al.*, 2016). The luciferase reporter had the *PGF* promoter region (720bp from -1144bp to -424bp upstream of the initial ATG start codon), cloned upstream of a luciferase gene so that activation of the promoter resulted in a quantitative luminescence signal. In this study, the same experimental approach was used to determine if the p.(Gly194Asp) variant identified in patient F1374 had any effect on FOXD1's ability to transactivate the *PGF* promoter, compared to WT FOXD1.

For this experiment, expression clones were created to generate WT and variant p.(Gly194Asp) FOXD1. Briefly, the parent entry clone for *FOXD1* (*FOXD1*-pENTR233.1) (Accession number: BC160026) containing the entire ORF of *FOXD1* (with a stop codon) was obtained from the plasmid repository DNASU (The Biodesign Institute, Arizona State University), and was sequence verified (Section

2.3.9). This entry clone was used to facilitate the introduction of the gene into the destination vector pDEST40 using Gateway technology (Section 2.3.17). SDM was performed on the *FOXD1* entry clone to introduce the p.(Gly194Asp) variant (Section 2.3.16) (SDM primers listed in Appendix 10). All clones were sequence verified (Section 2.3.9, Appendix 11) to ensure the *FOXD1* ORF was full-length, contained no variants and that the SDM reaction had worked. The resulting expression constructs were named pDEST40_*FOXD1*-WT and pDEST40_*FOXD1*-Mut. The control (pGL3-Basic) and *PGF* (pGL3-*PGF*) promoter reporter constructs were kindly provided by Dr Daniel Vaiman (INSERM, Paris, France).

To optimise this experiment, COS7 cells were transiently transfected with pDEST40_*FOXD1*-WT in combination with the pGL3-*PGF* and a renilla luciferase transfection control, pRL-TK (Section 2.3.21). Controls included cells transfected with the expression construct backbone vector (pDEST40), cells transfected with the promoter backbone vector (pGL3 basic) and mock transfection cells (cells treated with transfection reagent only). Cells were transfected when they were 70-80% confluent and each well of the 24-well plate was transfected with the same amount of total DNA (800ng per well). 48 hours post-transfection, cell lysis was performed and *Renilla* and firefly luciferase levels were recorded using the Promega Luciferase Assay System (Section 2.3.24). For every well, the firefly signal was normalised to the *Renilla* signal and pathway activation levels were calculated using firefly:*Renilla* ratio, conveyed as relative luciferase units (RLU) (Section 2.3.21). The assay was performed in triplicate, with RLU activity of the constructs shown in Figure 4-15.

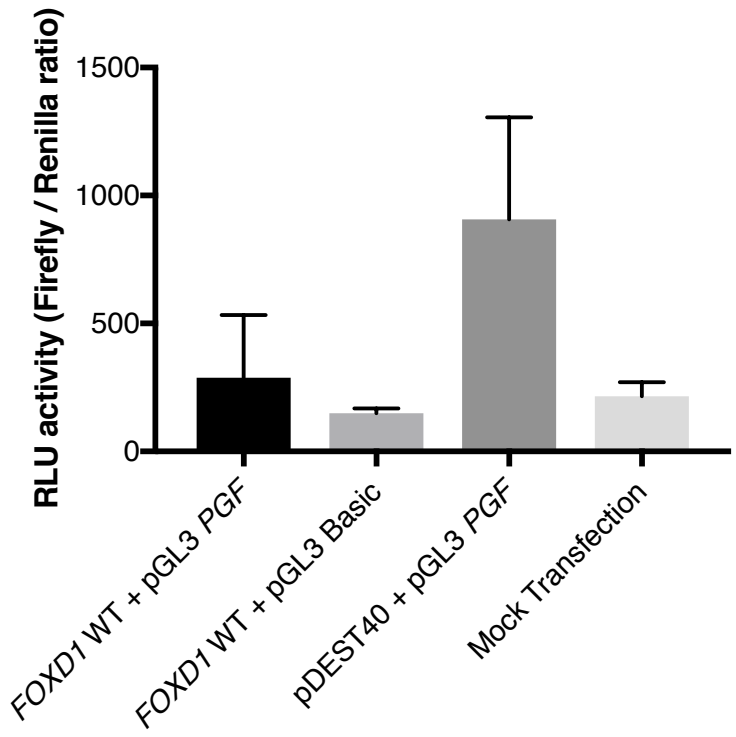


Figure 4-15 *FOXD1*-WT dual-luciferase reporter assay in COS7 cells.

Luciferase activity was measured and expressed as Relative Luciferase Units (RLU). "Mock transfection" was used as the experimental transfection control and was COS7 cells treated with transfection reagents only. "pDEST40 + pGL3 *PGF*" COS7 cells contain the *PGF* promoter construct only. "FOXD1 WT + pGL3 Basic" COS7 cells contain the FOXD1 expression construct, but no *PGF* promoter reporter construct. "*FOXD1* WT + pGL3 *PGF*" COS7 cells were transfected with both the *FOXD1* construct and *PGF* reporter. Error bars show the standard error of the mean.

The results did not show the expected pathway transactivation results as previously reported by Laissue and colleagues for WT FOXD1 with the PGF promoter, but this was a first pass experiment and transfection efficiencies and plating efficiencies had not been optimised.

At this stage of the project, as optimisation of this assay was ongoing, the control genome database gnomAD was released. *FOXD1* coverage was greatly improved in the genomic data (Figure 4-16) and interrogation of variants in *FOXD1* in the

gnomAD database showed a high frequency (n=11) of LoF variants throughout this gene, one of which was homozygous. This new data made the candidacy of *FOXD1* as a new gene mutated in foveal hypoplasia less plausible and this experiment was therefore not continued.

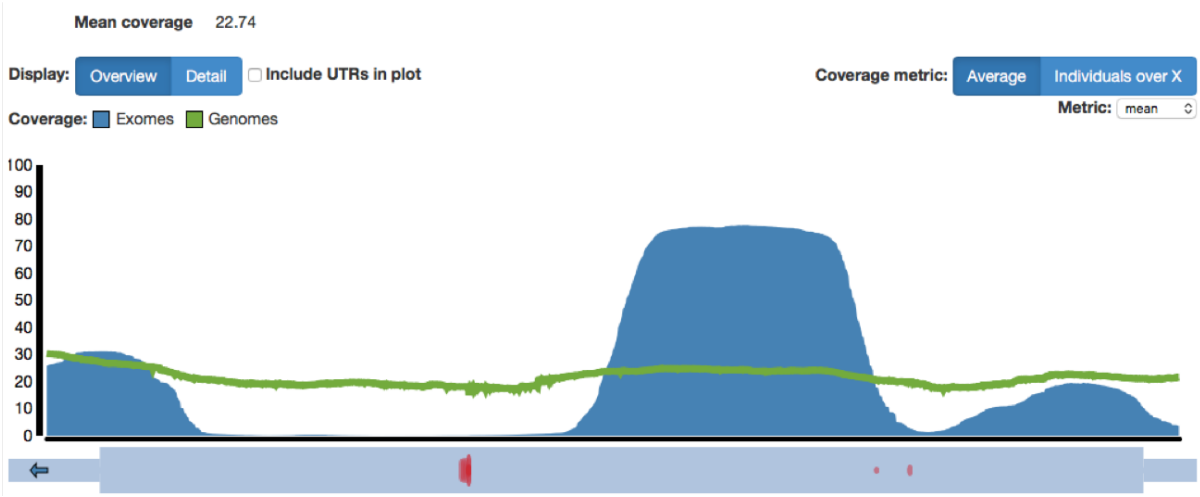


Figure 4-16 Mean depth of coverage over *FOXD1* in the gnomAD database.

The improved mean coverage (22.74X) over *FOXD1* as visualised in the gnomAD (green line) (http://gnomad.broadinstitute.org) publicly available database (last accessed 12/12/2017). The mean coverage over the single-exon gene was recorded as 18.06X in the ExAC database of exomes (blue). LoF variants shown in red at the base of the figure.

4.2.3.2 CNV ExomeDepth analysis identifies a homozygous duplication in *SLC38A8* in F1335

Investigation of the homozygous regions in patient F1335 revealed a region of approximately 7.49 Mb on chromosome 16 (GRCh37, chr16:78,198,192-85,691,294) which overlapped with *SLC38A8* (Figure 4-17). This data supported the previous haplotype analysis showing that case F1335 harboured a homozygous region of approximately 3.4Mb on chromosome 16 surrounding *SLC38A8* (Section 3.2.2). However, the WES data confirmed the previous Sanger sequencing result and identified no variants in the coding region or splice sites of *SLC38A8* in this patient.

These data support patient F1335 harbouring a missing mutation in *SLC38A8,* prompting a search for CNVs in *SLC38A8* in the WES data.

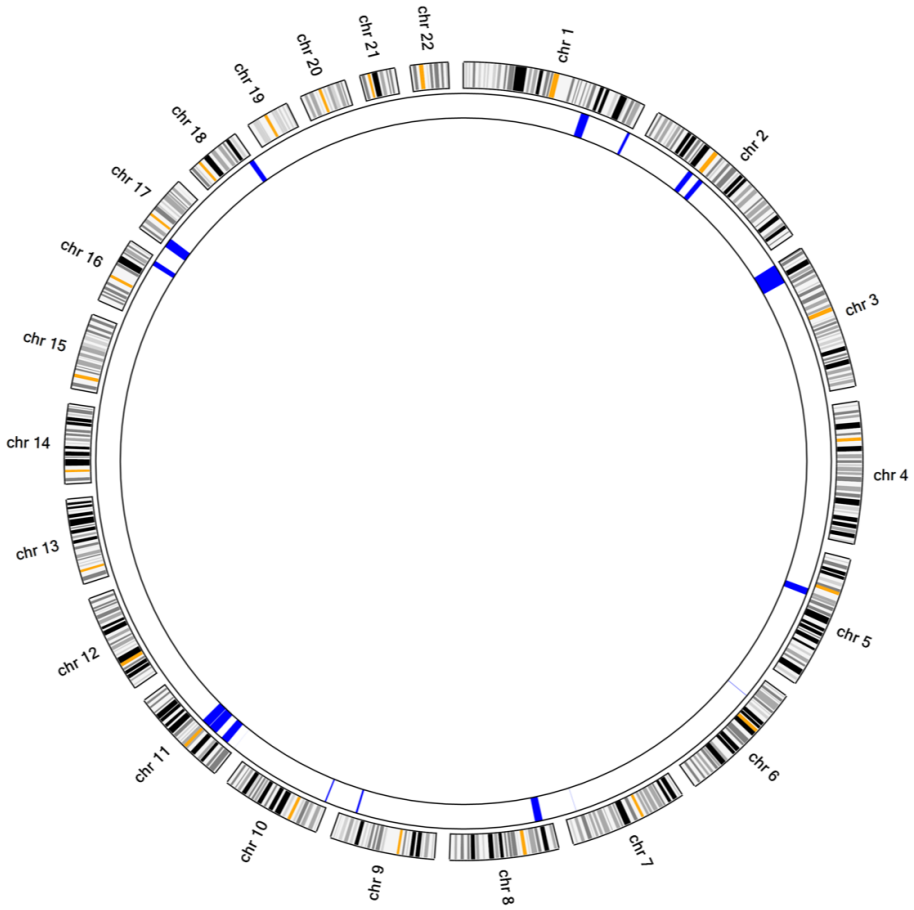


Figure 4-17 Homozygosity mapping in case F1335.

Locations of homozygous regions (blue) identified by AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (Section 2.4.6) from exome sequence data of patient F1335 against a circular ideogram of chromosomes 1-22. A region of approximately 7.46 Mb was identified on chromosome 16 (chr16: 78,198,192-85,691,294), spanning *SLC38A8*.

Alongside the basic WES analysis to identify candidate variants, the WES data for each individual was investigated for CNVs using the software ExomeDepth (Section 2.3.26.1). This software recommends ≥10 reference exomes are compared for optimal function; however, the software selects which of these reference exomes are used in the analysis. The number of references available, and those selected for

each ExomeDepth CNV analysis are shown in Appendix 17. The CNV lists were filtered to remove any CNVs previously reported in the paper by Conrad and co-workers (Conrad *et al.*, 2010). Analysis of ExomeDepth data in all 11 individuals (of the 12 sent for WES), revealed a number of CNVs that were not classified as common (Appendix 17). The resulting CNV calls were then ranked based on the highest Bayes Factor (BF), representing $\log_{10}$ of the likelihood ratio for the CNV call compared to that of normal copy number for that location (Section 2.3.25.4) (Plagnol *et al.*, 2012). The number of remaining CNVs was highly variable between patients, ranging from only 2 CNVs (case F1352) to 100 (case F1371). In the first instance, the CNV lists were interrogated to check if loci were highlighted that contained known genes responsible for foveal hypoplasia including *SLC38A8*, *PAX6*, *FRMD7* and the albinism genes. This analysis highlighted a homozygous duplication in case F1335 that encompassed exon 6 of *SLC38A8* (Table 4-8).

CNV predicted location	No. of exons	BF	Reads expected	Reads observed	Reads ratio
chr16: 84,050,746-84,056,494	1	17.3	152	348	2.29

Table 4-8 ExomeDepth identifies a homozygous duplication over exon 6 of *SLC38A8* in case F1335.

The homozygous duplication identified in patient F1335 using ExomeDepth. The predicted location (GRCh37), Bayes Factor (BF), Number of reads expected/reads observed and also the Reads ratio are shown. The BF is $\log_{10}$ of the likelihood ratio for the CNV call divided by normal copy number call at the same position. It is therefore a ratio by which a higher number suggests that the CNV is more likely to be true.

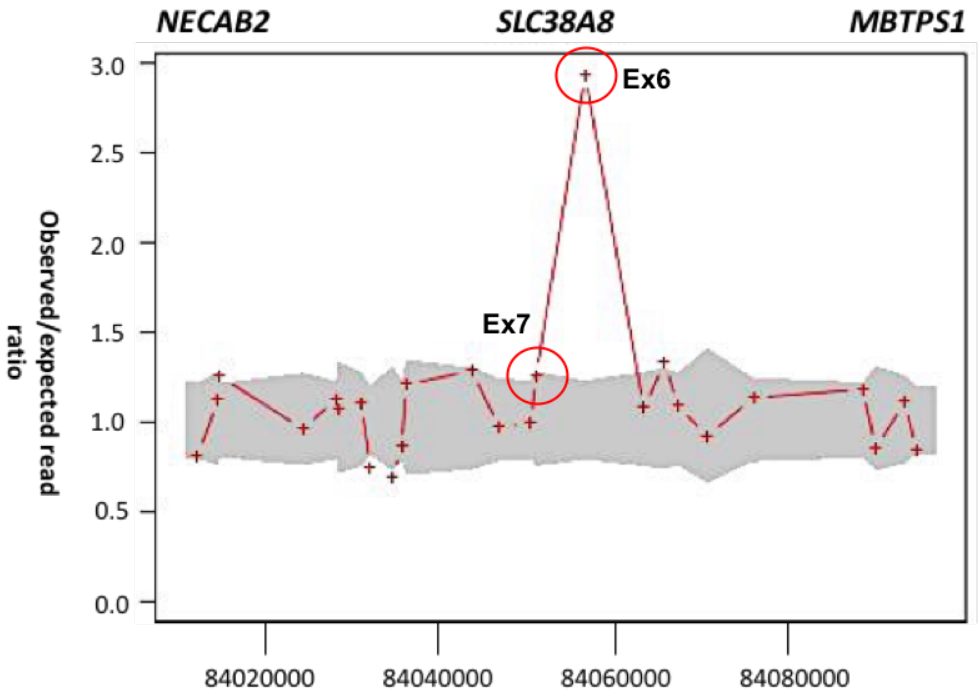


Figure 4-18 ExomeDepth data plot of the predicted duplication over *SLC38A8* exon 6.

The homozygous duplication over exon 6 (red circle) identified in patient F1335 using ExomeDepth software presented using the plot function. The red crosses indicate the ratio of the observed/expected reads for each exon. The grey shaded area represents the estimated 95% confidence interval of the ratio. The presence of an exon outside of this confidence interval indicates the existence of a duplication in this individual.

The ExomeDepth reads ratio for case F1335 was 2.29, suggesting that the duplication was homozygous, according to the ExomeDepth software guide, which is consistent with the homozygosity mapping data. The graph output shows the data based on the depth of coverage (Figure 4-18). This finding was confirmed using the GATK “DepthofCoverage” function. A BED file containing all exons of *SLC38A8* was added to the command in order to calculate the coverage over all coding regions. The read depth results were compared with five unrelated exomes that were processed in an identical manner to individual F1335 (aligned to GRCh37, indels sorted and realigned, PCR duplicates removed and recalibrated BAM files). This

analysis revealed an increase in read number (220.13 average reads/base, SD 50) over the coordinates chr16:84,056,380-84,056,494, representing *SLC38A8* exon 6, compared with an average of 61.62 reads/base (range 49.7-72.89) in the five comparison exomes at the same position.

Investigation of CNVs in the region of *SLC38A8* in the ExAC database revealed a number of large duplications (>60 Kb), however they were not found to overlap with the duplication identified in patient F1335. Similarly, in the Decipher database (https://decipher.sanger.ac.uk) there were 33 large duplications/deletions, but breakpoints did not overlap with those identified in F1335. The identified duplicated region was absent from the CNV browser (https://personal.broadinstitute.org/handsake/mcnv_data/) and the Database of Genomic Variation (http://dgv.tcag.ca/dgv/app/home?ref=GRCh37/hg19).

4.2.3.3 The identification of *LAMP1* as a new candidate foveal hypoplasia gene

Consanguineous case F1288 was a South Asian female with foveal hypoplasia. The parents of F1288 were not reported to have any disease phenotype but the patient had a second affected sib (case F1287). Previous haplotype analysis (Section 3.2.2) showed that the affected sib pair were unlikely to harbour a variant in *SLC38A8*. Homozygosity mapping using F1288's exome data showed that 7 candidate homozygous variants lay in regions of homozygosity (Figure 4-19 and Table 4-9). Four of these variants were not prioritised as they were not predicted to be pathogenic by SIFT, PolyPhen-2 and CADD (Table 4-10).

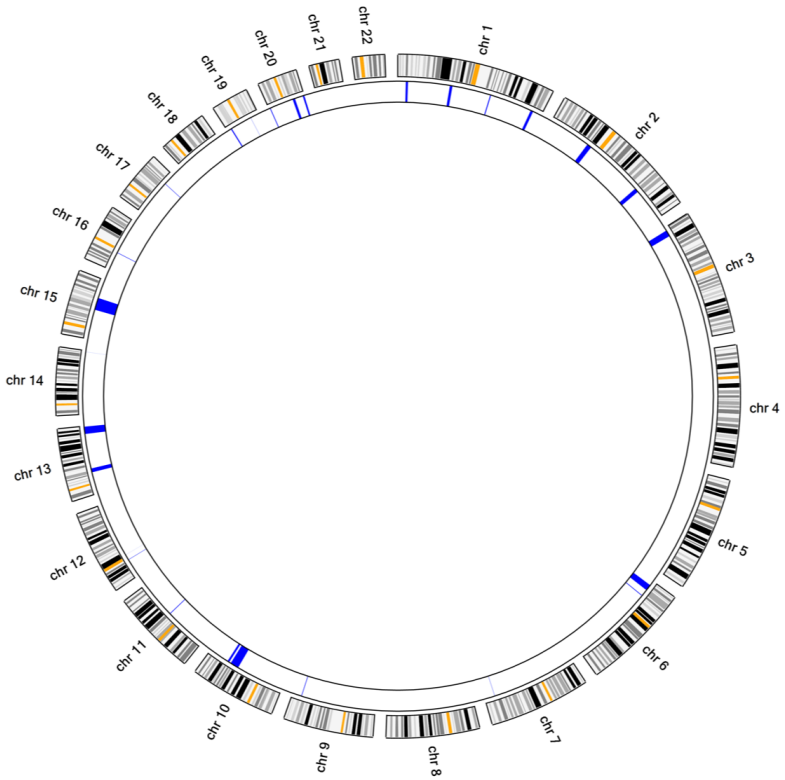


Figure 4-19 Homozygosity mapping in case F1288.

Locations of homozygous regions (blue) identified by AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (Section 2.4.6) from exome sequence data of patient F1288 against a circular ideogram of chromosomes 1-22.

Chromosomal location	Size (Mb)
15:50,474,766-70,991,925	20.52
10:81,286,427-95,720,501	14.43
13:102,344,867-114,620,051	12.28
3:5,241,223-15,900,973	10.66
6:8,026,577-18,427,784	10.40
13:36,940,028-43,463,299	9.52
2:71,417,065-79,255,426	7.84
1:229,663,064-234,041,422	4.38
1:87,458,695-91,859,795	4.40
20:42,302,371-46,365,571	4.33
1:13,183,573-16,802,890	3.62
10:98,808,866-102,053,166	3.24
20:60,145,672-62,858,844	2.71

Table 4-9 Location (GRCh37) and size of homozygous regions in patient F1288.

The homozygous regions were identified by AgileMultiIdeogram (https://www.dna.leeds.ac.uk/agile/AgileMultiIdeogram).

In the 3rd largest (12.28Mb) region of homozygosity on chromosome 13 (Figure 4-19), a homozygous frameshift variant was identified in the gene *Lysosomal associated membrane protein 1* (*LAMP1)*. The frameshift variant in *LAMP1* (MIM 153330) (NM_005561.3); c.1108delG, p.(Gly370Afs*14), number of reads WT/Mut: 2/113, was not reported in dbSNP, ExAC or gnomAD (last accessed 07/07/2018). The variant was prioritised as the main candidate in this patient and it segregated appropriately in the four available family members, including the sibling F1287 (Figure 4-20) (primers listed in Appendix 20). No variants or CNVs in *LAMP1* were identified in any additional patients in the cohort.

Gene	RefSeq	Chr position	cDNA change	aa change	Variant consequence	No. reads WT/Mut	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
LAMP1	NM_005561.3	13:113,976,036	c.1108delG	p.(Gly370Alafs*14)	Frameshift	2/113	.	.	.	.	.	.	23
CHRNA4	NM_000744.6	20:61,981,844	c.919C>T	p.(Gly307Ser)	Missense	8/789	rs764586079	.	1/246050 4.06E-06	0	probably_ damaging (1)	damaging	32
HELZ2	NM_001037335.2	20:62,192,521	c.7060G>C	p.(Val2354Leu)	Missense	0/159	rs780967372	4/120318 3.33E-05	6/250802 2.39E-05	0	possibly_ damaging	damaging	24.7
LOXL4	NM_032211.6	10:100,016,631	c.1334G>A	p.(Arg445His)	Missense	0/112	rs775864686	.	8/282710 2.83E-05	0	benign (0.094)	damaging	24
TADA3	NM_001278270.1	3:9,828,763	c.737G>T	p.(Ala246Asp)	Missense	0/42	.	.	.	.	benign (0.016)	tolerated	23.1
MTMR14	NM_001077526.2	3:9,730,388	c.1244G>A	p.(Ser415Asn)	Missense	0/42	.	.	.	.	benign (0.184)	tolerated	18.36

Table 4-10 Rare homozygous variants identified in case F1288.

Rare homozygous variants with a MAF ≤0.0001 identified in homozygous regions using AgileMultiIdeogram software (http://dna.leeds.ac.uk/agile/AgileMultiIdeogram/) (Section 2.4.6). The gene identifier, chromosome (chr) position, amino acid (aa) change, variant consequence, dbSNP rs ID number, allele frequency & count from the gnomAD database (last accessed 12/12/2017), the number of homozygous (hom) alleles present in the ExAC (http://exac.broadinstitute.org) and gnomAD (http://gnomad.broadinstitute.org) databases and the pathogenicity prediction scores from PolyPhen-2, (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al.*, 2010), SIFT (http://sift.jcvi.org/) (Ng and Henikoff, 2003) and CADD (http://cadd.gs.washington.edu/score) (Kircher *et al.*, 2014) are presented.

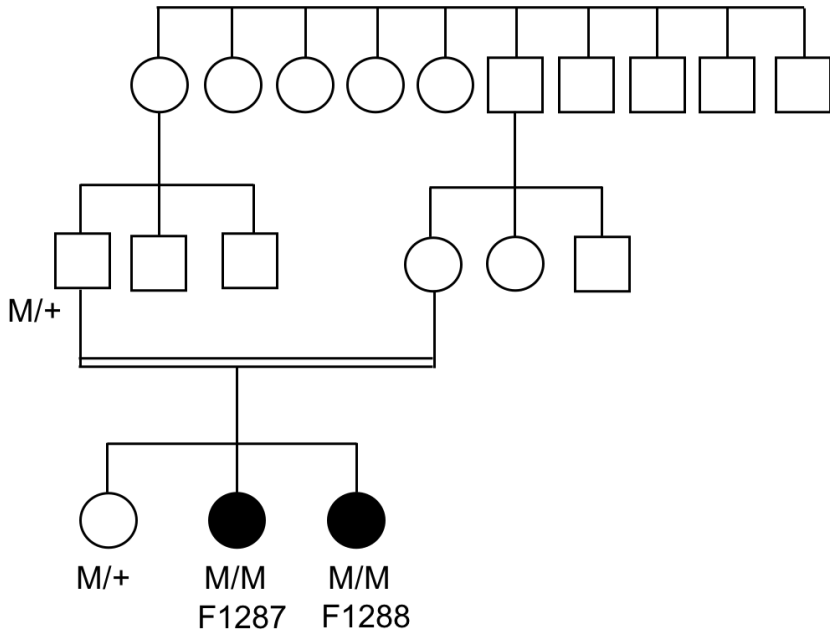


Figure 4-20 Pedigree and mutation status of patient F1288 and family.

Affected individuals F1287 and F1288 are indicated by the shaded symbols. The mutant (M) or wild-type (+) alleles are shown for each individual screened.

A protein schematic of LAMP1 was generated using the online prediction tools Protter and SMART (Figure 4-21). LAMP1 is a heavily glycosylated lysosomal membrane protein with 18 motifs for N-linked glycosylation (Figure 4-21). The 417 amino acid protein has a predicted single membrane-spanning section and 11 amino acid cytosolic tail (Figure 4-21). The frameshift variant identified in patient F1288 was located in the last exon of the gene and is predicted to escape NMD. The variant was predicted to alter the C-terminal domain of the protein, removing the transmembrane domain (Figure 4-21). The C-terminal domain is conserved down to the zebrafish.

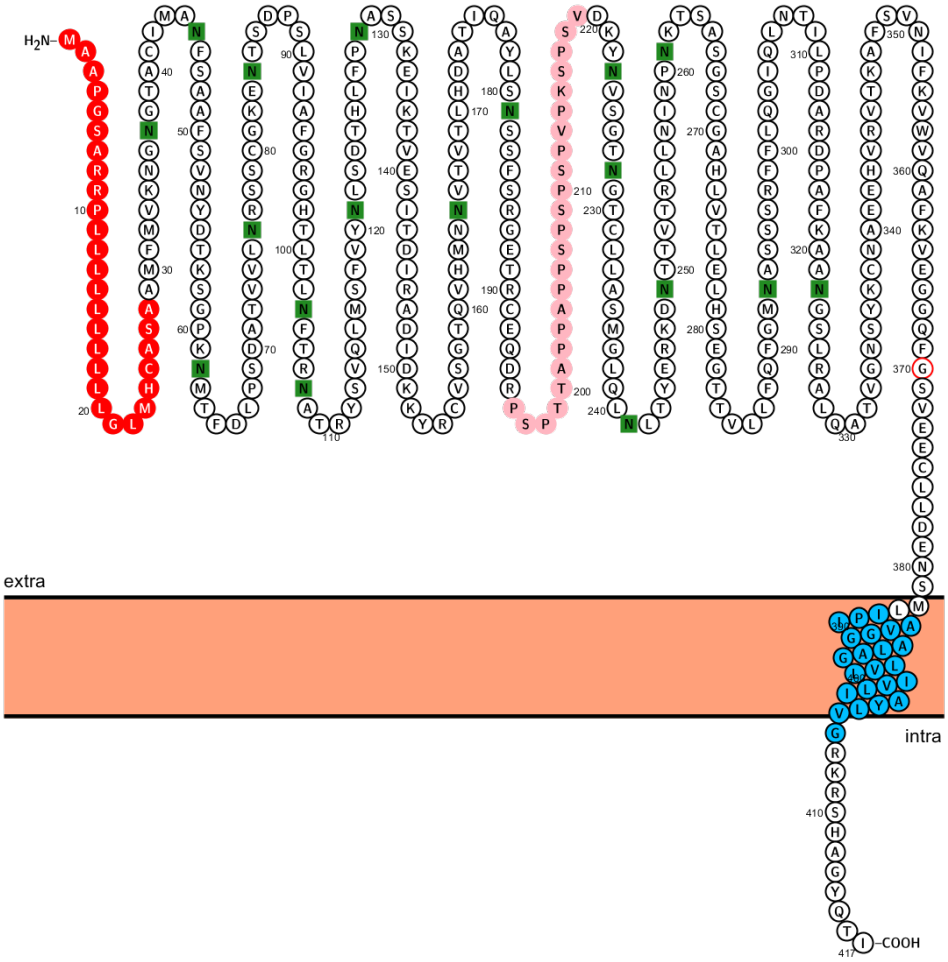


Figure 4-21 The p.(G370Afs*14) *LAMP1* variant mapped onto a protein schematic.

Schematic representation of the 417 amino acid protein structure of LAMP1 (consensus shown from SMART and Protter tools, Section 2.4.7). Unknown regions are outlined in black, the low complexity region is coloured pink (196-219aa), the signal peptide is coloured red (1-28aa) and the transmembrane domain is coloured blue (384-406aa). The 18 sites for N-linked glycosylation are coloured green. The p.(Gly370Alafs*14) variant is outlined in red.

Interrogation of LAMP1 expression revealed the protein is highly expressed in a number of tissues including the brain (The Human Protein Atlas, EMBL-EBI Expression Atlas, last accessed 12/12/2017). To explore *LAMP1* expression in the retina, *in silico* analysis was performed using the GENEVESTIGATOR platform (Section 2.4.8). HG-U133 Plus 2.0 GeneChip microarray (Affymetrix, Santa Clara,

USA) datasets of the human fetal retina were available for interrogation (Kozulin and Provis, 2009; Kozulin *et al.*, 2009b) which confirmed high expression of *LAMP1* at 19-20 weeks in the macula, nasal retina and surrounding retina (Figure 4-22).

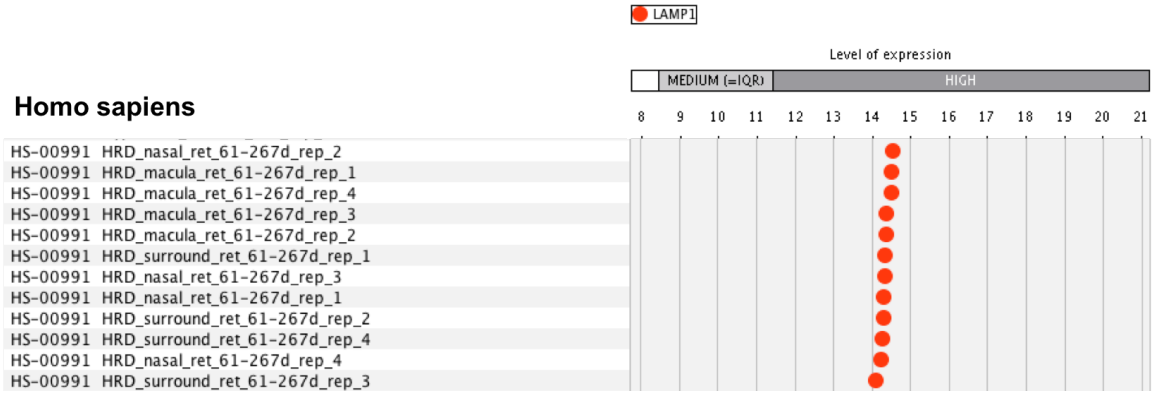


Figure 4-22 Human fetal expression of *LAMP1* in macula, nasal and surrounding retina.

GENEVESTIGATOR (Section 2.4.8) was used to investigate *LAMP1* expression in human microarray datasets (Kozulin and Provis, 2009; Kozulin *et al.*, 2009b). Fetal tissue (with no known disease) used was 19-20 weeks post gestation. *LAMP1* is highly expressed in fetal macula and both nasal and surrounding retina. Medium (inter-quartile range) and High (fourth quartile) are shown on the plot. The software states that these ranges are determined by looking at all expression values of all genes over all samples.

In the thirteenth largest homozygous region spanning 2.71Mb on chromosome 20 (Figure 4-19 and Table 4-9), a homozygous missense variant was identified in *Cholinergic Receptor, Neuronal Nicotinic, Alpha Polypeptide 4* (*CHRNA4)* (NM000744.6) (MIM 118504); c.919C>T, p.(Gly307Ser), number of reads WT/Mut: 8/789. The variant, p.(Gly307Ser), was present in dbSNP (rs764586079) and the allele was reported once in the gnomAD database in an African individual (1/246050) (Table 4-10) (last accessed 25/05/2017). In gnomAD there were no reported homozygotes (Table 4-10). The variant was predicted to be highly pathogenic

(PolyPhen-2 score 1 and SIFT score deleterious) and has a CADD score of 32 (Table 4-10). The affected amino acid was fully conserved in all available species (Figure 4-23). No variants or CNVs in *CHRNA4* were identified in any additional patients in the cohort.

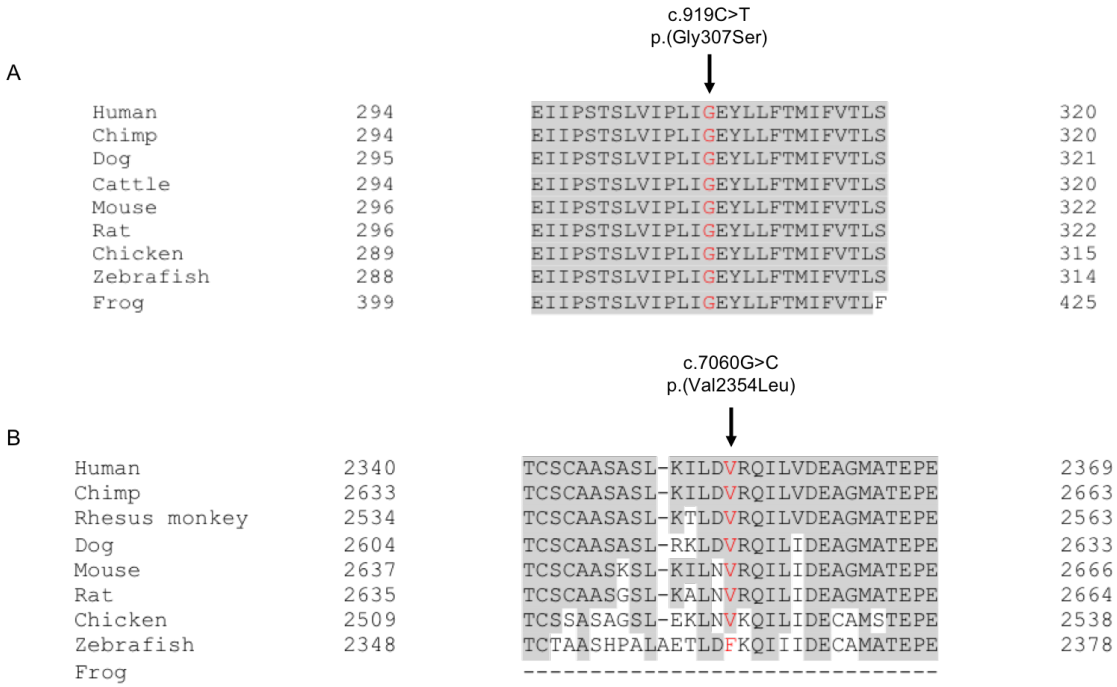


Figure 4-23 Protein sequence alignment of human CHRNA4 and HELZ2 with homologues.

Mutliple sequence alignment was calculated using HomoloGene (https://www.ncbi.nlm.nih.gov/homologene). The shaded amino acid residues indicate those that are shared with the human residue at the same position. The residues substituted in patient F1288 are coloured red. **A** CHRNA4 sequences used: Human (*Homo sapiens*) NP_000735.1; Chimpanzee (*Pan troglodytes*) NP_001029286.1; Dog (*Canis lupus familiaris*) XP_543097.3; Cattle (*Bos taurus*) NP_001179829.1; Mouse (*Mus musculus*) NP_056545.3; Rat (*Rattus norvegicus*) NP_077330.1; Zebrafish (*Danio rerio*) XP_692148.2, Frog (*Xenopus laevis*) NP_001107315.1. **B** HELZ2 sequences used: Human NP_001032412.2; Chimp XP_003317092.1; Rhesus monkey XP_002798110.1; Dog XP_005635349.1; Mouse NP_898985.2; Rat XP_230961.5; Chicken XP_004947338.1; Zebrafish XP_003198895.2; Frog XP_004920654.1.

There are medium levels of expression of *CHRNA4* in the cerebral cortex of the brain (data not shown). Autosomal dominant mutations in *CHRNA4* can cause nocturnal frontal lobe epilepsy-1 (MIM 600513) (Philips, 1995; Steinlein, 1995; Saenz, 1995; Steinlein 1997; Hirose, 1999; Hwang, 2011) or susceptibility to nicotine addiction (MIM 188890) (Feng *et al.*, 2004; Li *et al.*, 2005).

In the same homozygous region on chromosome 20 (2.71Mb), a homozygous missense variant was identified in the gene *Helicase with zinc finger 2 (HELZ2)* (MIM 611265), NM_001037335.2; c.7060G>C, p.(Val2354Leu), number of reads WT/Mut: 0/159. The variant was present in dbSNP (rs780967372) and the allele was reported six times in the gnomAD database in South Asian individuals (6/250802) (Table 4-10) (last accessed 25/05/2017). In gnomAD there were no reported homozygotes (Table 4-10). The variant was predicted to be highly pathogenic (PolyPhen-2 scored as possibly damaging and SIFT scored as damaging) and has a CADD score of 24.7 (Table 4-10). The affected amino acid was conserved in all available species down to the chicken (Figure 4-23). No variants or CNVs in *HELZ2* were identified in any additional patients in the cohort. There are low levels of *HELZ2* expression in the cerebral cortex of the brain (data not shown).

Further to analysing the regions of homozygosity in this patient, the whole dataset was interrogated without any assumptions on inheritance pattern. No further candidates were identified in this patient.

4.2.3.4 Investigating rare variants in the non-consanguineous sib pair (F1372 and F1373)

Homozygosity mapping was performed in the sibling pair using the AgileMultiIdeogram tool (Section 2.4.6), which suggested that there was no unreported or unknown consanguinity, due to the low number of autozygous regions. Shown in Table 4 11 are all rare variants (MAF 0.0001) that were identified in both sibs, and that were not present in the two control exomes. The variants shown had a CADD score >15 and were scored as pathogenic by at least one of the prediction tools, PolyPhen-2 or SIFT. The top compound heterozygous candidates were in the gene *Poly(A) binding protein cytoplasmic 3 (PABPC3)* (MIM 604680). The first variant was a frameshift, (NM_030979.2); c.977delA, p.(Phe335Leufs), number of reads 83/37 (F1372), 85/28 (F1373). The variant was predicted to be pathogenic by PolyPhen-2 and SIFT and had a CADD score of 34 (Table 4-11). The second heterozygous variant identified in *PABPC3* was a missense change; c.691A>G, p.(Lys231Glu), number of reads: 54/24 (F1372), 55/34 (F1373). The variant was scored as pathogenic by PolyPhen-2 and SIFT and had a CADD score of 24.6 (Table 4-11). Segregation analysis was not performed on these candidates. Further investigation of this gene is required, however at this stage it was not prioritised as a top candidate when compared to the findings in other patients in the cohort. The search was widened to dominant mutations but none were prioritised as top candidate genes.

4.2.4 Candidates of note from the remaining WES data

A selection of top candidate(s) from the remaining cohort (n=7), are listed in Table 4-12. These variants were selected as they were rare (cut off <0.01%) or unreported and were not present in the two control exomes. The variants shown had a CADD score >15 and were scored as pathogenic by at least one of the prediction tools, PolyPhen-2 or SIFT.

Gene	RefSeq	Chr position	cDNA change	aa change	Variant consequence	No. reads WT/Mut (F1372, F1373)	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
PABPC3	**NM_030979.2**	**13:25,671,333**	**c.977delA**	**p.(Phe335Leufs)**	**Frameshift**	**83/37, 85/28**	**rs371570689**	**.**	**.**	**.**	**probably_damaging**	**damaging**	**34**
	NM_030979.2	**13:25,671,027**	**c.691A>G**	**p.(Lys231Glu)**	**Missense**	**54/24, 55/34**	**rs78826513**	**239/120102 1.99E-03**	**.**	**0**	**probably_damaging**	**damaging**	**24.6**
GXYLT1	NM_001099650.1	12:42,538,334	c.115G>T	p.(Gly39Trp)	Missense	5/5, 7/6	rs181558534	.	1/77388 1.29E-05	0	possibly_damaging	tolerated	23.2
TRBV7-7	ENSG00000253291	7:142,119,881	c.285delA	p.(Leu96Ter)	Nonsense	197/16, 198/22	.	.	.	.	.	.	22.8
MUC16	NM_024690.2	19:9,000,169	c.40588G>A	p.(Gly13530Ser)	Missense	113/28, 130/21	.	.	26/26324 9.87E-04	0	probably_damaging	.	22.4

Table 4-11 Rare heterozygous variants identified in the sib pair F1372 and F1373.

Summary of frequency and predicted pathogenicity scores for the rare variants identified in the sib pair (F1372 and F1373). The gene identifier, chromosome (chr) position, amino acid (aa) change, variant consequence, dbSNP rs ID number, allele frequency & count from the gnomAD database (last accessed 12/12/2017), the number of homozygous (hom) alleles present in the ExAC (http://exac.broadinstitute.org) and gnomAD (http://gnomad.broadinstitute.org) databases and the pathogenicity prediction scores from PolyPhen-2, (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al.*, 2010), SIFT (http://sift.jcvi.org/) (Ng and Henikoff, 2003) and CADD (http://cadd.gs.washington.edu/score) (Kircher *et al.*, 2014) are presented. Highlighted are the candidate compound heterozygous variants.

Patient	Gene	RefSeq	Chr position	cDNA change	aa change	Variant consequence	No. reads WT/Mut	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
F1352	*BOD1L1*	NM_148894.2	4:13,602,111	c.6413T>A	p.(Lys2138Ile)	Missense	157/133	.	.	.	.	probably_ damaging (1)	damaging	28
		NM_148894.2	4:13,583,936	c.8519-3T>G	.	olice accept	49/71	.	.	.	.	.	.	23.7
F1369	*FRMD4A*	NM_018027.4	10:13,702,491	c.1723G>A	p.(Arg575Trp)	Missense	69/88	rs376696301	4/12148 3.3E-05	11/276800 3.97E-05	0	probably_ damaging (0.99)	damaging	34
		NM_018027.4	10:13,696,479	c.2987G>A	p.(Pro996Leu)	Missense	46/54	rs199934183	10/120884 8.27E-05	10/277060 3.62E-05	0	probably_ damaging (1)	damaging	32
F1371	*DNA2*	N_001080449.2	10:70,206,143	c.967C>T	p.(Glu323Lys)	Missense	32/35	.	.	1/30968 3.23E-05	0	possibly_ damaging (0.83)	damaging	33
		N_001080449.2	1:70,209,870	c.854T>C	p.(His285Arg)	Missense	77/71	.	.	1/30866 3.24E-05	0	probably_ damaging (0.98)	tolerated	23.3

Table 4-12 Rare compound heterozygous variants identified in the remaining foveal hypoplasia cohort.

Summary of frequency and predicted pathogenicity scores for the rare variants identified in the remaining foveal hypoplasia cohort. The gene identifier, chromosome (chr) position, amino acid (aa) change, variant consequence, dbSNP rs ID number, allele frequency & count from the gnomAD database (last accessed 12/12/2017), the number of homozygous (hom) alleles present in the ExAC (http://exac.broadinstitute.org) and gnomAD (http://gnomad.broadinstitute.org) databases and the pathogenicity prediction scores from PolyPhen-2, (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al.*, 2010), SIFT (http://sift.jcvi.org/) (Ng and Henikoff, 2003) and CADD (http://cadd.gs.washington.edu/score) (Kircher *et al.*, 2014) are presented.

4.2.5 Investigating rare heterozygous variants in the patient cohort

Although the inferred mode of inheritance was autosomal recessive, heterozygous variants were also investigated to see if there was a shared gene in the cohort (Table 4-13). The variants shown in Table 4 13 were selected as they were rare (MAF <0.00001) or unreported and were not present in the two control exomes. The data was filtered to retain variants with a CADD score >15 and those scored as pathogenic by at least one of the prediction tools, PolyPhen-2 or SIFT. This analysis was separate to the known RetNet/Albinism gene investigation described previously. Two novel heterozygous variants were identified in *Midasin* (*MDN1*) in patients F1369 and F1352 (Table 4-13). In patient F1369, the missense change, NM_014611.2; c.2461C>A, p.(Gly821Cys), number of reads WT/Mut: 15/5, was predicted to be pathogenic by both PolyPhen-2 (score 1) and SIFT (damaging) and had a CADD score of 34. The heterozygous change identified in patient F1352, NM_014611.2; c.8555G>A, p.(Ala2852Val), number of reads WT/Mut: 74/62, was predicted to be pathogenic by both PolyPhen-2 (score 0.92) and SIFT (damaging) and had a CADD score of 27.8 (Table 4-13). This is one of the largest genes in the genome, which might account for the two hits.

Each individual was also analysed for rare heterozygous variants on a case-by-case basis. Although longer lists of variants were generated in this analysis, two interesting heterozygous candidate variants (identified in patients F1405 and F1352) that fulfilled the criteria described above, are described below. These variants were included following literature investigation using PubMed and GeneCards, which revealed that the genes are involved in early eye development.

Patient	Gene	RefSeq	Chr position	cDNA change	aa change	Variant consequence	No. reads WT/Mut	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
F1369	*MDN1*	NM_014611.2	6:90,468,679	c.2461C>A	p.(Gly821Cys)	Missense	15/5	.	.	.	.	probably_ damaging (1)	damaging	34
F1352		NM_014611.2	6:90,410,448	c.8555G>A	p.(Ala2852Val)	Missense	74/62	.	.	.	.	possibly_ damaging (-.92)	damaging	27.8

Table 4-13 Rare heterozygous variants present in at least two patients in the cohort.

Summary of frequency and predicted pathogenicity scores for the rare heterozygous variants identified in the patient cohort. The gene identifier, chromosome (chr) position, amino acid (aa) change, variant consequence, dbSNP rs ID number, allele frequency & count from the gnomAD database (last accessed 12/12/2017), the number of homozygous (hom) alleles present in the ExAC (http://exac.broadinstitute.org) and gnomAD (http://gnomad.broadinstitute.org) databases and the pathogenicity prediction scores from PolyPhen-2, (http://genetics.bwh.harvard.edu/pph2/) (Adzhubei *et al.*, 2010), SIFT (http://sift.jcvi.org/) (Ng and Henikoff, 2003) and CADD (http://cadd.gs.washington.edu/score) (Kircher *et al.*, 2014) are presented.

4.2.5.1.1 A prioritised heterozygous candidate variant in *VSX2* identified in case F1405

Case F1405 was a Caucasian male who was diagnosed with foveal hypoplasia and posterior embryotoxon but had no decussation defects.

A heterozygous nonsense variant was identified in the gene *Visual System Homeobox 2* (*VSX2)* (MIM 142993) (NM_182894.2); c.892C>T, p.(Gln298Ter), number of reads WT/Mut: 37/27 (Appendix 9). The variant was not present in publicly available databases (last accessed 12/12/2017) (Appendix 9). The nonsense variant was located in the last exon of the gene (exon 5), so may not be targeted for NMD. Analysis of other loss of function LoF variants on gnomAD revealed 8 variants in each of the 5 coding exons of the gene, of which 2 are likely to escape NMD (last accessed 12/12/2017). A low complexity region would be deleted if the protein was truncated. *VSX2* is highly expressed in the retina and is a well-known key transcriptional regulator in eye organogenesis. Recessive mutations have been reported in patients with isolated microphthalmia (MIM 610093), or microphthalmia with coloboma (MIM 610092). Analysis of remaining patient exome data did not reveal any further variants in *VSX2*. There were no CNVs identified in *VSX2* in the ExomeDepth analysis. The variant was confirmed using primers listed in Appendix 19.

4.2.5.1.2 A heterozygous candidate variant in *EPHA2* identified in case F1352

In case F1352, an East Asian male with foveal hypoplasia, a novel heterozygous missense change was identified in the gene *Ephrin type-A Receptor 2 Precursor*

(*EPHA2)* (MIM 176946) (NM_004431.4); c.1148T>G, p.(Val383Gly), number of reads WT/Mut: 57/62 (Appendix 9). The variant was predicted to be pathogenic by SIFT and the CADD score of deleteriousness was 28.6 but was scored as 'benign' by PolyPhen-2 (Appendix 9). The variant was present in the dbSNP database (rs7510403034) and was reported once in the heterozygous state in an East Asian individual in the gnomAD database (1/245434 4.07e-6) (Appendix 9). Conservation analysis revealed that the valine residue was not conserved in the chicken or frog (Figure 4-24).

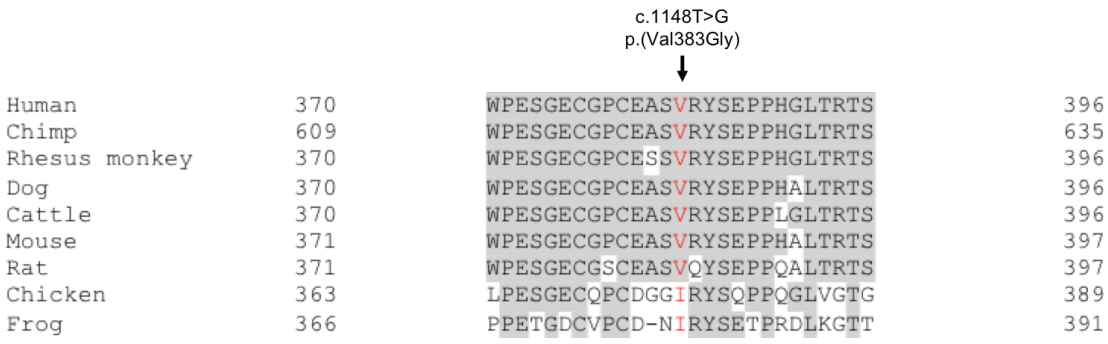


Figure 4-24 Multiple sequence alignment of human EPHA2 and homologues.

The shaded amino acid residues indicate those that are shared with the human residue at the same position. The residues substituted in the patient is coloured red. Patient F1352; c.1148T>G, p.(Val383Gly). Sequences of EPHA2 used: Human NP_004422.2; Chimp XP_513064.4; Rhesus monkey NP_001035768.1; Dog XP_544546.2; Cattle NP_001192660.1; Mouse NP_034269.2; Rat NP_001102447.1; Chicken XP_001234814.2; Frog NP_001006765.1.

AD mutations in *EPHA2,* a receptor tyrosine kinase, have been identified in patients with cataracts (type 6) (MIM 116600) (Ionides *et al.*, 1997; McKay *et al.*, 2005; Shiels *et al.*, 2008; Zhang *et al.*, 2009). A high number of the reported mutations are located at the C-terminal end of the EPHA2 protein and the putative change identified in

patient F1352 is located in exon 5/17. There were no CNVs identified in *EPHA2* in the ExomeDepth analysis.

4.3 Discussion

In this chapter, 12 patients diagnosed with foveal hypoplasia in a pattern consistent with autosomal recessive inheritance were analysed by WES in order to identify the gene(s) and mutation(s) involved.

4.3.1 WES technology and filtering strategy for variant identification

WES is a well justified, efficient strategy to capture the highly interpretable coding region of the genome in the search for alleles underlying rare Mendelian disease, at an affordable cost (Ng *et al.*, 2010). The Agilent SureSelect All Exon XT exome capture kit with Illumina sequencing technology was chosen for this study as it was widely used with success in the Vision Research Group at The University of Leeds (El-Asrag *et al.*, 2015; Arno *et al.*, 2016; Fiorentino *et al.*, 2017; Fiorentino *et al.*, 2018; Panagiotou *et al.*, 2017). In comparison studies, Agilent's SureSelect All Exon XT platform has been found to have the highest target enrichment efficiency and detect a greater number of total variants with increased sensitivity towards detection of SNVs and small indels, compared to other kits, including Roche/Nimblegen's SeqCap EZ Exome Library v2.0 kit, Illumina's Nextera Rapid Capture Exome (v1.2) kit and Agilent's SureSelectQXT kit (Clark *et al.*, 2011; Shigemizu *et al.*, 2015). It is thought that the high target enrichment efficiency of the Agilent kit results from the

relatively long baits and/ or RNA methodology (Clark *et al.*, 2011). In the current study, on average, 97.7% of genomic targets were covered at a sequencing depth of ≥10X and 93.9% were covered at a depth of ≥20X. This is similar to figures generated in kit comparison studies. For example, Agilent XT's average coverage has been reported to be 96% at ≥10X sequencing depth (from a randomly selected set of 75M sequencing reads) in a study comparing Roche/NimbleGen's SeqCap EZ Human Exome Library v3.0, Illumina's Nextera Rapid Capture Exome (v1.2), Agilent's SureSelect XT Human All Exon v5 and Agilent's SureSelect QXT (Shigemizu *et al.*, 2015). Comparison of the average depth of coverage seen in this study to similar assays using the same enrichment and sequencing chemistry also gave similar coverage results (Mackay *et al.*, 2014; Koboldt *et al.*, 2014; Guo *et al.*, 2017).

The bioinformatics pipeline used in this study was optimised by the Leeds Vision Research group for the analysis of WES samples run on the Illumina platform. The analysis pipeline was chosen in this study due to the high variant identification success rate in the group (El-Asrag *et al.*, 2015; Panagiotou *et al.*, 2017). BWA was used as the aligner of choice over alternatives such as Bowtie-2 and Novoalign, as it has been shown to correctly align a higher number of reads compared to Novoalign (Li *et al.*, 2009; Li, 2013). Although Novoalign has been identified as the most accurate aligner, as conceded by Heng Li's paper, BWA has improved sensitivity (Highnam *et al.*, 2015). To compare all three aligners, both Novoalign and BWA significantly outperform Bowtie2 in alignment accuracy when a relatively small or large number of indels are present and they also outperform Bowtie2 in the identification of both novel and common variants (Li *et al.*, 2009; Li, 2013). BWA also

outputs a data format that can be input into Picard tools without the need for additional alterations (Auwera *et al.*, 2013). For variant calling, the GATK was used to call indels and SNVs in the GVCF format and was the preferred choice of variant caller for this study due to its high sensitivity and lower calling rate of false positives when compared to other variant callers including SAMtools, Atlas2 and glftools (Liu *et al.*, 2013). The vcfhacks package (Perl) was used for variant filtering and annotation to aid the identification of disease-causing variants. The Perl programs, which include a number of scripts, have been optimised on data generated using the GATK variant callers. The programs were originally developed in-house by Dr David Parry (now at the University of Edinburgh) and have been used to successfully identify a number of genes mutated in Mendelian disease (Parry *et al.*, 2012; Acevedo *et al.*, 2015; Logan *et al.*, 2015).

The variant in *FOXD1* in case F1374 was initially missed in the analysis. The variant was only identified after comparing pipelines and datasets with a collaborator at the start of this study (UKIRDC collaboration, Dr Nikolas Pontikos, UCL). After troubleshooting the alignment pipeline, the issue was circumvented by removing the Agilent SureSelect XT V5/V6 BED file from GATK's HaplotypeCaller command, when calling the variants. The BED file is enriched for exonic regions that are baited for capture and is specific to the version of the kit used. The calling of the SNPs and indels was therefore not restricted to the specific genomic target regions corresponding to the Agilent XT capture kit used to generate the exome libraries. However, this resulted in the generation of longer candidate variant lists.

Further to this, investigation of the % coverage of the known IRD/albinism genes listed in Appendix 6 revealed that the Agilent V5/V6 probes do not capture each of the 270 genes at 100%. It is therefore evident that capture kit chemistry may introduce measurement bias, with potential missed mutations in genes that are poorly covered, or genes that have no coverage at all, such as *FOXD1*.

In order to prioritise pathogenic variants, three main parameters were employed: (1) the MAF in publicly available control databases of variation including dbSNP, ExAC and gnomAD (when it was released at a later stage of the project); (2) *in silico* pathogenicity prediction tools including PolyPhen-2, SIFT and CADD; and (3) conservation of the residue among homologues. Genes of interest were also examined globally in the ExAC and gnomAD databases to see how many LoF alleles were found in a heterozygous and homozygous state to deduce if there was genetic redundancy. Further to these parameters, literature investigations were used to understand any known association of the candidate gene with the retina, specifically if the candidate gene played a role in early development or organogenesis, such as *VSX2* and *EPHA2*.

At present, the prevalence or frequency data of foveal hypoplasia without albinism is not known, although it is clearly very rare due to the lack of reports in the literature. A statistical framework for frequency-based filtering developed by James Ware and Daniel MacArthur, lead developers of the ExAC and gnomAD databases, has sought to address variant interpretation for Mendelian disease by giving an estimated maximum tolerated allelic count for a variant in a reference database, such as ExAC

or gnomAD (http://cardiodb.org/allelefrequencyapp/) (Whiffin *et al.*, 2017). The team recognised that investigators can afford to be much more stringent with their filtering than many researchers currently are being (Whiffin *et al.*, 2017). In the current study, a more conservative approach was initially applied in the first line of exome data filtering when a BED file of known IRD and albinism genes was used to retain potentially causative variants. In this primary analysis, a MAF cut-off value of 0.0001 (0.01%) was applied to detect heterozygous and biallelic variants. Use of such a conservative cut-off value when analysing all of the exome in the case-by-case biallelic and heterozygous analyses would have generated long lists of variants that would have been difficult to prioritise. Therefore, a more stringent filtering frequency was applied in order to reduce the number of variants analysed downstream, using the tool developed by James Ware as a guide (Whiffin *et al.*, 2017). To summarise the use of the tool, the calculation is based on several parameters including inheritance, disease prevalence, maximum allelic contribution, penetrance and the number of screened reference alleles. In the application of the tool in this study for an estimated maximum credible population allele frequency, biallelic inheritance was set, penetrance was set at 90% (most likely 100% based on other *SLC38A8* families, but 90% was chosen here as a precaution), only one mutation has been found more than once (allelic heterogeneity was set at 0.1), and given that the disease is rare, a prevalence of 1/1,000,000 (people, not chromosomes), was applied. The use of the tool with these settings gave an estimated (maximum) possible MAF of 0.0001 (0.01%). This was the maximum MAF cut-off used for the analysis of biallelic variants. No candidate variants reported in the tables exceeded this predicted maximum allele count in the reference databases. Comparative mutation identification studies in recessive disease have employed cut off values of ≤0.1%

(Kobayashi et al., 2017). There have even been recommendations for setting the maximum cut off at 1% (Bamshad et al., 2011), despite more recent tools suggesting that more stringent thresholds (such as ≤0.01%), as applied in this study, are justified without potentially discarding true pathogenic variants (Whiffin et al., 2017).

The combined foveal hypoplasia cohort analysis sought to identify a clear candidate gene in which more than two patients harboured rare variants. This strategy was prioritised as it can be difficult to successfully narrow down candidate genes from long lists generated for single patients. As foveal hypoplasia without albinism is currently homogeneous, it was assumed that a common gene might be identified in the cohort. Although this approach was not successful in this study, similar studies have identified a common gene from a cohort of rare cases (Panagiotou *et al.*, 2017; Fiorentino *et al.*, 2018).

In the absence of a common gene in the foveal hypoplasia cohort, all cases were analysed on a case-by-case basis. Specifically, attention was paid to the consanguineous patients and the sib pair, as it was possible to reduce the candidate gene lists for these cases. The rare variants shared in the sib pair (listed in Section 4.2.3.4) were not in particularly interesting candidate genes. However, with the bioinformatics pipeline used in this study and the strict filtering criteria applied, they were the top variants identified. Homozygosity was analysed in all cases, in case patients had unreported or unknown consanguinity (Littink *et al.*, 2012), but no additional consanguineous cases were found. Autozygosity mapping is a valuable tool when investigating autosomal recessive conditions, using regions identical by

descent (Alkuraya, 2013; Carr *et al.*, 2013). The approach has been used as a powerful method in IRD investigations (Littink *et al.*, 2012; Collin *et al.*, 2011). The practice is more commonly applied to small inbred families, whereby candidate disease loci can be pinpointed from regions autozygous in all affected individuals. However, it has also been used in families with little or no consanguinity. An example is the study of 186 different Dutch families to identify the genetic defect underlying AR RP (Collin *et al.*, 2011). In the study by Collin et al, ~94% of the affected individuals harboured large homozygous sequences, showing that homozygosity mapping is a powerful tool. In the current study, homozygosity mapping was used to help narrow-down the candidate disease-gene harbouring regions in cases from inbred families. SNP microarray genotyping-based homozygosity mapping is the traditional method and has been very successful to map and identify disease genes (den Hollander *et al.*, 2007b; Baere *et al.*, 2011). In the current study, WES data was used to infer regions of homozygosity (ROHs) in the consanguineous cases but this method has limitations. Exome data is enriched for SNPs in the coding regions only and will therefore show gaps in the data. Nevertheless, there are many examples of the successful application of WES SNP analysis in ophthalmic research (Van Schil *et al.*, 2015; Pierrache *et al.*, 2017). The strategy was successful for case F1335, as it led to the identification of further *SLC38A8* variation, to be discussed in more detail below. Homozygosity mapping using WES data in the current study identified a common region in the two consanguineous cases F1374 and F1288. The region on chromosome 10 overlapped in both patient (10:81,286,427-102,056,166, 20.7Mb). However, no shared candidate gene was identified in this shared region of homozygosity.

Although a definite novel gene was not identified using this methodology, this is because foveal hypoplasia is more heterogeneous than anticipated and not due to problems with the strategy applied. The work has highlighted a number of exciting novel candidate genes mutated in foveal hypoplasia.

4.3.2 *PAX6* mutations in patients with foveal hypoplasia

A BED file was generated containing all known IRD or albinism genes using RetNet and the literature in order to analyse a wide variety of retinal genes and not just those with a similar phenotype. This strategy was applied due to the growing number of examples of "expanding phenotypes" reported in the literature. Case F1351 was found to harbour a heterozygous variant in *PAX6* (c.275T>C, p.(Val92Ala)). Subsequent screening of *PAX6* in the remaining *SLC38A8* mutation-negative cohort revealed a *PAX6* variant in case F1378 (c.278C<A, p.(Ala79Glu)). Patient samples were referred to this study with a minimum phenotype of foveal hypoplasia (with no pigmentation defects). It is in the clinician's best interest to send the samples for screening for every possible diagnosis to help their patient, therefore such findings are not unexpected. Cases were referred to this study with a minimum phenotype of foveal hypoplasia and neither F1351 nor F1378 were tested for nystagmus or misrouting defects by the referring clinician.

Therefore, further clinical investigations need to be performed on cases F1351 and F1378, to resolve if they have chiasmal decussation defects, or the more likely scenario, that they have isolated foveal hypoplasia. Optic nerve decussation defects have been investigated in cases with *PAX6* mutations, but visually evoked potentials

recorded as “normal” (Schroeder *et al.*, 2003). Lack of detailed chiasmal routing phenotyping is evident in the literature, including a report of a *SLC38A8* mutation (c.95T>G, p.(Ile32Ser) variant segregating in 3 families of Israeli Jewish Indian ancestry with reported isolated foveal hypoplasia (Perez *et al.*, 2014) and the report of *PAX6* variants in patients with isolated foveal hypoplasia (Azuma *et al.*, 1996). There is a clear need to formalise the assessment of patients with foveal hypoplasia. Very few studies have used strict criteria to fully exclude a diagnosis of foveal hypoplasia with optic nerve misrouting, with visual evoked potential examination rarely performed in patient cohorts (Azuma *et al.*, 1996; Perez *et al.*, 2014; Curran and Robb, 1976; Oliver *et al.*, 1987). Instead, tests for nystagmus (Curran and Robb, 1976; Perez *et al.*, 2014; Azuma *et al.*, 1996; O'Donnell and Pappas, 1982; Giocanti-Auregan *et al.*, 2014), optic nerve size (Curran and Robb, 1976; O'Donnell and Pappas, 1982; Recchia, Carvalho-Recchia and Trese, 2002; Giocanti-Auregan *et al.*, 2014), iris transillumination (Curran and Robb, 1976; Perez *et al.*, 2014; Giocanti-Auregan *et al.*, 2014) and ocular and cutaneous observations of pigmentation (Perez *et al.*, 2014; O'Donnell and Pappas, 1982; Recchia, Carvalho-Recchia and Trese, 2002), are the basic tests performed, with the latter three used to exclude albinism. Only two reports have excluded chiasmal misrouting in patients with isolated foveal hypoplasia through the use of visual evoked potential recordings (Oliver *et al.*, 1987; Schroeder *et al.*, 2003).

The p.(Ala79Glu) variant identified in patient F1378 has already been published in 1999, in a patient diagnosed with mild aniridia (stromal hypoplasia of the iris with eccentric pupil), along with mild nystagmus and early cataracts at the age of 21 (Grønskov *et al.*, 1999). In contrast to this published case, case F1378 had no

diagnosis of cataracts or nystagmus and visual evoked potential analysis was not performed. DNA was available from the mother and three uncles of F1378, who were all unaffected. It is not possible to conclude if the detected mutation in patient F1378 is *de novo,* although more than one third of mutations reported in *PAX6* are *de novo* (The *PAX6* Allelic Variant Database, LOVD http://pax6.hgu.mrc.ac.uk/). It is unlikely that the heterozygous variant was inherited from the deceased father of F1378, as he had no recorded visual disease phenotype and there is high penetrance but variable expression of *PAX6* mutations.

Genotype-phenotype correlation studies of *PAX6* have revealed that missense mutations are rare in classical aniridia (often associated with deletions or premature termination codons), but are more common in milder, non-aniridia phenotypes, including isolated foveal hypoplasia (Hingorani *et al.*, 2009; Tzoulaki, White and Hanson, 2005; Yokoi *et al.*, 2016; Hanson *et al.*, 1999; Azuma *et al.*, 2003). In 2 independent cases of isolated foveal hypoplasia, a recurrent *PAX6* missense mutation; c.382C>T, p.(Arg128Cys) has been identified (Azuma *et al.*, 1996; van Heyningen and Williamson, 2002). Looking at the location of missense mutations in the *PAX6* protein in isolated foveal hypoplasia cases, the R128C mutation, alongside 78 and 79 identified in patients F1351 and F1378, are located in the paired domain (PD) (exons 5, 5a, 6 and 7) (Figure 4-25). This region is well conserved among PAX6 homologs. X-ray crystallography assessment of the *Drosophila* PD has shown that it has a bipartite structure with a highly conserved N- (AAs 4-75) and variable C-terminal region (AAs 80-131), with residues of the N-terminal region making sequence-specific contact with the DNA in a helix-turn-helix (HTH) conformation (Czerny, Schaffner and Busslinger, 1993; Epstein *et al.*, 1994a; Tang, Chao and

Saunders, 1997; Xu *et al.*, 1995). The variants identified in F1351 and F1378 (AAs 78 and 79) are located outside of the 4th alpha helix where the C-terminal region starts and both residues contact the DNA backbone (Figure 4-25).

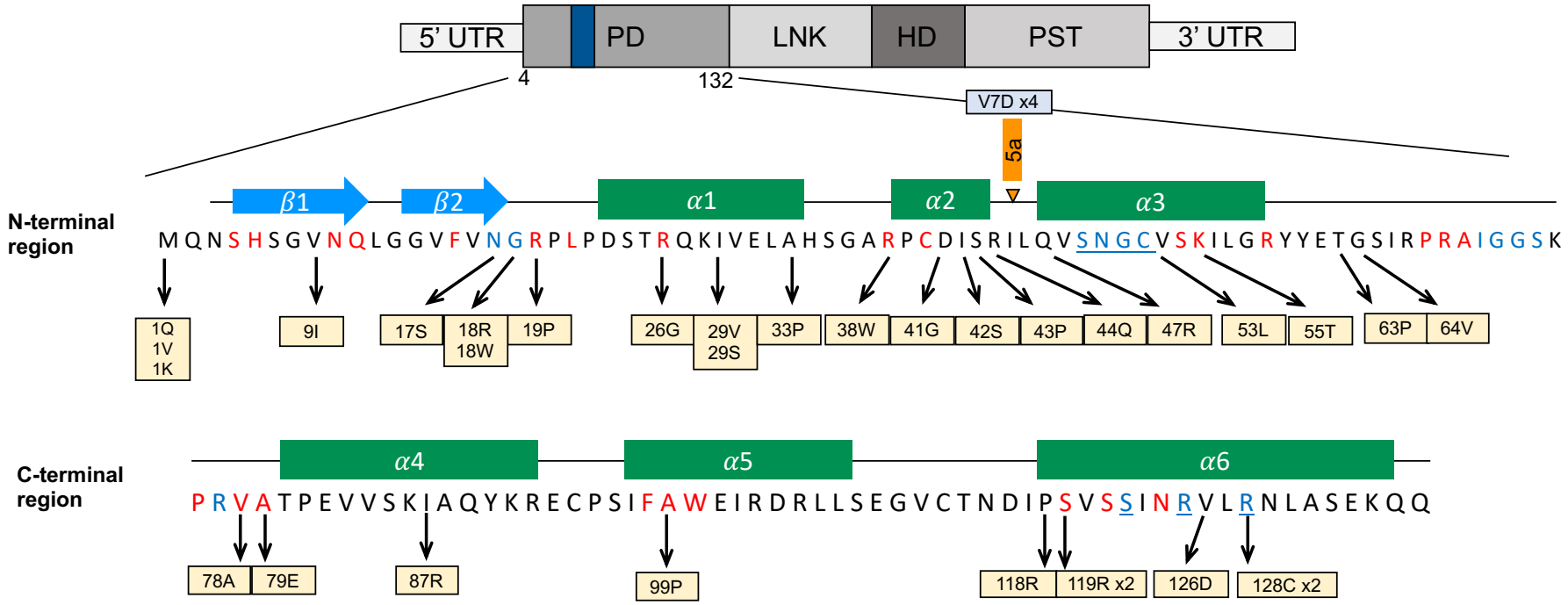


Figure 4-25 Reported mutations in the Paired Domain of PAX6.

Missense mutations (yellow) in the paired domain (PD) as reported in the PAX6 LOVD are shown. The PAX6 protein consists of the PD including the navy blue alternatively spliced exon 5a; the homeodomain (HD), separated from the PRD by a linker region; and the proline, serine and threonine-rich transactivation domain (PST). The expanded paired domain includes a1–a3, denoting the N-terminal subdomain alpha helices, and a4–a6, C-terminal subdomain alpha helices. The amino acid sequence is shown, with residues in red contacting the DNA backbone and those in blue contacting the minor groove or the major groove (underlined). Variants identified in the current study are as follows: p.(V78A) in patient F1351 and p.(A79E) in patient F1378. Image adapted with permission from Copyright Clearance Centre on behalf of Oxford University Press: Human Molecular Genetics (van Heyningen and Williamson, 2002).

Functional analysis of PD missense mutations has shown that some alleles are hypomorphic, such as the R26G in the N-terminal of the PD in a case of Peter's anomaly. Others are described as haplo-insufficient, such as the I87R in the C-terminal region of the PD reported in a case of aniridia (Tang, Chao and Saunders, 1997; Hanson *et al.*, 1999). The difference between these two missense mutations is hypothesised to be due to the R26 residue being in contact with the DNA backbone, perhaps similar to the two missense changes identified in F1351 and F1378 in this study (Tang, Chao and Saunders, 1997; Hanson *et al.*, 1999).

Both mutations in the coding region and DRR of *PAX6* have been reported to cause the same phenotype of aniridia, and even individuals with gross deletions in *PAX6* do not show a difference in phenotypic severity (Vasilyeva *et al.*, 2017). However, there are other key regulatory genes involved in development that display pleiotropic functions. An example of a pleiotropic gene is *SHH,* whereby mutations in a SIX3 binding site in a *SHH* enhancer cause holoprosencephaly, but pre-axial polydactyly is caused by mutations in a limb-specific enhancer located approximately 800kb away from *SHH* (Lettice *et al.*, 2002; Lettice *et al.*, 2008). Therefore, fifteen *cis*-acting elements of the *cis*-DRR that have been associated with the aniridia phenotype were screened in the patient cohort to look for potential mutations (Plaza *et al.*, 1995; Kammandel *et al.*, 1999; Xu *et al.*, 1999; Williams *et al.*, 1998; Kleinjan *et al.*, 2001; Griffin *et al.*, 2002; Kleinjan *et al.*, 2006; Bhatia *et al.*, 2013; Ravi *et al.*, 2013; Bhatia *et al.*, 2014). This revealed a number of variants unlikely to be the cause of disease due to their high allele frequencies in the control databases of variation.

4.3.3 Searching for novel foveal hypoplasia genes

4.3.3.1 The prioritisation *of FOXD1* as a candidate gene for foveal hypoplasia

WES data analysis in combination with homozygosity mapping prioritised *FOXD1,* after a novel homozygous missense variant (c.581C>T; p.(Gly194Asp)) was identified in patient F1374. FOXD1 was highlighted as the best candidate foveal hypoplasia gene from the large number of candidates because *FOXD1* is expressed in the ventral diencephalon and temporal retina in the CNS and is one of the earliest cues in the development of the visual system. FOXD1, alongside FOXG1 (also named *Brain Factor 1 (BF-1*)), marks the distinction between the nasal and temporal (N-T) retina, respectively (Hatini, Tao and Lai, 1994). FOXD1 expression was noted in microarray data of fetal macula and surrounding retina (Figure 4 14). The data suggests that a pattern of refined spatial dosage of FOXD1 may be important in macular development in humans. It was hypothesised that perturbations in the gene might therefore explain the foveal hypoplasia phenotype seen in foveal hypoplasia without albinism patients alongside deformities in chiasmal routing, during early eye development.

The Forkhead box gene, *FOXD1,* previously known as *Brain Factor 2 (BF-2)*, is characterised by its winged helix/forkhead DNA binding domain, evolutionarily conserved across the FOX family (Clark *et al.*, 1993; Lai *et al.*, 1993). FOXD1 is known to play an important role in cell cycle progression and cell proliferation (Jackson *et al.*, 2010; Koga *et al.*, 2014). The interaction of the FOXD1 and FOXG1 forkhead transcription factors has been deduced in animal models using a combination of loss- and gain-of-function experiments of *FOXG1* in the mouse

(Marcus *et al.*, 1999; Huh *et al.*, 1999), chick (Yuasa *et al.*, 1996), frog (Bourguignon, Li and Papalopulu, 1998; Hardcastle and Papalopulu, 2000) and zebrafish (Picker *et al.*, 2009). FOXG1 defines the nasal character of the retina and *FOXD1* is complimentarily expressed in the ventro-temporal (VT) retina. Here, ipsilateral projecting axons arise, fulfilling the partial decussation of RGC axons seen in mammals, but not in the chick (Herrera *et al.*, 2004). Nevertheless, in the chicken, misexpression of FOXG1 and FOXD1, also known as *chick brain factor 1 and 2* (*Cbf-1* and *Cbf-2*), causes abnormal regional specificity of axon projection along the naso-temporal (N-T) axis in the tectum (Yuasa *et al.*, 1996; Takahashi *et al.*, 2003). It is the misexpression of chick FOXD1 that leads to disrupted expression patterns of asymmetrical downstream effectors in the temporal retina, such as the repression of chick FOXD1*,* SOHO1, GH6 and ephrinA5, but the induction of *EphA3* in the retina (Takahashi *et al.*, 2003; Takahashi *et al.*, 2009). In the mouse model, the molecules vital for ipsilateral RGC axon projection, including *EphB1* and *Zic2*, are downregulated in *Foxd1*-deficient embryos, and *Foxg1* has been shown to expand from its nasal retinal domain into the V-T region (Herrera *et al.*, 2004). This change to the regionalisation of the ventral diencephalon is an additive effect causing improper formation of the optic chiasm, a cardinal phenotypic feature seen in patients with FHONDA.

Interestingly, in mice, expression of *Foxd1* and *Foxg1* in the optic vesicles is lost in *PAX6* null mutants (*PAX6*$^{-/-}$) (Bäumer *et al.*, 2002). This suggests the importance of *PAX6* in the initiation of N-T determinants.

When assessing the functional characterisation of a novel gene it is more accurate to do so *in vivo*. However, as only one variant had been identified in *FOXD1* in this study, a luciferase reporter assay was used, similar to the experiment applied by Laissue *et al*. This research associated variants in *FOXD1* with adverse pregnancy outcomes, including recurrent spontaneous abortion and embryonic resorption in humans and mice (Laissue *et al.*, 2016). These published findings were only associations and in this study the clinician who referred case F1374 did not report any parental miscarriages. In the study by Laissue and colleagues, a dual-luciferase assay was used to investigate the transactivation capacities of both human and mice *FOXD1* mutations on the two promoter targets involved in mammalian implantation, *Pgf* and *C3, in vitro* (Laissue *et al.*, 2016). An assay to test the p.(Gly194Asp) variant identified in F1374 was started but was soon abandoned following the advent of the gnomAD database. The discovery of homozygous null alleles in the gnomAD database shifted attention away from *FOXD1* as a prime candidate, as genetic redundancy was implied from these findings.

It should be noted that *FOXD1* has not been completely ruled out as a foveal hypoplasia gene, However, the prioritisation of *FOXD1* in this study has highlighted how important it is to check how many exomes or genomes are covered in the publicly available databases surrounding the variant of interest, prior to stating that it is a rare variant. The benefits of using population-scale measurements of variant density and allele frequencies in such a gene-identification study allows interpretation of variants in the context of large collections of variation from the “normal” population. The gnomAD dataset in particular is an extremely powerful tool when used for variant prioritisation, more so than the Exome Variant Server and

1000 Genomes, as it comprises two data sets, exome sequence data from 123,136 individuals and 15,496 individual genomes.

4.3.3.2 The identification of a homozygous frameshift variant in *LAMP1*

LAMP1 was identified as a top candidate disease gene following the discovery of a homozygous frameshift variant (p.(Gly370Afs*14)) in patient F1288 that was unreported in the publicly available databases.

LAMP1 (and LAMP2) are major structural components of the lysosomal membrane (Yogalingam *et al.*, 2008). The highly glycosylated protein has a sialylated N-terminal region orientated toward the lumen, one transmembrane domain and a cytoplasmic C-terminal tail (Chen *et al.*, 1985; Lewis *et al.*, 1985; Lippincott-Schwartz and Fambrough, 1986). The LAMP proteins maintain the integrity of the lysosome membrane and protect the limiting membrane from auto-digestion by forming a continuous carbohydrate lining on the inner leaflet known as a glycocalyx (Kundra and Kornfeld, 1999). LAMP proteins are present in late endosomes and lysosome and the proteins are also required for phagolysosome formation (Huynh *et al.*, 2007).

Lysosomes are important membrane-bound intracellular organelles with an acidic interior that are crucial regulators of cell homeostasis (Appelmans, Wattiaux and De Duve, 1955). They contain acidic hydrolases and specific membrane proteins and facilitate the degradation and recycling of proteins (autophagy) and extracellular molecules (phagocytosis) (Kornfeld and Mellman, 1989). However, lysosomes can also be secretory, for example as found in melanocytes and hematopoietic cells

(Yogalingam *et al.*, 2008). Lysosome-related organelles (cell-type specific), including melanosomes (which are found in melanocytes and RPE cells), release melanin to keratinocytes and thus produce pigmentation (Marks and Seabra, 2001; Stinchcombe, Bossi and Griffiths, 2004). Defects in melanosomes, such as protein transport and morphogenesis, cause a group of genetic disorders including Hermansky-Pudlak, Chediak-Higashi and Griscelli syndromic forms of albinism (Section 1.8.5.2) (Marks and Seabra, 2001). Lysosomal exocytosis is a Ca^{2+}-regulated mechanism whereby lysosomes are docked to the cell surface and fuse with the plasma membrane, emptying their content extracellularly (Andrews, 2000; Bossi and Griffiths, 2005). Lysosomal exocytosis is important in processes such as immune response, cell signalling and plasma membrane repair (Andrews, 2000). LAMP1 is reported to play a role in lysosomal exocytosis, as overexpression of the protein leads to an increased abundance of LAMP1 at the lysosome plasma membrane and increased lysosomal exocytosis of β-hexosaminidase (essential for glycoprotein metabolism in the maintenance of cell homeostasis) (Andrews, 1995; Kima, Burleigh and Andrews, 2001). Accumulation of LAMP1 on the plasma membrane has also been associated with increased exocytosis of lysosomal hydrolases in a mouse model of human sialidosis (MIM 256550), a lysosomal storage disease (Yogalingam *et al.*, 2008).

The role of LAMP-1 in the foveal hypoplasia cohort is still relatively unclear. It is possible that impaired autophagy or disrupted lysosomal exocytosis are the pathomechanism for the potential LAMP1-related disease seen in patient F1288. Impaired cargo clearance and processing in the endocytic/phagosome and autophagy pathways in the RPE leads to the accumulation of aggregates that

contribute to oxidative stress in blinding diseases such as AMD (Ferrington, Sinha and Kaarniranta, 2016; Wang *et al.*, 2009; Rakoczy *et al.*, 1999), Stargardt's disease (Tan, Toops and Lakkaraju, 2016; Anderson *et al.*, 2017; Bhattacharya *et al.*, 2017) and choroideremia (Wavre-Shapton *et al.*, 2013).

In LAMP-1 and LAMP-2 deficient cells, the interaction between phagosome and lysosome and the motility of lysosome and phagosome along microtubules is affected (Malicdan *et al.*, 2008). Both autophagic vacuoles and unesterified cholesterol levels are raised in embryonic fibroblasts, when both LAMP proteins are disrupted. Protein degradation rates, however, are not affected (Eskelinen *et al.*, 2004; Eskelinen, 2006). The ablation of both *lamp-1* and *lamp-2* in mice leads to embryonic lethality, suggesting that the proteins may have redundant function (Andrejewski *et al.*, 1999). However, *lamp-1*-deficient mice have a near-normal phenotype with normal lysosome function and morphology (Andrejewski *et al.*, 1999). The two proteins only share approximately 37% sequence homology, but protein topology and the heavy glycosylation is conserved (Malicdan *et al.*, 2008).

As detailed in Section 4.2.3.3, *LAMP1* expression is high during early development and was reported in fetal macula and nasal regions and the surrounding retina at 19-20 weeks. There is currently no known link to *SLC38A8*, or important transcription factors involved in early embryonic development, but the gene remains an interesting candidate.

4.3.4 Further interesting candidate variants identified in the foveal hypoplasia cohort

As we are all carriers of recessive alleles, the heterozygous findings may well be incidental (Gao *et al.*, 2015). The study by Gao et al estimated that each haploid set of human autosomes carries on average 0.29 recessive alleles. However, this study analysed lethal alleles so the average for non-lethal alleles will be higher (as they won't be selected against). It is thought that 1/10 humans carry an IRD recessive allele (Personal Communication, Prof Graeme Black, Manchester).

4.3.4.1 A heterozygous nonsense variant in *VSX2* in case F1405

In case F1405, a candidate nonsense variant, (p.(Gln298Ter)), was identified in the gene *VSX2* (also known as *CHX10).* The variant was prioritised as it has known ocular developmental function and is the earliest domain-specific marker expressed in the developing neural retina. *VSX2* encodes a homeobox transcription factor with the consensus DNA binding sequence TAATTAGC and is expressed in the neuroepithelium of the developing eye and the INL of the mature retina (Liu *et al.*, 1994). The gene was not present in the RetNet database and so was not identified through the BED file analysis of known IRD/albinism genes but was identified in the heterozygous variant analysis.

In OMIM, AR mutations in *VSX2* are reported to cause microophthalmia with coloboma 3 (MIM 610092) and microophthalmia, isolated 2 (MIM 610093). Reviewing the literature, mutations have been identified in 14 families. A number of mutations have been identified at the 200th residue Arginine, which lies in the *VSX2*

homeodomain, including two AR mutations (R200Q and R200P) identified in two families with non-syndromic microphthalmia, cataracts, and severe abnormalities of the iris (Ferda Percin *et al.*, 2000). A homozygous R200P mutation has also been identified in two consanguineous Qatari families with isolated microphthalmia and cloudy corneas (MIM 611093) (Faiyaz-Ul-Haque *et al.*, 2007). In the same residue, a homozygous nonsense mutation has been identified (R200Ter) in a consanguineous Afghan child with bilateral anophthalmia and an absent visually evoked potential, suggesting a mutation hotspot (Iseri *et al.*, 2010). A recurring homozygous R227W mutation has been identified in two consanguineous Arab families with isolated microphthalmia/clinical anophthalmia (MIM 610093) (Bar-Yosef *et al.*, 2004) and in a consanguineous Pakistani child with severe microphthalmia and absent visual evoked potential in one eye (Iseri *et al.*, 2010). Deletions have also been identified in this gene, including a recurring homozygous Leu84Serfs57 frameshift in two affected cases with microphthalmia and coloboma (Iseri *et al.*, 2010) and an approximate 4 kb deletion of exon 3 in a case of microphthalmia/anophthalmia without iris abnormalities (Bar-Yosef *et al.*, 2004). The only mutation report of a case without microphthalmia is a homozygous Lys258Serfs44 frameshift mutation in *VSX2* identified in a unique phenotype resembling Knobloch syndrome. The child had poor vision from birth and had smooth irides, lens subluxation, cone-rod dysfunction, and high myopia, however, the case did not have partial dislocation of the lens and there was no distinct macular atrophy (Khan *et al.*, 2015). The AR mutations identified in VSX2 in the literature are predicted to be null alleles due to impaired DNA binding, including the R200Q and R200P alleles (Ferda Percin *et al.*, 2000). In the current study, the Q298Ter nonsense variant identified in F1405 was heterozygous and is not located in the homeodomain (AAs 148-200). The VSX2 protein has a length of

361 amino acids, and the p.(Gln298Ter) variant is predicted to introduce a stop codon at position 298. As the stop codon is located in the last exon the protein is unlikely to be targeted for NMD and therefore this variant may elicit a different mechanism of disease compared to the null alleles reported in the literature.

In the literature there is only one family in which the mother of the affected individual had a retinal disease phenotype (Iseri *et al.*, 2010). This parent showed severe right macular dysfunction and mild-moderate left macular dysfunction (Iseri *et al.*, 2010). The parent was a carrier of the p.(Leu84Serfs57) frameshift mutation, a 1 bp deletion in exon 1 of *VSX2.* The parent's phenotype in the heterozygous state was suggested to be a case of incomplete dominance (Iseri *et al.*, 2010). There are no further reports of a disease phenotype in the heterozygous state in the literature. In the current study, case F1405 had foveal hypoplasia and posterior embryotoxon (a corneal abnormality), with no chiasmal misrouting.

VSX2, a pro-retina transcription factor, is reported to function in a signalling pathway with MITF, a pro-RPE transcription factor (Figure 4-26) (Bharti *et al.*, 2012). VSX2 is involved in the maintenance and proliferation of the neural retinal progenitor cell pool, differentiation of bipolar cells and timing of photoreceptor production (Widlund and Fisher, 2003; Vetrini *et al.*, 2004). In the regulatory circuit, PAX6 expression is essential for the proper expression of MITF which stimulates melanin synthesis by triggering the expression of melanogenic target genes, TYR and TYRP1 (Luo, Chen and Jimbow, 1994; Widlund and Fisher, 2003; Raviv *et al.*, 2014). Biallelic mutations

in *MITF* have been reported to cause albinism (Lowings, Yavuzer and Goding, 1992; George *et al.*, 2016).

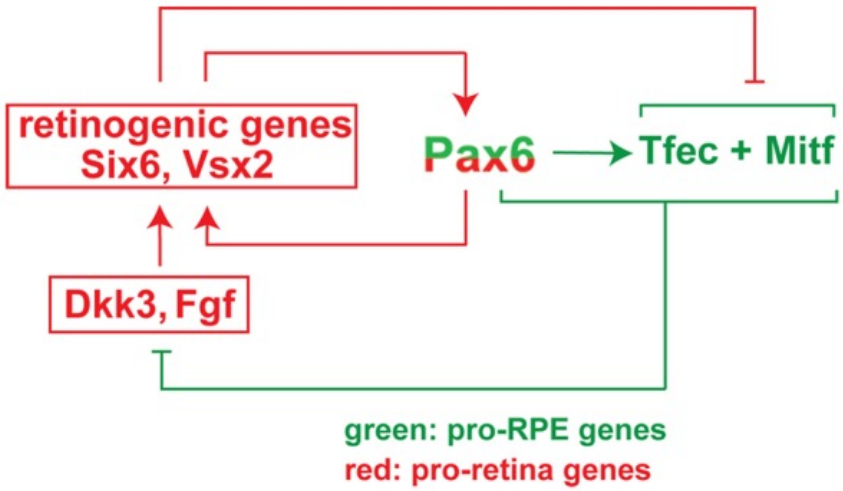


Figure 4-26 Schematic of the regulatory feedback loop involving *Vsx2*, *Pax6* and *Mitf* during the development of the mouse eye.

Co-expression of Pax6 with either Mitf or Tfec suppresses the expression of the pro-retina genes Fgf15 and Dkk3. This leads to activation of canonical WNT signalling, and subsequent upregulation of Mitf and Tfec expression, in a positive feedback loop which enables RPE differentiation and suppresses the retinal fate. Canonical WNT signalling is inhibited if Pax6, Mitf and Tfec are unable to inhibit Dkk3 and Fgf, leading to the upregulation of several retinogenic genes including Six6, Lhx2, and Vsx2. These retinogenic genes will increase the expression of Pax6 in pro-retina and anti-RPE feed-forward loops and Mitf expression is suppressed. Image reproduced with permission under the Creative Commons CC0 public domain dedication (Bharti *et al.*, 2012).

4.3.4.2 A heterozygous missense variant in *EPHA2* in case F1352

In patient F1352 a rare heterozygous variant was identified in the gene *EPHA2* (p.(Val383Gly)). The EphA receptors play a role in axon guidance during development (Section 1.2.5) and this led to prioritisation of this variant as a candidate in F1352. There are 14 known members of the Eph receptors in mammals, divided into two classes (EphA and EphB), depending on whether they

ligate A- or B- type ephrins (Lisabeth, Falivelli and Pasquale, 2013). Ephrin/Eph pairings regulate many biological processes including cell adhesion (as the Eph receptors and ephrins are both membrane -anchored), cell morphology, synaptogenesis and angiogenesis. EPHA2 has been detected in the vasculature of xenografted tumours and has been linked to VEGF expression and endothelial cell migration (Ogawa *et al.*, 2000; Brantley-Sieders *et al.*, 2004; Hunter *et al.*, 2006). EPHA2 is widely expressed in the human, including the brain, with levels highest in the esophagus (The Human Protein Atlas). In the brain, the expression of the protein is highest in the pituitary gland (The Human Protein Atlas).

The ephrins/EphA contribute to the specificity of the connections in the establishment of the topographic map (Pasquale, 2005). The EphAs mediate repulsive responses that prevent axons from growing into areas they shouldn't (Pasquale, 2005; Hansen, Dallal and Flanagan, 2004). Patient F1352 was not tested for optic nerve misrouting, therefore it will be important that they undergo visual evoked potential examination. If the V383G variant was a null, this could propose a mechanism for the chiasmal misrouting seen in FHONDA cases. Mutations in *EPHA2* have been identified to cause AD congenital cataract (Dave *et al.*, 2013). Mice deficient in Epha2 develop adult-onset cataracts, highlighting the importance of the gene in mammalian lens development and maintenance (Jun *et al.*, 2009). Most of the causative mutations identified so far reside in the cytoplasmic domains of the protein, with four mutations located in the sterile-α-motif domain and two in the tyrosine kinase domain (Bennett *et al.*, 2017). Similar to those mutations causing AD cataract, the AD V383G variant identified in patient F1352 lies in the cytoplasmic region of the protein, in the fibronectin type III repeat.

4.3.5 The identification of a potential *SLC38A8* duplication in case F1335

ExomeDepth CNV analysis revealed a potential homozygous duplication of *SLC38A8* exon 6. ExomeDepth was chosen as there were the appropriate number of control exomes available that had been sequenced in the same batches (same sequencing run and the libraries were prepared at the same time), as the cohort of patients.

Studies have suggested that CNV analysis should be undertaken alongside SNV and indel variant analysis, as CNV regions have been found to equate to approximately 12% of the human genome (Redon *et al.*, 2006), with >1000 CNVs accounting for approximately 4 million bp in the average human (Conrad *et al.*, 2010; Malhotra and Sebat, 2012; Abel and Duncavage, 2013).

CNVs, which are unbalanced alterations ranging from 50bp to whole chromosomes, are part of a larger collection of variation in the genome known as structural variations (SVs). There are many WES meta-analysis algorithms available, likewise to pathogenicity prediction softwares, that will aid in the detection of SVs on a larger scale. Such examples include CoNIFER (Krumm *et al.*, 2012), Exome Hidden Markov Model (XHMM) (Fromer *et al.*, 2012), FishingCNV (Shi and Majewski, 2013) and ExomeDepth (Plagnol *et al.*, 2012), that function by comparing read depths at the same target regions between affected and unaffected individuals, outputting variations in copy number format. Relying solely on WES read depth, however, results in high variability of coverage over individual exons, dependent on both PCR

bias and probe capture efficiency. Duplications on the whole have been described as especially difficult to clinically interpret, often with an inability to determine location, phase or effect, such as an assumed homozygous duplication identified in the gene *PRPF31* as a cause of AD RP (Ellingford *et al.*, 2018). To circumvent such issues, the use of WGS allows for wider scope of CNV detection, through the inclusion of non-coding sequence, analysis of split reads, paired end mapping, read depth and assembly based methods (Pirooznia, Goes and Zandi, 2015).

It is difficult to choose a WES CNV-assessment tool for which the positives outweigh the reported negatives. ExomeDepth was chosen for the analysis as it has been described as most effective tool when the samples are well-correlated, with minimal technical variability in the exome library preparation and therefore resulting in more normalised read count data (Kadalayil *et al.*, 2015). Studies have reported that ExomeDepth shows similarities in duplication detection to tools such as CoNIFER (Krumm *et al.*, 2012). ExomeDepth uses control sample BAM files, sequenced in the same run as the test sample and analysed using the same alignment pipeline, to build an optimised aggregate reference dataset of read count data. The program recommends the use of ten control samples, but, as shown in Appendix 17, the average reference dataset totalled only 7 out of 11 available references. ExomeDepth is not a very successful tool in the detection of novel genes (unless there is a homozygous deletion) as it generates many false positives (Guo *et al.*, 2013). The tool is dependent on high sample read depth and well-correlated samples (de Ligt *et al.*, 2013; Tan *et al.*, 2014), therefore suggesting variability in its performance and effectiveness (Ellingford *et al.*, 2017). However, it has been shown to be a very successful tool in identifying CNVs in candidate genes which are

specifically searched for. Case F1335 was referred for WGS, to be detailed in Chapter 5.

4.3.6 Summary

In this chapter, two patients (F1351 and F1378) with heterozygous *PAX6* missense mutations were solved. One patient (F1335) was identified to have a potential *SLC38A8* duplication CNV, to be investigated further by WGS (Chapter 5). Rare variants were identified in *LAMP1* and *VSX2*, identifying these as candidate foveal hypoplasia genes. This work also generated lists of rare candidate variants for the rest of the foveal hypoplasia cohort, but given the lack of a strong candidate identified in more than one case and the clear potential for missing mutations in *SLC38A8* (as seen in F1335), WGS was performed before further new candidate genes were investigated (Chapter 5). *SLC38A8*-related disease was also successfully excluded from some of the cases, including the consanguineous sib pair F1287 and F1288.

Chapter 5 Using WGS to investigate structural and non-coding variation in *SLC38A8*

5.1 Introduction

During this study, it became evident that mutations in *SLC38A8* not detectable using the conventional methods used might be prevalent in the foveal hypoplasia cohort, after Sanger sequencing of the gene failed to detect a second mutation in Case F1377 (Chapter 3). This case, alongside the patient who had a large region of homozygosity spanning *SLC38A8* but no causative mutation (F1310) (Poulter *et al.*, 2013), were prioritised for WGS to try and find the mutation not yet identified by Sanger sequencing. Also put forward for WGS was patient F1335 with the suspected duplication, as detailed in Chapter 4, in order to confirm and map the breakpoints of the CNV. The work of this chapter concludes with the analysis of the *SLC38A8* locus in the remaining *SLC38A8* mutation-negative individuals.

5.2 Results

5.2.1 WGS of foveal hypoplasia patient DNA

Whole genome DNA libraries were prepared from 10 consented individuals by Edinburgh Genomics (Section 2.3.24, Table 2.4). Included within this cohort was F1335 (with the suspected duplication), F1310, F1377, F1374, F1352, F1369,

F1298, F1071, F1403 and F1404. Libraries were prepared and sequenced in separate batches over the course of 18 months, with alignment to the GRCh37/hg19 human reference sequence using a similar approach to the pipeline for WES analysis (Sections 2.3.25.2, 2.3.27.2, 2.3.27.3).

5.2.1.1 WGS QC and metrics

QC was performed for DNA samples and libraries by Edinburgh Genomics to check the DNA was high quality and also that the insert size was as expected. The QC was also performed to identify any contaminating adapter-dimers, which are known to cause clustering issues. The QC steps undertaken are detailed in Section 2.3.27.4. All of the WGS samples passed the QC checks, as processed by Edinburgh Genomics. As shown in Table 5-1, the sample yield, or the number of bases generated per run, was as expected for 150 bp paired-end sequencing. The number of those bases with a quality score of 30 or above is represented by the % Q30.

Individuals F1310 and F1377 were sequenced in the first batch of genomes sent to Edinburgh Genomics. Another 8 genomes (including F1335), were sequenced at a later stage of the project. The yield (the total number of giga bases read for the sample per run) for both F1310 and F1377 is notably higher than for the individuals from the second batch of sequencing, with an average yield of 150.19 giga bases compared to 141.24 giga bases, respectively. All samples had an average read depth >30. The mean coverage was the highest for individual F1377, with an average read depth of X38.85.

Patient	Read pairs sequenced	Yield (Gbp)	% Q30	No. of reads in BAM file	Mapping rate (%)	Properly mapped reads rate (%)	Duplicate rate (%)	Mean coverage
F1310	510,061,390	152.36	86	1,020,122,780	99.63	98.07	24.79	36.22
F1377	495,634,399	148.01	88	991,268,798	99.76	97.96	16.66	38.85
F1335	467,571,613	139.67	83	935,143,226	89.37	87.41	10.77	33.58
F1071	461,272,426	137.84	81	922,544,852	99.65	97.54	11.70	37.05
F1298	487,910,450	145.79	83	975,820,900	99.66	97.61	22.08	34.58
F1352	476,586,451	142.40	82	953,172,902	99.64	97.58	21.68	33.95
F1369	483,975,381	144.61	82	967,950,762	99.64	97.62	22.89	33.98
F1374	486,692,815	145.43	82	973,385,630	99.60	97.40	20.55	35.05
F1403	437,543,140	130.73	81	875,086,280	99.57	97.52	19.67	31.98
F1404	453,383,643	135.44	81	906,767,286	99.62	97.50	20.95	32.54

Table 5-1 WGS sample summary metrics.

WGS summary metrics for all 10 foveal hypoplasia individuals analysed by Edinburgh Genomics. The data shows the following metric calculations per genome: Number of read pairs sequenced, Yield (the total number of bases read for the sample per run), % Q30 (the proportion of the bases in the yield that had a Q score >30), Number of reads in the BAM file, Mapping rate percentage, Percentage of properly mapped reads, Percentage of duplicate reads and Mean sample coverage.

5.2.2 Whole-genome analysis of case F1310

As detailed in Chapter 3, individual F1310 was selected as an ideal candidate for WGS. Previous SNP genotype investigation by collaborators had mapped a large region of homozygosity spanning *SLC38A8*, and microsatellite analysis performed in Chapter 3 confirmed this.

Clinical photographs were available for this individual and her father and brother (Figure 5-1). Individual F1310 (Figure 5-1 A, B) shows no hypopigmentation and there is absence of the foveal pit. F1310's brother and father are unaffected (Figure 5-1 C, D).

Exploration of the paired-end reads covering *SLC38A8* was undertaken using IGV which revealed a homozygous paracentric inversion spanning exons 6, 7 and 8 of *SLC38A8* (Figure 5-2). From this global overview, it is clear that the pair orientations are anomalous, as shown with the mismatched read colours of blue (reverse reads) and green (forward reads). Enlarging a portion of the screenshot taken from the IGV output, as shown in Figure 5-3, highlights the failure of the paired-end mate reads to match.

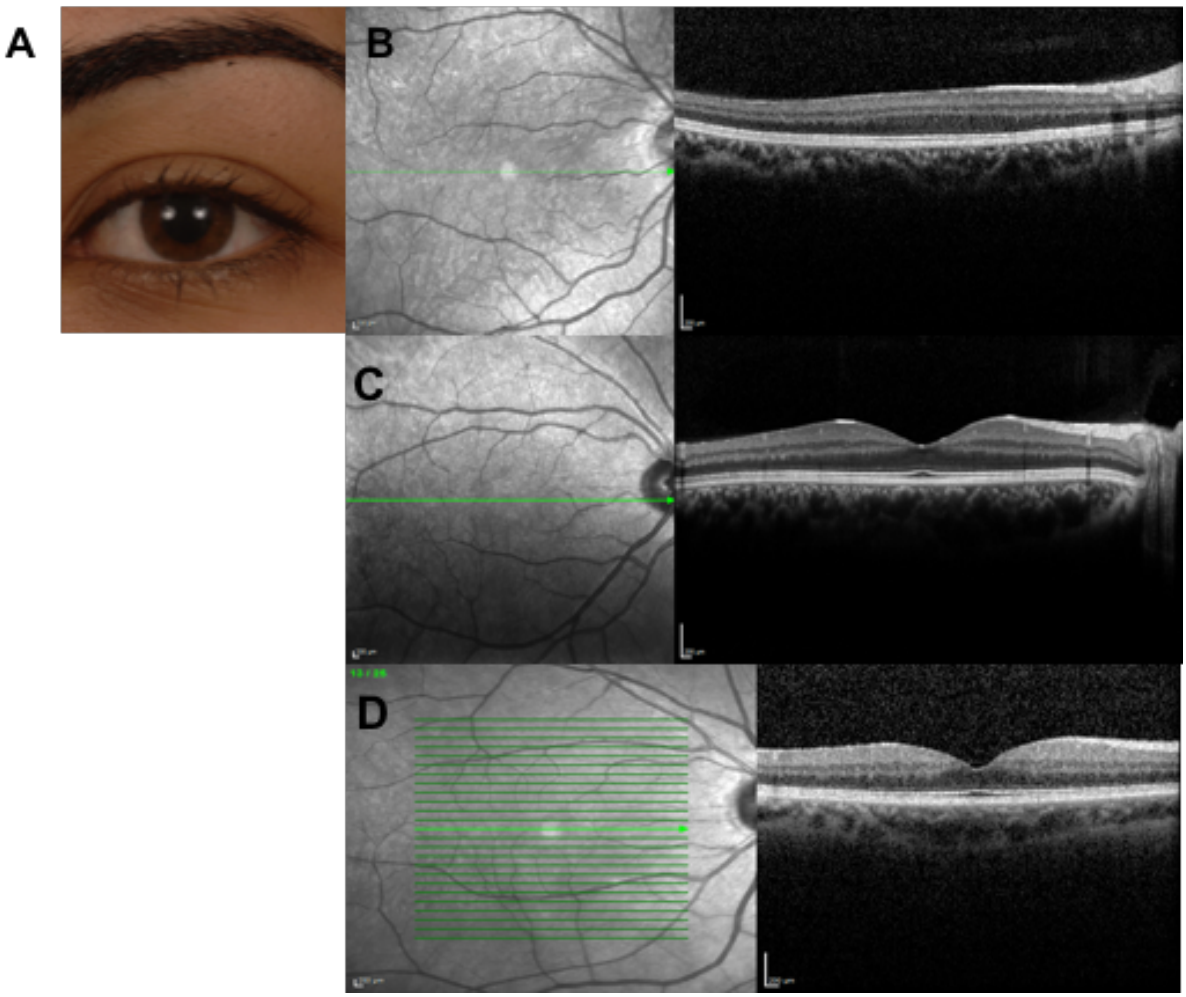


Figure 5-1 Clinical data for affected case F1310 and her unaffected brother and father.

Case F1310 presents with foveal hypoplasia. (A) Individual F1310 presents with naturally pigmented brown eyebrows and brown irises. (B) Horizontal OCT of the right eye of individual F1310 shows foveal hypoplasia. (C,D) Horizontal OCT of the right eye of brother and father. Their foveal pits have developed normally. The green line shows the location of the cross-section. Images reproduced with the kind permission of Mr Kamron Khan (Leeds).

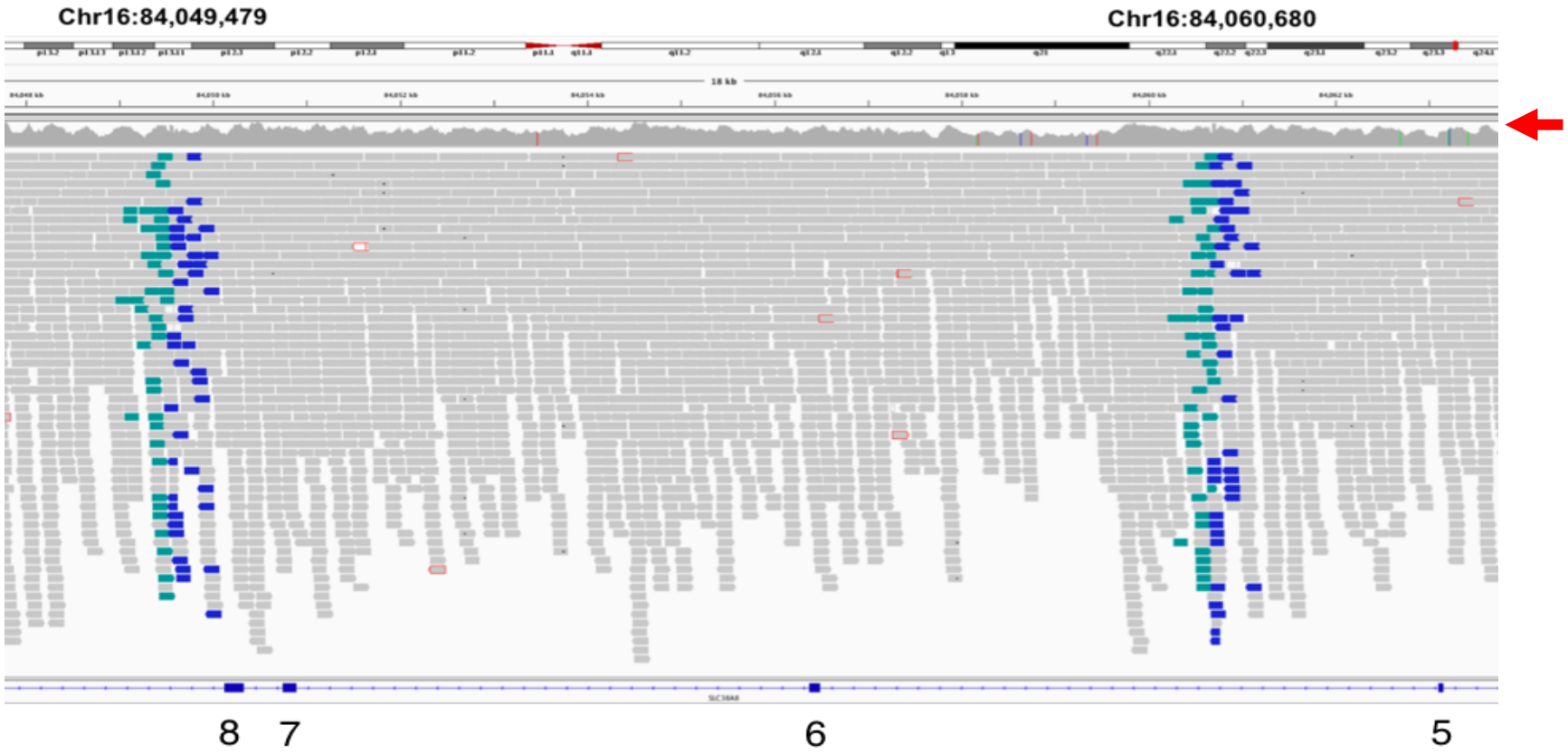


Figure 5-2 IGV global overview of the *SLC38A8* inversion identified in case F1310.

IGV alignment of WGS reads over *SLC38A8*. The inversion approximately spans from chr16:84,049,479-84,060,680, encompassing *SLC38A8* exons 6, 7 and 8. Individual mismatched bases can be selected for using IGV and are represented as coloured base pairs. The mismatched bases are highlighted by colour (blue: reverse reads and green: forward reads) and intensity according to base call and quality and are positioned on the grey polygons (reads). The red arrow highlights the read depth.

Further to highlighting the paired mates, when zooming below the ~50kb range in IGV, individual aligned reads can be interrogated. Visualisation of the mismatched base pairs in the two discordant regions revealed that the anomalous mate pairs are located on either side of a 42bp palindromic repeat sequence (Figure 5-3).

IGV also showed a visualisation of the read depth, shown as a coverage histogram plot (red arrow, Figure 5-3). The read coverage histogram shows approximately double the read depth over the two locations of the palindromic sequence. It was thought that the doubling of read depth over both palindromes was most likely the result of an alignment issue, as the sequence appears at both breakpoints.

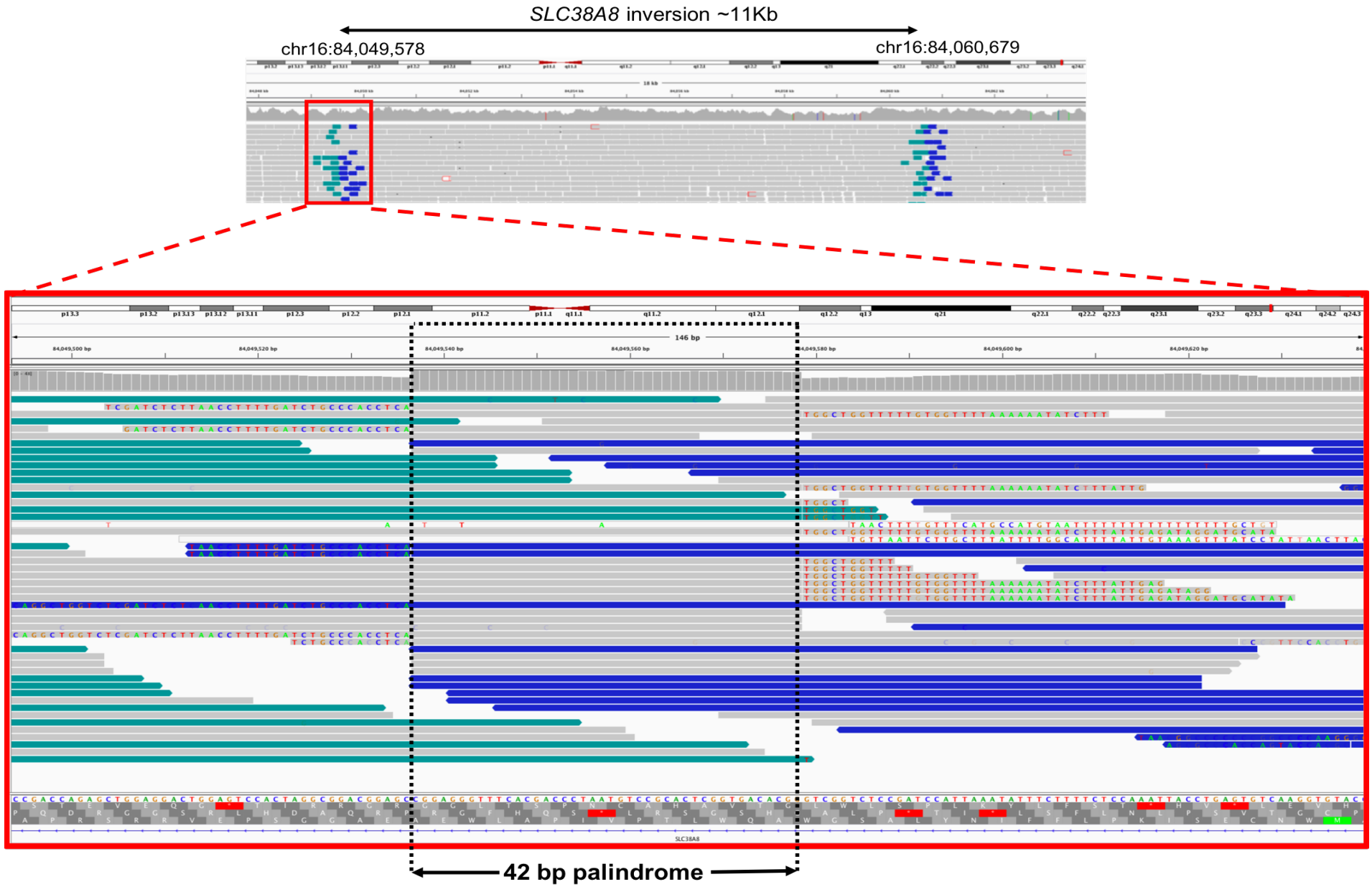

SLC38A8 inversion ~11Kb
chr16:84,049,578
chr16:84,060,679
146 bp
42 bp palindrome

Figure 5-3 IGV alignment of paired-end reads over *SLC38A8* in Case F1310.

Schematic of IGV alignment of WGS reads over *SLC38A8*. The coloured reads show the paired mates with the homozygous inversion spanning a region of approximately 11.2Kb. The output highlights the mismatched base pairs coloured on reads positioned either side of a 42bp palindromic repeat sequence (arrows and black box).

Due to the mismatched base pairs, it was clear that the exact location of the inversion breakpoints were positioned within the palindromic sequences but the exact location could not be determined. As recommended by the HGVS sequence variant nomenclature guidelines (v15.11) for repeats (http://varnomen.hgvs.org/recommendations/DNA/variant/inversion/), the most 3' position possible of the sequence should be assigned for the inverted sequence. In this instance, the arbitrary coordinates are therefore g.84,049,578_84,060,679inv (GRCh37) or g.84,015,974_84,027,074inv (GRCh38), an inversion of 11,101 bp.

BreakDancerMax, a read-pair algorithm (Section 2.3.23.2) and DELLY2, a split-read and read-pair algorithm (Rausch *et al.*, 2012) (Section 2.3.23.3), were applied following the identification of the inversion in the IGV. The BreakDancerMax output returned a single pair of predicted coordinates (chr16:84,016,486-84,027,618), with a predicted inversion size of 10,692 bp. Investigation of SVs using DELLY2 produced more accurate *SLC38A8* breakpoint coordinate results than BreakDancerMax. Two breakpoints, on either side of the palindromic sequence, were presented: g.84049536_84060721inv and g.84049579_84060680inv (GRCh37), of which the latter were the most 3' (according to the HGV recommendations).

5.2.2.1 Validation of chromosomal breakpoint positions

In order to confirm the inversion breakpoints in F1310, PCR primers were designed using ExonPrimer to amplify the novel sequence across the two breakpoints (Appendix 14). A total of 4 primer pairs were designed to either amplify the inverted *SLC38A8* sequence (as identified in F1310), or WT sequence without the inversion (Appendix 14). Primers were designed to amplify products spanning the inversion breakpoints, with flanking sequence of <500 bp each side. The primer combinations are listed as follows and are presented in Figure 5-4:

BP1 and BP2 amplified the first breakpoint of the inversion identified in F1310 (chr16:84,049,579), yielding a product of 946 bp.

BP3 and BP4 amplified the second breakpoint of the inversion in F1310 (chr16:84,060,680), yielding a product of 884 bp.

BP1 and BP3 amplified the WT sequence near breakpoint 1 (with no inverted DNA sequence), yielding a product of 961 bp.

BP2 and BP4 amplified WT sequence near breakpoint 2 (with no inverted DNA sequence), yielding a product of 869 bp.

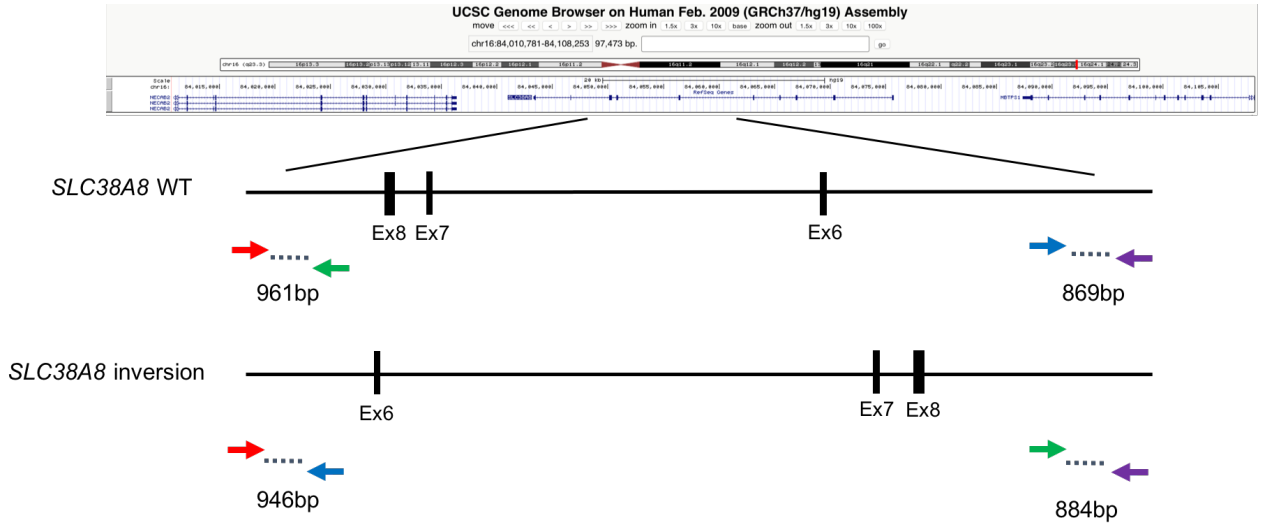
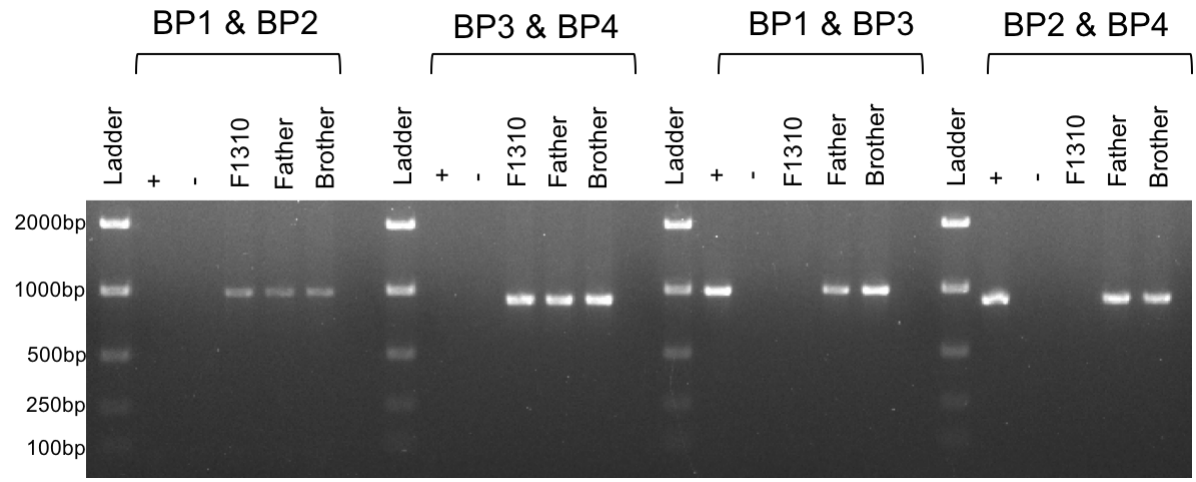


Figure 5-4 Schematic representation of the *SLC38A8* inversion PCR amplification.

The UCSC Genome Browser (GRCh37) chromosome 16 location of the *SLC38A8* inversion identified in F1310, encompassing exons 6-8. PCR amplification over coordinates chr16:84,049,578-84,060,680 using the 4 primer combinations: Red (BP1), Blue (BP2), Green (BP3), Purple (BP4).

As expected, products of approximately 946 bp and 884 bp were detected after PCR amplification of the inverted sequence using F1310 DNA and subsequent size fractionation using gel electrophoresis (Figure 5-5). Sanger sequencing of the amplified PCR products then confirmed the same inversion breakpoints in patient F1310. The optimisation of this PCR also allowed screening for this chromosomal rearrangement in family members. The results showed that the inversion was present in a heterozygous state in the unaffected father and brother of F1310 (Figure 5-5).



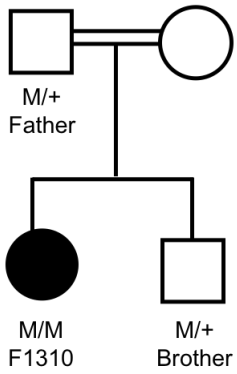


Figure 5-5 Segregation of the *SLC38A8* inversion in the family of Case F1310.

The inversion was PCR-amplified in the DNA of affected individual F1310 and DNA of available family members (father and brother). Top pane, from left to right: Primer combination BP1 and BP2 to amplify the first inversion breakpoint of F1310, showing a band of approximately 946 bp; Primer combination BP3 and BP4 to amplify the second inversion breakpoint of F1310, showing a band of approximately 884 bp; Primer combination BP1 and BP3 to amplify the WT sequence in the region of breakpoint 1, showing a band of approximately 961 bp; Primer combination BP2 and BP4 to amplify the WT sequence in the region of breakpoint 2, showing a band of approximately 869 bp. The DNA marker used was EasyLadderI. Bottom pane, pedigree of affected individual F1310 and unaffected father and brother. + Unaffected control sample, - Water control.

Further to segregation analysis of the inversion in the family, the *SLC38A8* mutation-negative patient cohort was also screened. The primer pairs detailed above were used in a duplex PCR alongside standard PCR amplification of a second gene that was unrelated to this project. This duplex PCR was performed so that there was a positive PCR control for each DNA sample (Figure 5-6). Screening failed to reveal any similar genomic rearrangement in the remaining mutation-negative cohort, as shown in Figure 5-6, A, B. PCR of WT sequence instead amplified products of approximately 961 bp and 869 bp in all *SLC38A8*-negative patients, excluding F1310 where the inversion was present (Figure 5-6, C, D).

Exploration of the regulatory landscape across the inversion breakpoints using RepeatMasker revealed short interspersed nuclear element (SINE) repeats spanning the regions of the 5‘ and 3’ inversion junctions (g.84049578_84060679inv) (Figure 5-7). At the 3’ breakpoint the SINE spanned the region chr16:84,049,316-84,049,583, a genomic size of 268 bp and in the region of the 5’ breakpoint (chr16:84,060,679), the SINE spanned the region chr16:84,060,676-84,060,962, a size of 287 bp (Figure 5-7). There were no long terminal repeats (LTRs), locus control regions (LCRs), microsatellites or other repeat elements in the region of the breakpoints.

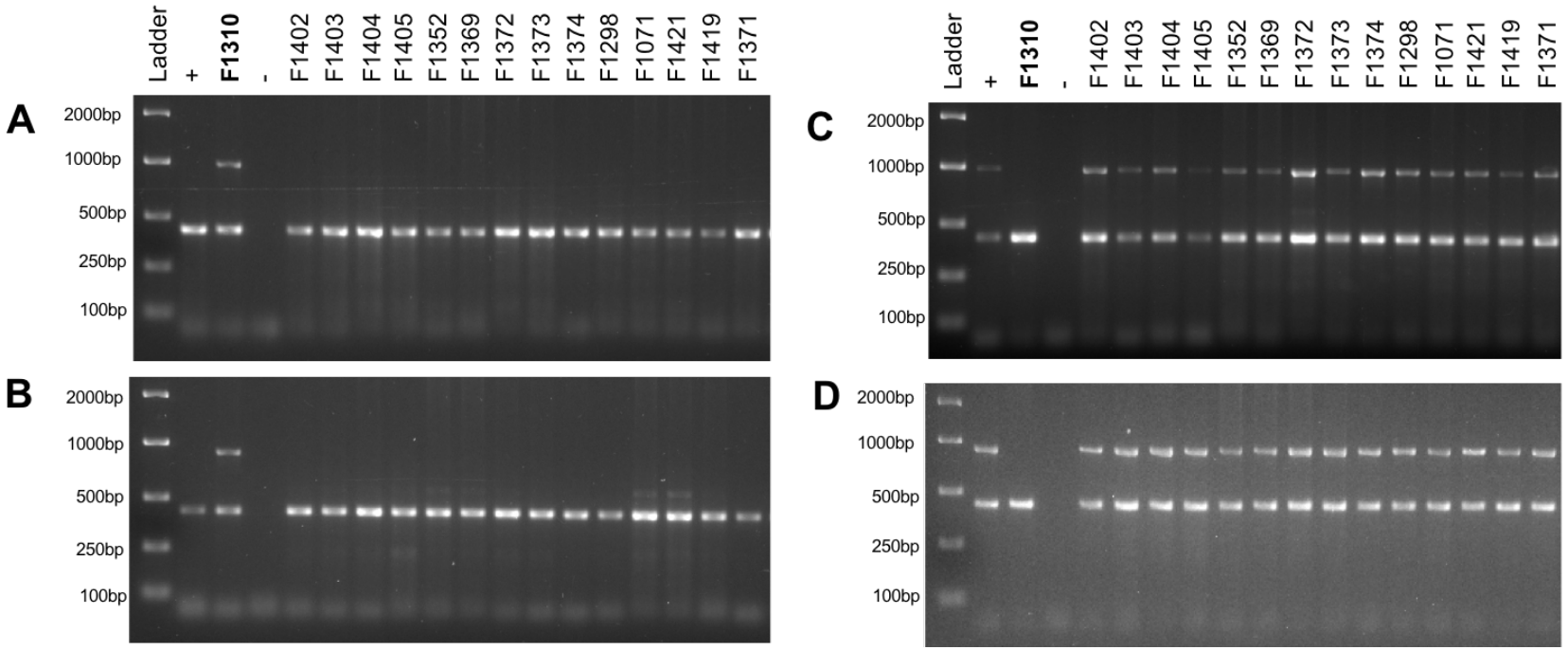


Figure 5-6 Screening for the detection of the *SLC38A8* inversion using duplex PCR.

All *SLC38A8*-negative individuals and case F1310 (with homozygous *SLC38A8* inversion), shown in bold font, were screened in a duplex PCR with breakpoint primer pairs alongside primers amplifying 417 bp of an unrelated gene as a positive control. A Primer combination BP1 and BP2 to amplify the first breakpoint of F1310. A band of approximately 946 bp is visible for case F1310 only. B Primer combination BP3 and BP4 to amplify the second breakpoint of F1310. A band of approximately 884 bp is visible for case F1310. C Primer combination BP1 and BP3 to amplify the WT sequence in the region of breakpoint 1, showing a band of approximately 961 bp in all cases except F1310. D Primer combination BP2 and BP4 to amplify the WT sequence in the region of breakpoint 2, showing a band of approximately 869 bp in all cases except F1310. The duplex control product of 417 bp can be seen for all DNAs. The DNA marker used was EasyLadderI. + unaffected control sample, - water control.

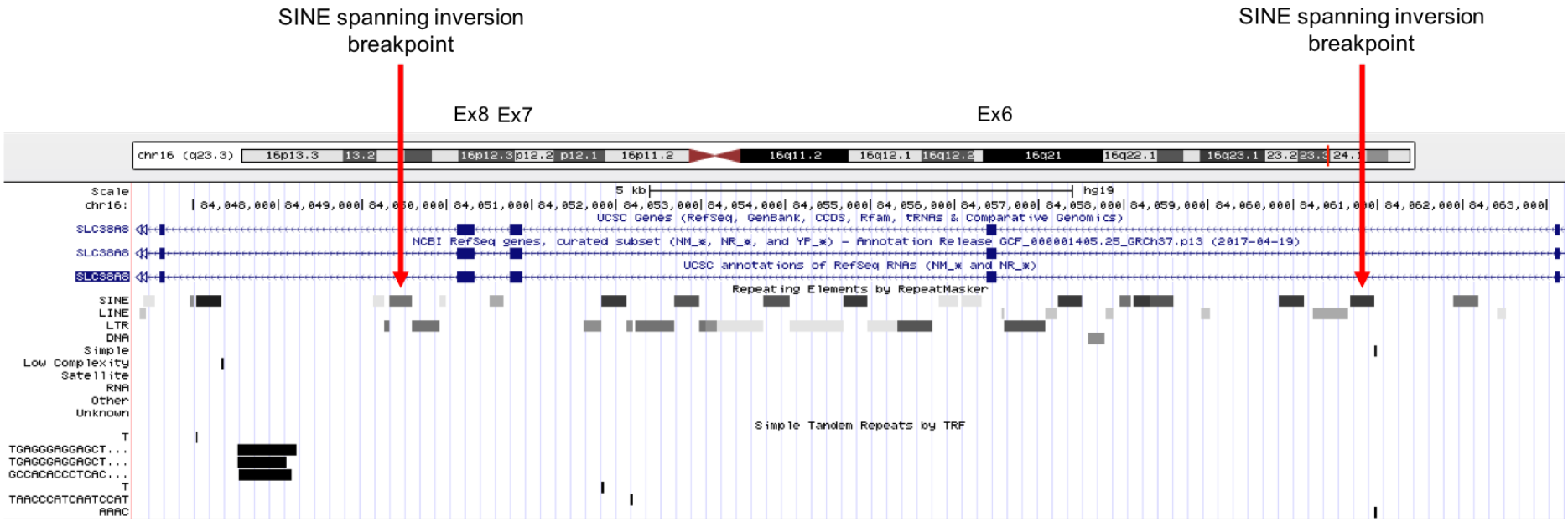


Figure 5-7 SINE repeat elements span the *SLC38A8* inversion breakpoints.

UCSC Genome Browser overview of the g.84049578_84060679inv breakpoint junctions and repeating elements highlighted by RepeatMasker.

5.2.3 Searching for the missing mutation in patient F1377

The initial screen of *SLC38A8* in the novel cohort of patients revealed a heterozygous missense change in the Methionine translational start site (c.2T>C; p.(Met1?)), but no second mutation was identified using this screening method (Section 2.3.9).

Sanger sequencing identified six heterozygous SNPs across *SLC38A8* suggesting that there was not a heterozygous deletion of the full gene (Section 3.2.1, Appendix 13). It was therefore hypothesised that the second mutation might not be detectable by Sanger sequencing (such as the inversion), or might be present in regions outside of the annotated coding sequence (enhancers, promoters, 5' and 3' untranslated regions (UTRs) or novel exons)). The aim of this experiment was to search for this second mutation.

5.2.3.1 Defining the 5' UTR of *SLC38A8*

Presently, the official annotation of *SLC38A8* (NM_001080442.1) does not contain any 5' UTR and only has 116 bp of 3' UTR spanning chr16:84,043,387-84,043,271 (UCSC Genome Browser, last accessed 07/12/2017). The aim of this experiment was to determine if *SLC38A8* had a 5' UTR or any novel exons, so that these regions could be screened for mutations.

UCSC Genome Browser (Section 2.4.2) was used to interrogate all of the available mRNA sequences (both human and non-human) (GenBank, UniGene, H-Inv, Gene

Bounds, Poly(A), SIB Alt-Splicing), sequence conservation (Multiz Alignments, GERP, Primate Chain, Vertebrate Chain) and gene predictions (UCSC Genes, NCBI RefSeq, Ensembl Genes, GENCODE, UniProt, AceView Genes, AUGUSTUS, Consensus CDS, Genscan, MGC Genes), but no evidence of additional exons or UTR sequences were observed in the human. Investigation of ESTs (non-human) revealed a new EST which shows two upstream exons of 5' UTR in the monkey (thymus tissue, DK581820), but they did not appear to overlap with the 5' UTR sequence of 290 bp sequenced in the human in this study. Interrogation of in-house RNA-seq datasets (human adult cornea, undifferentiated and differentiated SHSY-5Y and RPE1 cell line RNA-seq) did not reveal any evidence for additional exons or UTR sequences, in line with the data presented on UCSC Genome Browser. Interrogation of the UTR database, UTRdb (http://utrdb.ba.itb.cnr.it/) (Grillo *et al.*, 2010) revealed that the mouse and frog have *SLC38A8* 5' UTR sequences of 177 bp and 259 bp respectively and 3' UTR sequences of 1147 bp and 103 bp, respectively. Attempts were made to map these sequences onto the human genome, however there were major differences in homology and so this was not feasible.

In order to define the 5' UTR of *SLC38A8*, RNA ligase-mediated rapid amplification of cDNA ends (RLM-RACE) was performed (Section 2.3.14). Brain cDNA was chosen for the experiment based on the expression data results in the original gene identification study (Poulter *et al.*, 2013). Total and poly(A)+ adult human brain cDNA was synthesised from commercial sources of RNA using random primers (Section 2.3.13).

Following the verification of the cDNA using p53 primers, the expression of *SLC38A8* was confirmed in both samples of adult human brain RNA using previously published primers spanning *SLC38A8* exons 7-8 (Poulter *et al.*, 2013) (Appendix 1), yielding a product of 188 bp (Figure 5-8). Based on these results, the human brain poly(A)+ RNA was used for the RLM-RACE experiment.

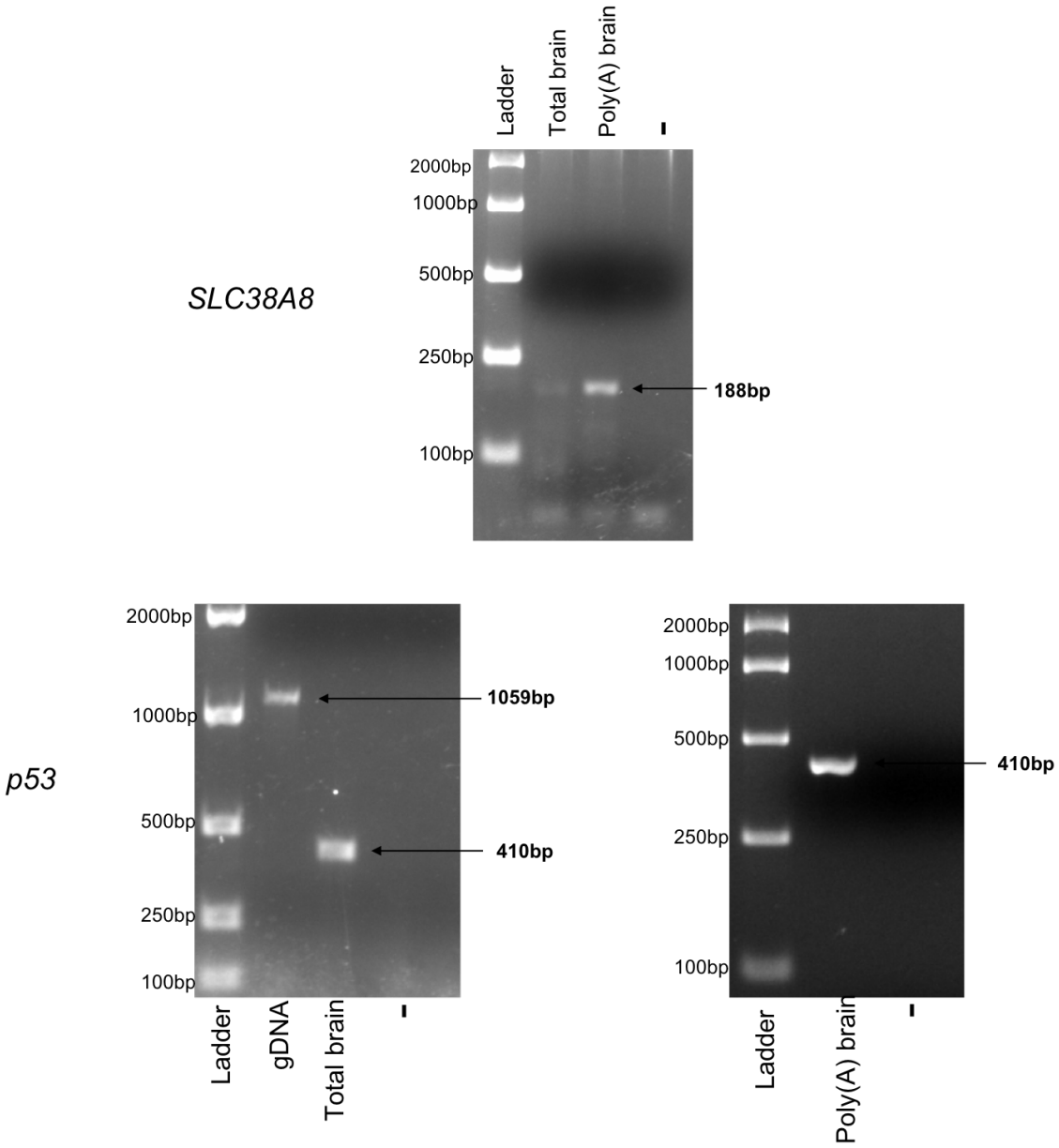


Figure 5-8 Amplification of the SLC38A8 gene using primers designed to amplify exons 7-8, in human adult brain total and poly(A)+ RNA.

The PCR products (188 bp) generated by the amplification of cDNA created from human adult brain total RNA and poly(A)+ RNA using previously published primers spanning *SLC38A8* exons 7-8 (Poulter *et al.*, 2013) (Appendix 1) (upper pane). *p53* expression (lower panes) was tested in all cDNA as a positive control (size 410 bp). The DNA marker used was EasyLadderI. gDNA: genomic DNA.

The Ambion FirstChoice RLM-RACE kit was chosen for the experiment (Section 2.3.13). This protocol is a PCR-based technique designed to clone the full-length 5' UTR (Figure 5-9).

Selection of only full-length 5' UTR began by treating poly(A)+ RNA with calf intestinal phosphatase (CIP) to remove the 5'-phosphate from all molecules which contain free 5'-phosphates (ribosomal RNA, fragmented mRNA, tRNA and any contaminating gDNA). Any full-length 5' UTRs (with cap structure found on intact 5' ends) are unaffected by treatment with CIP. The RNA is then treated with tobacco acid pyrophosphatase (TAP) to remove the 5' cap structure from the full-length 5' UTR. A 5'-monophosphate then remains. T4 RNA ligase-mediated ligation of a synthetic 45 bp RNA adaptor to the RNA population then takes place. The majority of the decapped, full-length 5' UTRs will acquire the adapter sequence at the 5' end. To amplify the 5'-end of the transcript, a random-primed reverse transcription reaction and nested PCR are performed.

A schematic of the RLM-RACE primers used in this experiment is shown in Figure 5-9 (Appendix 1). A 5' RACE Outer Primer and 5' RACE Inner Primer was provided in the kit (Figure 5-10) (Section 2.3.13). Two *SLC38A8*-specific nested primers were

manually designed for the experiment, labelled 5' *SLC38A8* outer primer and 5' *SLC38A8* inner primer.

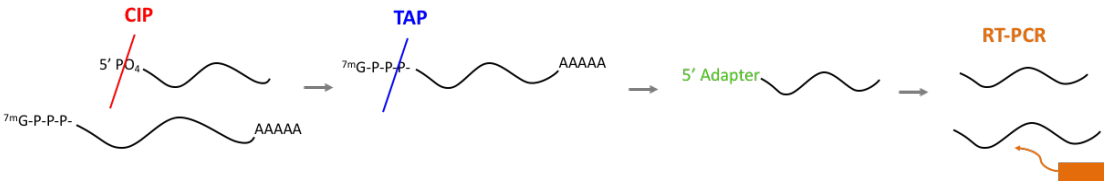


Figure 5-9 Schematic representation of the 5' RLM-RACE protocol.

Calf intestinal phosphatase (CIP) removes the 5' PO_4, followed by tobacco acid pyrophosphatase (TAP)-mediated removal of the cap structure from full-length 5' UTRs. The adapter is ligated to the 5' end of the decapped mRNA and the process is completed with reverse transcription PCR.

The nested *SLC38A8* primers were positioned as 5' as possible to the beginning of the known RNA transcript. The primers were also designed >100 bp apart in order to produce different sized PCR fragments that could be easily distinguished by agarose gel electrophoresis.

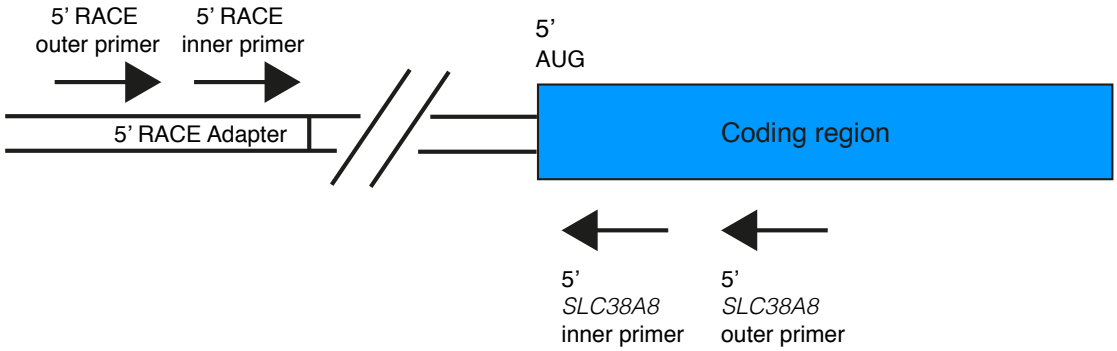


Figure 5-10 Positions of the RLM-RACE primers.

The positions of the nested primers used in the RLM-RACE experiment. Two primers were supplied by Ambion (5' RACE outer/inner primer) and two were designed by the author (5' RACE *SLC38A8* outer/inner primer). The RLM-RACE protocol is detailed in Section 2.3.14.

The poly(A)+ RNA was treated with TAP, CIP and the 5' RACE adapter was ligated. After RT-PCR using random decamers (Section 2.3.13), the cDNA was used in a nested PCR (Figure 5-11). Included in the nested PCR was a minus-template control (-), which included all of the PCR components used, except cDNA template. A minus-TAP control (TAP -) was also carried through the adapter ligation, reverse transcription and PCR (TAP -). This control was not expected to yield a band since the RNA was either dephosphorylated by CIP treatment, or retained its cap structure, since it was not treated with TAP, so would not be able to ligate to the adapter. Size fractionation by agarose gel electrophoresis was performed for all products (Section 2.3.5). The TAP - showed multiple faint bands ranging from <100-500 bp but no 5' UTR product was obtained using the RLM-RACE method.

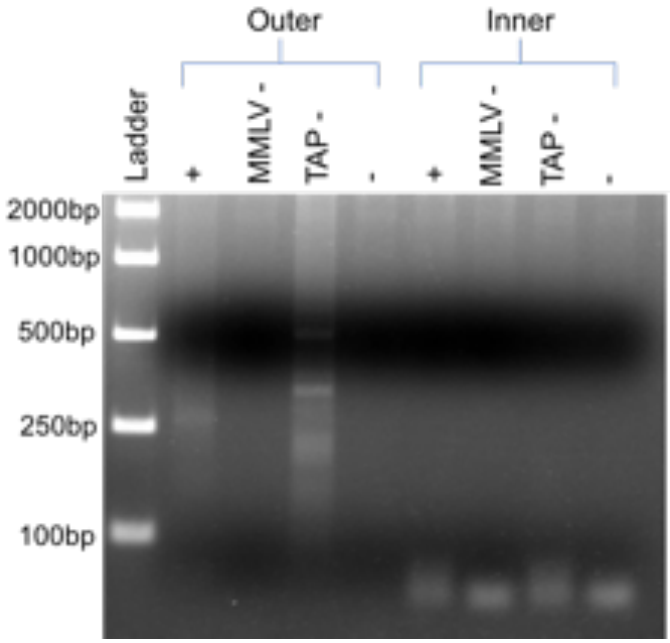


Figure 5-11 Agarose gel electrophoresis analysis of 5' RLM-RACE nested PCR products.

RLM-RACE nested PCR amplification of the 5' region of *SLC38A8* in poly(A)+ brain RNA treated with CIP, TAP and 5' RACE Adapter-ligated (+). A Moloney Murine Leukemia Virus reverse transcriptase negative control (MMLV -), tobacco acid pyrophosphatase negative control (TAP-) and minus-template control were included in both the first round of PCR (outer) and the nested reaction (inner).

As an alternative strategy to define the 5' UTR sequence of *SLC38A8*, primers were designed to amplify presumptive UTR from cDNA generated from total and poly(A)+ adult human brain, as used previously (Section 2.3.13). Forward primers were designed manually (Section 2.3.1) in the sequence upstream of the TIC of *SLC38A8* obtained from the RefSeq alignment in the UCSC Genome Browser (GRCh37/hg19). Five forward primers (labelled UTR-Primer) were designed to pair with cDNA_ex1-R, positioned in *SLC38A8* exon 1 (Appendix 1) (Figure 5-12). The most 5' primer (UTR-Primer_5) was located 714 bp upstream of the AUG start codon.

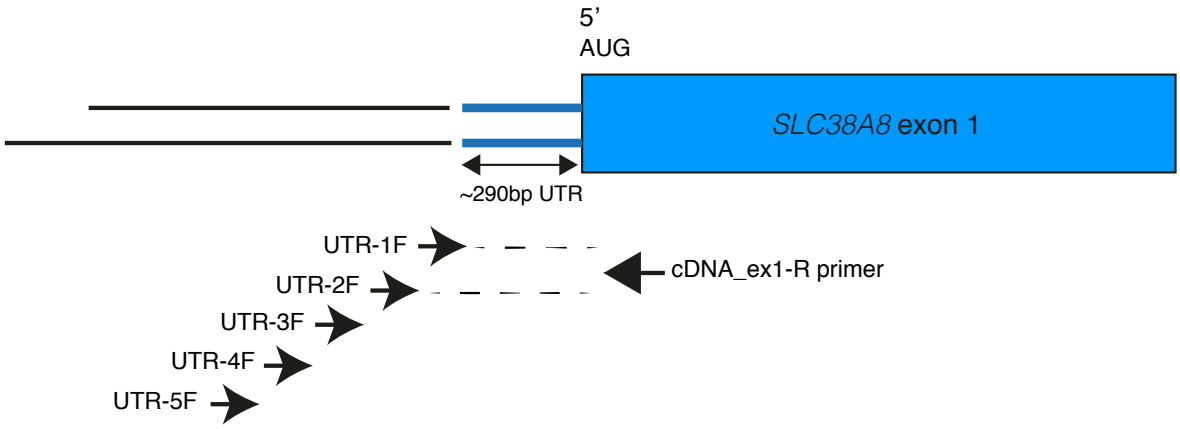


Figure 5-12 RT-PCR of the presumptive 5' UTR of *SLC38A8*.

Using primers designed in the presumptive 5' UTR of *SLC38A8*, 290bp was amplified in poly(A)+ adult human brain cDNA using primer combinations UTR-1F/2F with cDNA_ex1-R.

The PCR products generated from these 5 primer pairs are shown in Figure 5-13. A band of approximately 250-300 bp was attained with UTR-1F and a band of approximately 350-400 bp with UTR-2F in adult human poly(A)+ brain cDNA. Sanger sequencing (Section 2.3.9) of these products confirmed 290 bp of sequence immediately upstream of the ATG start codon of *SLC38A8*. As this sequence does not contain an ORF it was designated as 5' UTR. This UTR region was sequenced in

patient F1377 but no variants were found. There was a smaller band of approximately 200 bp present for UTR-4F with cDNA_ex1-R. Sanger sequencing of this PCR product revealed it was an artefact.

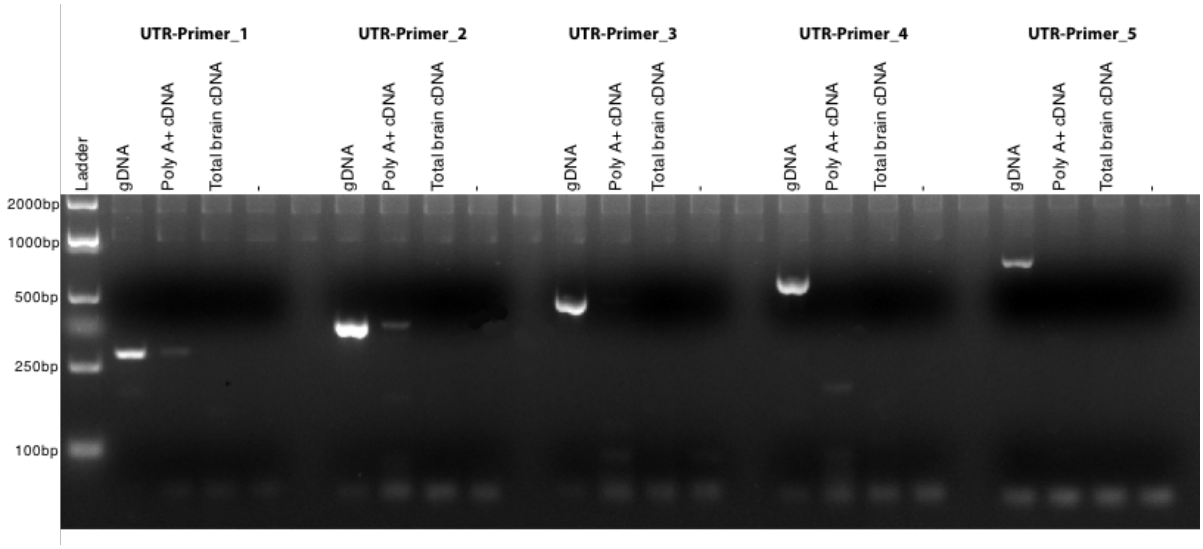


Figure 5-13 Defining the 5' UTR of *SLC38A8* using RT-PCR.

Using five forward primers designed in the presumptive 5' UTR of *SLC38A8*, approximately 290 bp was amplified in both gDNA and human poly(A)+ human adult cDNA using primers UTR-2F and cDNA_ex1-R and also UTR-1F with cDNA_ex1-R.

5.2.3.2 The search for a second mutation using WGS and splicing assays

Following alignment and variant calling (Section 2.3.27, Appendix 2), the final variant files were further filtered to retain only the variants that were present in the *SLC38A8* locus, chr16:84,036,380-84,087,365 (GRCh37), using the vcfhacks command FilterVariantsbyLocation.pl. This region contains all the genomic sequence between the two genes flanking *SLC38A8*; *NECAB2* and *MBTPS1*. The analysis identified a total of 224 variants. The list was further filtered to retain only heterozygous variants, generating a list of 164 variants. The UTR sequence identified in Section 5.2.3 was

investigated and confirmed that no variants were located in this region. The list of heterozygous variants was investigated further to establish if they were rare. Only 6/164 variants were rare (either unreported in publicly available databases or MAF ≤0.00001) and these variants were further analysed using the HSF online tool.

This analysis prioritised a single base-pair deletion located at 301 bp within intron 5 of *SLC38A8* (chr16:84,056,794 (GRCh37)) that was predicted to activate a cryptic intronic splice donor site, causing potential alteration of splicing and extending the length of the exon by 289 bp (Table 5-2). The variant was not present in the dbSNP or gnomAD publicly available databases (last accessed 12/12/2017).

cDNA position	Splice site type	Motif	New splice site	WT score	c.691-301 score	Exon length variation (bp)
-301	Donor site	gtgggtgtt	TGGgtgttt	40.97	70.88	+289

Table 5-2 HSF prediction of SLC38A8 intronic variant c.691-301delG.

The HSF (version 3.0) (Section 2.4.5) prediction for the variant (c.691-301delG) (highlighted in red), position chr16:84,056,794, identified in WGS data of individual F1377. The variant is predicted to cause activation of an intronic cryptic donor site, causing potential alteration of splicing. The variant occurs in a deep intronic region and the results of splicing and auxiliary sites that could be created by the variant are shown.

Using the HSF (version 3.0) online tool, the activation of the predicted cryptic donor splice site at position -301 is shown in a graphical schematic (Figure 5-14). At position -301, the HSF matrix score totalled approximately 70 (70.88), which at >65 (the consensus value threshold), signified the creation of a mutant donor splice site.

The WT score at the same position was below the consensus value threshold at 40.97 (Figure 5-14).

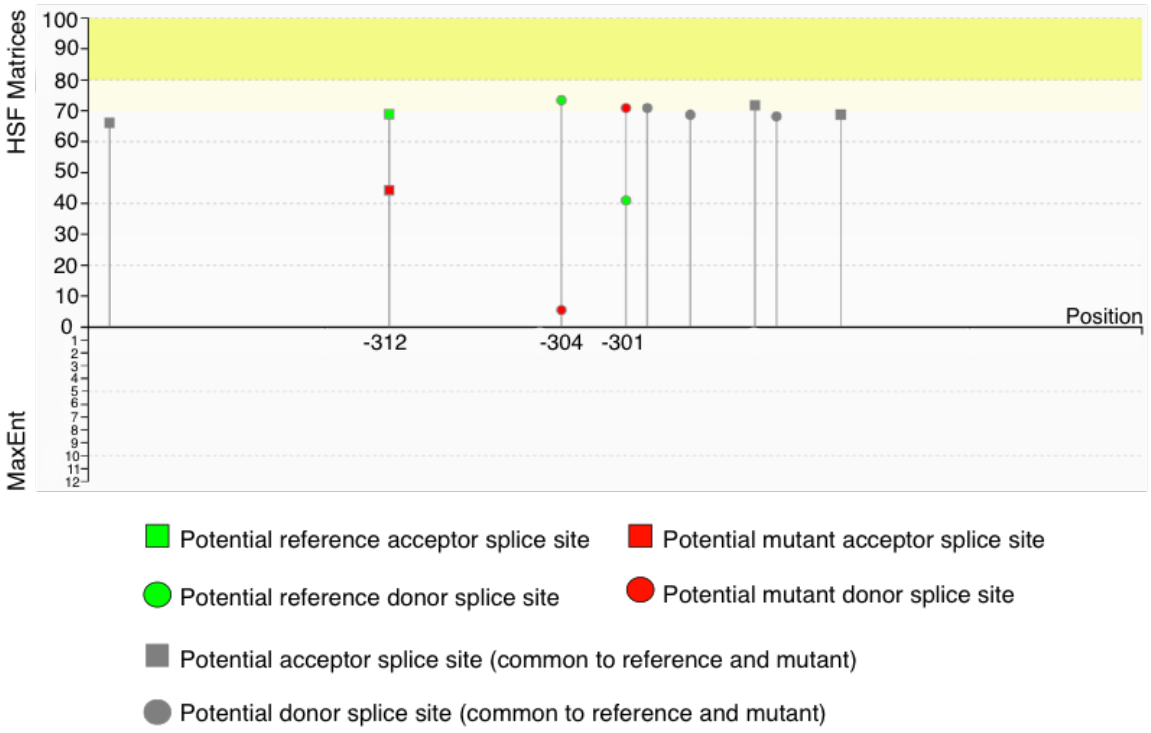


Figure 5-14 HSF graphical representation of the *SLC38A8* intronic variant c.691-301delG identified in Case F1377.

The introduction of a potential cryptic splice donor site at position c.691-301delG in Case F1377. The potential reference donor site has a HSF Matrix score of 40.97 compared to the HSF matrix score for the variant at 70.88. The representation was generated using the HSF online tool (Section 2.4.5).

5.2.3.3 Assessing the effect of potential intronic mutations using a 'midigene' multi-exon splicing assay

As *SLC38A8* is not widely expressed, and no patient RNA samples were available, the functional effect of the heterozygous *SLC38A8* intron 5 variant (c.691-301delG) was tested using an *in vitro* splicing assay system (Section 2.3.23). The splicing vector pCI-NEO-*RHO*_ex3-5/DEST was used to create a multiexon midigene splicing construct for *SLC38A8* encompassing the variant of interest. As the insert

sizes for midigene constructs are large, bacterial artificial chromosome (BAC) clones are used to generate the insert by PCR and SDM, rather than the traditional method of amplifying the insert from patient DNA used for minigene constructs.

UCSC Genome Browser was used to identify BAC clones which contained *SLC38A8* using the BAC End Pairs track. Six potential clones were identified and RP11-486I19 was chosen. The BAC was provided as a glycerol stock in E.coli which was streak purified and single colonies were picked and grown up before their DNA was isolated (Section 2.3.20). Primers were designed to amplify 7,484 bp of *SLC38A8* encompassing exons 5-6 (chr16:84,056,048-84,063,532) so that the c.691-301delG variant was flanked by at least one upstream and one downstream *SLC38A8* exon. As detailed in Section 2.3.23, the BAC DNA was amplified using attB-tagged primers (*SLC38A8*_BAC-1F and *SLC38A8*_BAC-1R), using a proof-reading, high-fidelity DNA polymerase (HiFi HotStart Polymerase) (Appendix 15). The PCR product was purified (Section 2.3.17). It was then sequence verified using Sanger sequencing analysis (Section 2.3.9) to ensure it was full-length (encompassing *SLC38A8* chr16:84,056,048-84,063,532, GRCh37) and contained no variants in the coding sequence. Six intronic variants were identified in the BAC and all were common on gnomAD and therefore likely to be non-pathogenic (Table 5-3).

Location on Chr16	dbSNP rs ID	Variant	Allele count & frequency (gnomAD)	No. of reported homozygotes (gnomAD)
84,063,419	rs4782578	c.633-264C>T	9118/30912 0.2950	1376
84,063,214	rs11863267	c.633-59C>T	3183/30940 0.1029	170
84,056,886	rs9319450	c.691-393G>A	4025/30930 0.1301	687
84,056,856	rs11861419	c.691-363T>C	1992/30870 0.06453	214
84,056,746	rs11865720	c.691-253A>G	8291/30814 0.2691	1125
84,056,660	rs11861366	c.691-167G>A	7044/30908 0.2279	853

Table 5-3 Heterozygous non-pathogenic variants identified in the BAC clone *SLC38A8* locus.

Non-pathogenic variants identified by Sanger sequencing the coding regions of the *SLC38A8* BAC clone (RP11-486I19). Variants annotated using the reference sequence NM-001080442.1 and human reference genome build GRCh37. For all SNVs, rs number (dbSNP rs ID), allele count & frequency and the number of homozygotes reported are listed using data resourced from the gnomAD browser (last accessed 12/12/2017).

A Gateway entry clone for *SLC38A8* was created by transferring the attB-tagged PCR product into a donor vector (pDONR201) using BP clonase (Section 2.3.17.1). The size of the resulting pENTR_*SLC38A8* construct totalled 9.74 kb.

The Gateway-adapted destination midigene vector, pCI-NEO-*RHO*_ex3-5/DEST, was provided as a kind gift from Erwin van Wijk (Radboud University Medical Centre, Nijmegen) (Appendix 11). The recombination sequences in this destination vector were located between rhodopsin (*RHO)* exons 3 and 4 and between exons 4 and 5. This destination vector was Sanger sequenced prior to cloning to ensure the sequence was as expected. A single variant was identified in the backbone of the vector, in intronic *RHO* sequence 145 bp away from *RHO* exon 5 (Table 5-4). The

variant had a dbSNP identifier (rs2625964) but no allele frequency information was available from 1000 Genomes/gnomAD.

Location on Chr3	dbSNP rs ID	Variant	Allele count & frequency (gnomAD)	No. of reported homozygotes (gnomAD)
129,252,305	rs2625964	c.937-145T>A	-	-

Table 5-4 SNP identified in the pCI-NEO-*RHO*_ex3-5/DEST construct.

Homozygous variant identified after Sanger sequencing of the pCI-NEO-*RHO*_ex3-5/DEST vector. The location on chromosome 3 in the *RHO* gene and the dbSNP rs ID number are shown. The variant was not present in the 1000 Genomes or the gnomAD database of control genomes (last accessed 12/12/2017).

To clone the pDONR_*SLC38A8* into pCI-NEO-*RHO*_ex3-5/DEST, an LR reaction using LR clonase (Invitrogen) was undertaken (Section 2.3.17.2). The resulting pCI-NEO-*RHO*_ex3-5/DEST splicing vector containing WT *SLC38A8* exons 5-6 will herein be referred to as pCI-NEO-*RHO*_ex3-5_*SLC38A8*-WT. Following this, SDM was performed on the pCI-NEO-*RHO*_ex3-5_*SLC38A8*-WT clone to introduce the candidate intronic deletion variant (c.691-301delG) (Section 2.3.15.1), herein referred to as pCI-NEO-*RHO*_ex3-5_*SLC38A8*-Mut. The constructs were Sanger sequence verified and confirmed to have the desired region of *SLC38A8* inserted between the *RHO* exons 3 and 5 (Appendix 11) (Section 2.3.9). The final size of the destination constructs was 13.79 kb.

Both pCI-NEO-*RHO_SLC38A8*-WT and pCI-NEO-*RHO_SLC38A8*-Mut vectors were separately transfected into HEK293T and COS7 cells (Section 2.3.22, 2.3.23). 48 hours post-transfection, RNA extraction was performed (Section 2.3.23.1) and RNA was converted to cDNA using random primers (Section 2.3.23.1). The cDNA was

amplified using *p53* and *RHO* exon 5 primers (these primers did not span the *SLC38A8* insert). These served as both RT-PCR and transfection efficiency controls (Figure 5-15) (Primers listed in Appendices 17 and 18).

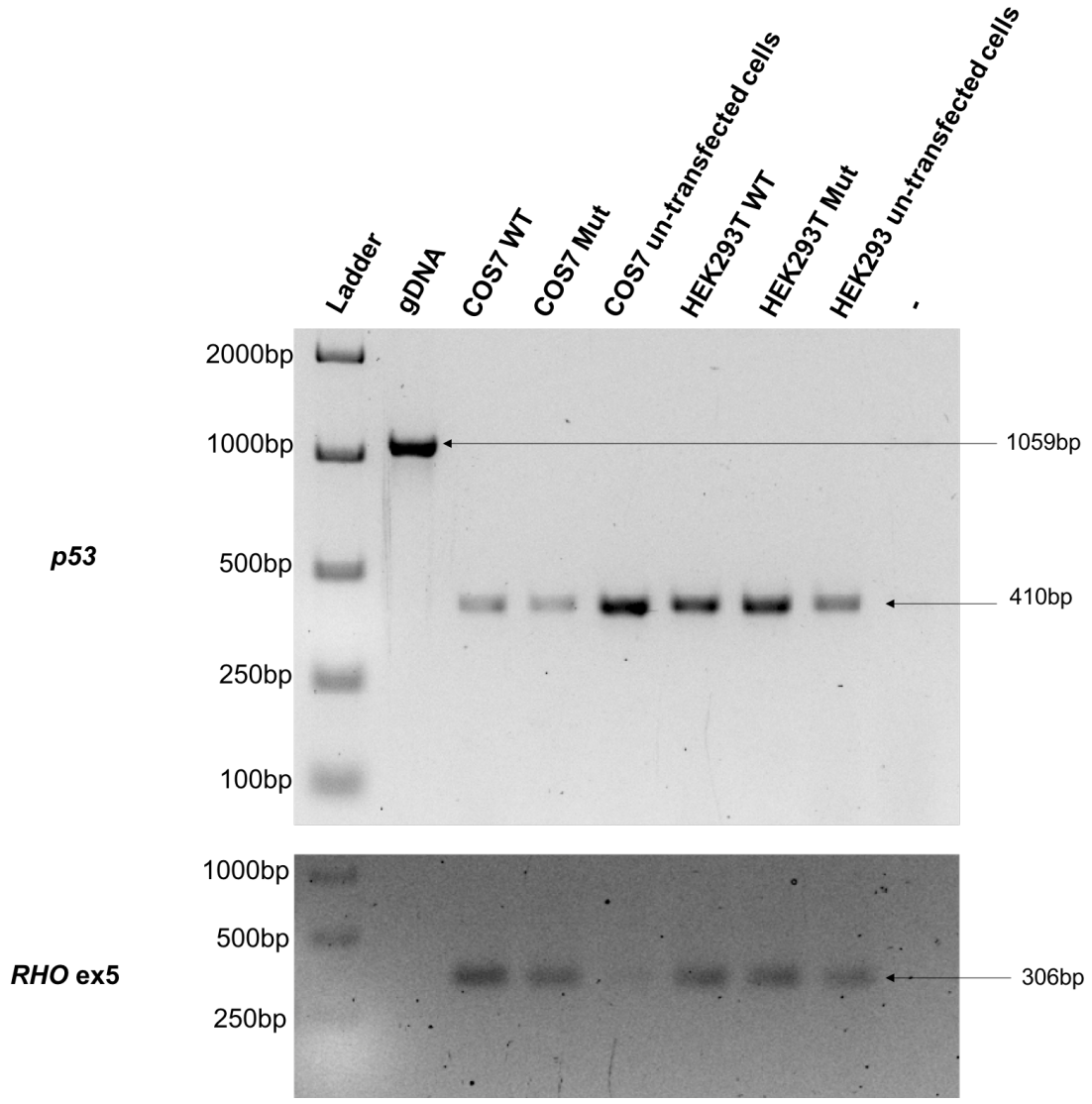


Figure 5-15 Amplification of *p53* and *RHO* exon 5 to assay cDNA quality and transfection efficiency.

RT-PCR using primers targeting *p53* (top pane) and *RHO* exon 5 (lower pane) were used as controls for transfection efficiency. Genomic DNA (gDNA) was used as a PCR control from a healthy control. Wild-type (WT), Mutant (c.691-301delG) (Mut) and untransfected cells are shown for both cell lines, COS7 and HEK293T.

Amplification of both the pCI-NEO-*RHO_SLC38A8*-WT and pCI-NEO-*RHO_SLC38A8*-Mut midigene cDNA was performed using multiple combinations of *SLC38A8* and *RHO* primers. RT-PCR was used to assay across the entire midigene *SLC38A8* insert using the primers designed in *RHO* exons 3 and 5 (Appendix 16). The primer *RHO*_exon3F was previously published in a similar midigene experiment assaying *ABCA4* intronic variants (Sangermano *et al.*, 2017). The splicing pattern predicted for pCI-NEO-*RHO_SLC38A8*-WT in both COS7 and HEK293T cells was *RHO* exon 3 > *SLC38A8* exon 5 > *SLC38A8* exon 6 > *RHO* exon 5. This would amplify a PCR product of 306 bp, as presented in the schematic (Figure 5-16).

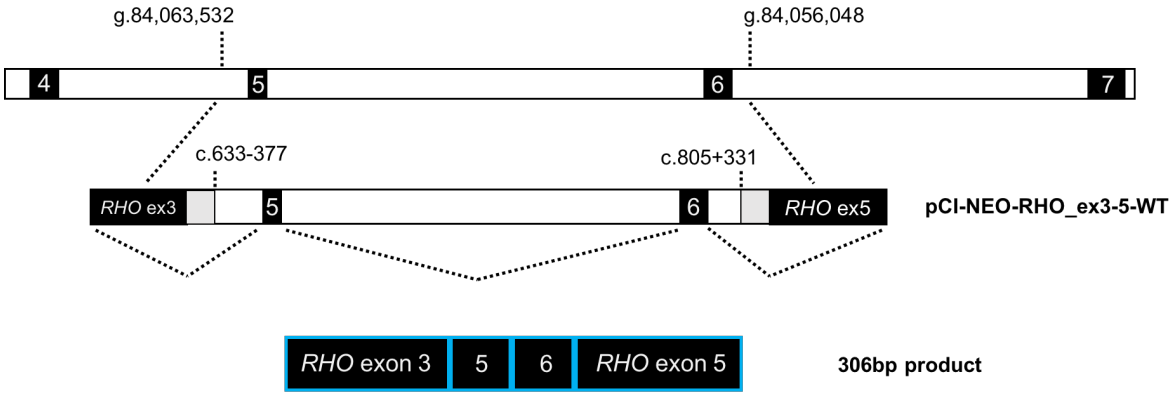


Figure 5-16 Schematic representation of the *SLC38A8* midigene generated in the study.

Schematic representation of the *SLC38A8* genomic region amplified from the BAC clone and the structure of the WT midigene generated using Gateway cloning technology.

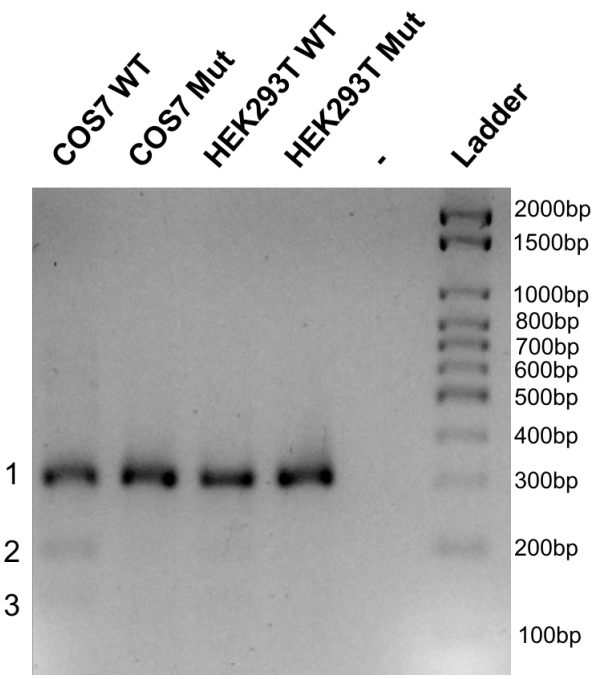


Figure 5-17 Amplification of cDNA to assess splicing in the *SLC38A8* WT and mutant midigene constructs.

RT-PCR using primers designed in the *RHO* exons 3 and 5 amplifying across the insert of WT (pCI-NEO-*RHO_SLC38A8*-WT) or c.691-301delG mutant expression construct (pCI-NEO-*RHO_SLC38A8*-Mut). The predicted WT splicing product of 306 bp (1) was also observed for all clones in both cell types. Smaller bands of approximately 200 bp (2) and 150 bp (3) were also visible in WT construct cDNA in COS7 cells.

The mutant, c.691-301delG, located near *SLC38A8* exon 6, was predicted to alter splicing by activating a cryptic intronic splice donor site, causing exon elongation by 289 bp (HSF v3.0 online tool). RT-PCR of pCI-NEO-*RHO_SLC38A8*-Mut in both COS7- and HEK293T-extracted cDNA generated a product of approximately 306 bp. This result was as observed in the amplification of the pCI-NEO-*RHO_SLC38A8*-WT construct indicating that no aberrant splicing was occurring (Figure 5-17, band 1).

To exclude the possibility that more than one splicing outcome had occurred, and a larger product missed, PCRs were also performed with insert-specific intronic

primers, amplifying across the insert as follows using intronic primers: *SLC38A8*_5F > *SLC38A8*_6R, *RHO* 3F > *SLC38A8*_5R and *SLC38A8*_6F > *RHO*_5R (Appendix 16). Amplification of cDNA from both WT and Mutant COS7 and HEK293T cell lines using these internal RT-PCR primers was unsuccessful.

5.2.3.4 Nested PCR to examine *SLC38A8* expression in blood and fibroblasts

The ideal way to determine if this c.691-301delG intronic SLC38A8 variant caused aberrant splicing was to analyse mRNA from case F1377. No patient RNA samples were available for F1377 but before re-sampling the patient, an experiment was performed to check if "illegitimately" transcribed *SLC38A8* transcripts could be amplified in RNA extracted from blood or fibroblasts using nested PCR. Previously published primers spanning *SLC38A8* exons 7-8 (Appendix 1) were used to amplify a region of the gene in whole blood, fibroblast and retinal RNA (as a control). Amplification of *SLC38A8* failed in both fibroblast and retinal cDNA and was extremely faint in whole blood cDNA. A nested PCR was performed to amplify a smaller product of 81 bp in whole blood cDNA, using the 188 bp as the template. The results showed that *SLC38A8* was successfully amplified in whole blood cDNA using this method (Figure 5-). Therefore, when this patient is sampled, the *SLC38A8* transcript can be amplified using nested PCR to look for splicing mutations in a blood sample. Further experiments are also planned to examine the *SLC38A8* locus in more detail and to expand the search area to look at epigenetic modifying mutations, for example.

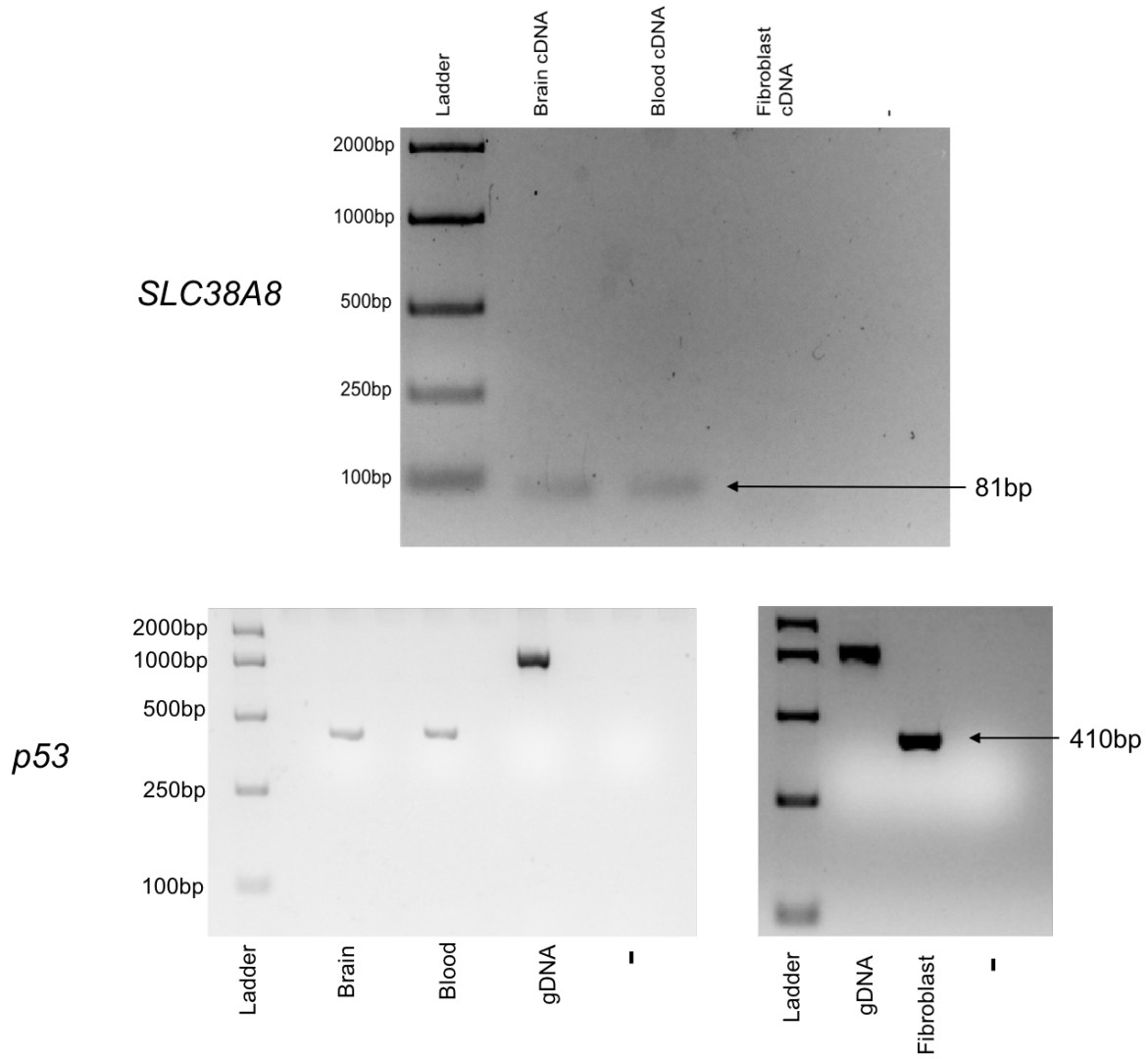


Figure 5-18 Nested PCR amplification of *SLC38A8* mRNA in brain and leukocyte cDNA.

SLC38A8 expression was visible in blood cDNA using nested PCR primers (upper pane), yielding an 81 bp fragment. *p53* expression (lower panes) was tested in all cDNA (primers listed in Appendix 1).

5.2.4 Whole-genome analysis of Case F1335

In Chapter 4, CNV analysis using ExomeDepth revealed a suspected homozygous duplication CNV in exon 6 of *SLC38A8* in patient F1335. In order to confirm this CNV and refine the breakpoints, WGS was undertaken.

To globally investigate the region of the predicted duplication and interrogate the aligned paired-end reads, the aligned WGS BAM file was uploaded to IGV. The use of the IGV tool also permitted examination of the quality of the coverage over the locus and comparison of the WES and WGS sequencing reads (Figure 5-19). Using WGS to more accurately define the precise breakpoints failed to present a result as obvious as originally hypothesised. The double peak on either side of exon 6, as visualised in the WGS data in IGV (Figure 5-19), suggested that the duplicated region is incomplete (that some of the sequence is perhaps deleted). Closer investigation of the region of the duplication showed that the patient harbored a heterozygous common dinucleotide deletion located downstream of exon 6 (c.805+53_805+54delAA), at position chr16:84,056,326-84,056,327 (Figure 5-19). The deletion, present in dbSNP (rs5818479), has a reported MAF of 0.23 in the 1000 Genomes dataset. To try and confirm the duplication and define the breakpoints, primers flanking the region of the duplication were designed, however, optimisation of the assay was unsuccessful.

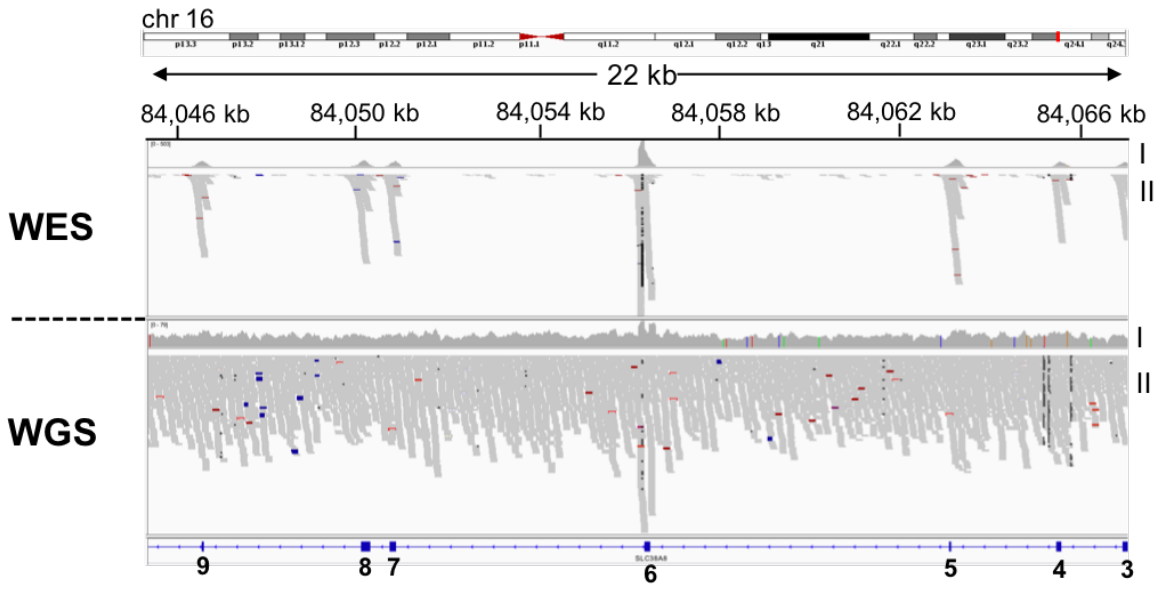


Figure 5-19 IGV representation of the putative homozygous duplication CNV identified in F1335 using WES and WGS.

IGV plots of WES and WGS BAM file sequencing reads of patient F1335. (I) The coverage track above the reads represents depth of coverage for all of the reads. (II) The grey bars represent individual sequence reads. Global representation of the duplication shows an increase in the depth of coverage over *SLC38A8* exon 6 in both WGS and WES data.

In order to further investigate the breakpoints, discordant reads from the full data set were marked using SAMBLASTER (version 0.7.12). It is evident from Figure 5-20 that there are no discordant paired end reads or split reads.

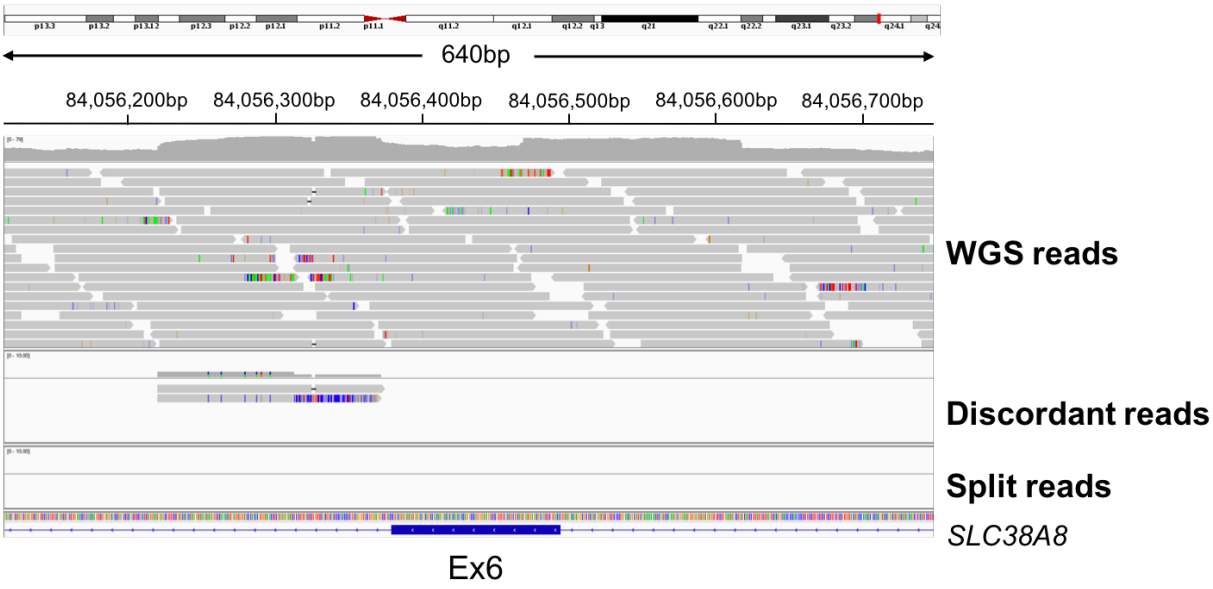


Figure 5-20 No evidence of discordant reads in the region of the *SLC38A8* duplication identified in patient F1335.

Discordant reads were isolated from WGS data using SAMBLASTER and uploaded to the IGV. There was no evidence of discordant reads in the region of increased reads. The two discordant reads shown are artefacts.

In collaboration with Dr Chris Watson (Yorkshire Regional Genetics, St James's Hospital, Leeds), work is ongoing to perform *de novo* assembly over the *SLC38A8* locus, as an alternative to reference genome alignment.

5.2.5 Investigation of remaining *SLC38A8* mutation-negative patient genomes

At the end of the project, a further 7 individuals from the *SLC38A8* mutation-negative cohort (F1071, F1298, F1352, F1369, F1374, F1403, F1404) underwent WGS to investigate if they contained undetected SV at the *SLC38A8* locus, as this locus might be susceptible to these types of mutations. The BAM files of these patients were analysed in IGV, but no SV were identified.

To look for additional variants which may have been missed by WES, the genomes were subjected to Alamut Batch analysis (version 1.5.2) (http://www.interactive-biosoftware.com/alamut-batch/) through a collaboration with Dr Chris Watson (Yorkshire Regional Genetics, Leeds, UK). Variants in the *SLC38A8* region of interest (chr16: 84,036,380-84,087,365, GRCh37) were pulled out of the genome VCF file using the BED format <chromosome, start, end>. Variants were then automatically annotated using a range of splice site prediction software including HSF, alongside MaxEntScan, SpliceSiteFinder, GeneSplicer and NNSPLICE (0.9). Alamut Batch interprets raw splice site recognition in the variant vicinity (local splicing effect predictions) and also predicts a score at the nearest splice site. The local splicing effect predictions are annotated by MaxEntScan, NNSPLICE and HSF in this interpretation algorithm. In contrast, for splicing predictions at the nearest splice site, only the raw prediction score data is provided and there is no interpretation of effect.

The total number of candidate heterozygous and homozygous variants are presented in Table 5-5. None of the patients had any unidentified non-synonymous variants in *SLC38A8*, confirming the results of Sanger sequencing in Chapter 3 and WES in Chapter 4 (for the cases put forward). There was no potential splicing effect of the synonymous variants identified. No patients had variants in the 5’ UTR region identified in Section 5.2.3.1.

Patient	Total rare variants	Intergenic variants	Intragenic variants coding		Intronic variants	Rare intronic variants (MAF 0.0001)	Rare intronic variants with potential splicing effect
			Synonymous	Non-synonymous			
F1071	51	37	1	0	13	7	0
F1298	49	33	0	0	16	10	0
F1352	87	53	0	0	34	8	1
F1369	81	37	0	0	44	6	0
F1374	66	41	0	0	25	5	1
F1403	45	30	1	0	14	8	1
F1404	53	31	0	0	22	6	0

Table 5-5 Summary of all rare variants identified using Alamut Batch in the *SLC38A8* locus.

Filtering of Alamut Batch output retained all variants with a MAF 0.0001 in the *SLC38A8* locus (chr16: 84,036,380-84,087,365 (GRCh37).

One unreported intronic variant was identified which potentially altered the WT splice site (Table 5-6). The intron 4 variant identified in individual F1352; c.633-770A>G, chr16: 84,063,926, was not listed in the dbSNP, ExAC or gnomAD publicly available databases of variation (last checked 12/12/2017). The variant was located -720bp away from the nearest 3' splice site and was predicted to introduce a novel splice donor site at position chr16: 84,063,927. The local splice site variant Max Ent score was recorded at 8.34 and local splice site variant HSF score at 85.39. The WT splice site motif 'aggatatgt' was predicted to change to the novel splice donor site 'AGGgtatgt', with the cryptic site introducing a variation of exon length of +761 bp. Alongside ongoing work to analyse the WES data (Chapter 4), work is underway to analyse this rare deep intronic variant using the midigene assay.

Patient	Position	Variant	dbSNP rs ID	Allele frequency data (gnomAD)	Splice prediction	MaxEnt score
F1352	16:84,063,926	c.633-770A>G Het	.	.	New donor site	8.34

Table 5-6 Unreported splice variant identified in *SLC38A8* in the WGS Alamut Batch analysis of all mutation-negative cases.

Unreported variant identified in the *SLC38A8* locus annotated using the location on chromosome 16 (reference sequence NM_001080442.1), human reference genome build GRCh37.

5.3 Discussion

The fundamental aim of the work undertaken in this chapter was to expand the mutation search of *SLC38A8* and interrogate the genomic landscape of three key foveal hypoplasia patients. Both SVs and mutations in non-coding regions were investigated. WGS analysis of individual F1310 revealed a homozygous inversion spanning exons 6, 7 and 8 of *SLC38A8*. Further investigation confirmed this to be a 11,101 bp paracentric inversion encompassing three *SLC38A8* exons and flanking intronic sequence. PCR and electrophoresis indicated that the inversion was not present in the remaining *SLC38A8* mutation-negative cohort. In Chapter 4, ExomeDepth analysis of Case F1335 identified a potential homozygous duplication, predicted to duplicate exon 6 of *SLC38A8*. WGS was applied to try and confirm the duplication and accurately characterise the breakpoints, but these have yet to be resolved.

Case F1377 was first studied in Chapter 3, whereby Sanger sequencing of *SLC38A8* revealed a heterozygous missense variant in the start codon of the gene, but failed

to reveal a second mutation. Using a WGS pipeline employing vcfhacks, prioritisation of all rare, heterozygous variants in the *SLC38A8* locus identified a single intronic base-pair deletion. The deletion was predicted to introduce a mutant splice donor site approximately 301 bp away from exon 6 of *SLC38A8*. However, functional assessment of this variant, using an *in vitro* midigene assay, failed to reveal any changes in splicing pattern compared to WT. The search for the second mutation in this patient is thus ongoing.

WGS analysis of 7 additional *SLC38A8* mutation-negative individuals was performed, but global analysis of the *SLC38A8* locus in IGV in these patients did not reveal any obvious SVs. Alamut Batch analysis of variants in the *SLC38A8* locus revealed one candidate intronic variant in patient F1352, however further investigation is required to confirm if this alters splicing, and to find the second variant in this case.

5.3.1 WGS analysis in studies of IRD

The genetic basis of IRD currently remains unknown in 20-50% of NGS-investigated cases (Audo *et al.*, 2012; Neveling *et al.*, 2012; Lee *et al.*, 2015; Weisschuh *et al.*, 2016). Mutations in novel genes that have yet to be ascertained could explain this low detection rate. In the study so far, this rationale has been investigated by using WES to explore the potential for genetic heterogeneity in foveal hypoplasia cases, as discussed in Chapter 4. In the present chapter, the research focused on the potential for mutations undetectable by conventional methods in the known gene *SLC38A8.* The implementation of WGS was hypothesised to increase the mutation detection

rate. So far only one foveal hypoplasia case has been solved (F1310 with the inversion). However, despite this, the application of WGS provides the opportunity to check for SV in other foveal hypoplasia genes and the albinism genes. Ultimately, the application of WGS will help to facilitate a confident approach when investigating novel rare genes as nothing will be missed due to the technical limitations of WES.

Although WES is the current dominant method for detecting causal alleles in Mendelian disease (Bamshad *et al.*, 2011), the employment of WGS has been proven to identify disease-causing mutations in non-coding regions or SV. A number of studies support the view that WGS is more reliable for CNV detection, given the fact the coverage isn't limited by non-contiguous analysis (Meynert *et al.*, 2014; Belkadi *et al.*, 2015). Interrogation of intronic and intergenic nucleotides in Mendelian disease is now being driven through large-scale studies, such as the Genomics England 100,000 Genomes Project (https://www.genomicsengland.co.uk/the-100000-genomes-project/). The employment of WGS, albeit on a smaller scale than WES at present, has so far demonstrated many advantages, including higher diagnostic yield than conventional methods of genetic testing (Gilissen *et al.*, 2014; Stavropoulos *et al.*, 2016). WGS is also more effective than targeted capture arrays (Ellingford *et al.*, 2016).

The motivation for interrogation of *SLC38A8's* genomic landscape was justified by three factors:

i. The discovery of a potential duplication in patient F1335 using WES in Chapter 4;

ii. The inability to identify a mutation in the coding region and splice sites of individual F1310, known to be homozygous across the locus;

iii. The fact that only one mutation was identified in *SLC38A8* in Case F1377.

5.3.2 Structural variation at the *SLC38A8* locus

Genomic rearrangements of regions >50 bp, known as SVs, encompass changes such as deletions, duplications, insertions, inversions, translocations and complex rearrangements. The complexity of SVs is defined by their vast differences in size and genomic architecture. These characteristics often lead to issues in the accuracy of read alignment to the control genome, due to the introduction of novel breakpoints in the affected genome and the short read length typical of NGS platforms (Tattini, D'Aurizio and Magi, 2015).

5.3.2.1 A structural inversion in *SLC38A8* identified by WGS

Using the IGV software in the absence of genetic SV, the orientation of read pairs is expected to display as forward-reverse. However, in the case of individual F1310, clusters of forward-forward and reverse-reverse pairs indicated an inverted region of 11,101 bp (Figure 5-2, Figure 5-21). In this instance, the location of the SV breakpoints was easily identified, located at g.84,049,578_84,060,679inv (GRCh37). The IGV is able to colour-code the paired-ends if their insert sizes are larger than expected, are located on different chromosomes (inter-chromosomal translocation events) or have abnormal pair orientations. The colour-coding is indicative of potential genomic rearrangements or potential misalignments, which may have a high occurrence in repeat regions. The use of the IGV to globally visualise raw

genome sequencing reads has been successful in a number of IRD-related studies, including the identification of a 55 kb deletion in the gene *EYS* in a patient with RP (Carss *et al.*, 2017). However, this is only useful if there is a gene of interest to focus the search on.

The use of BreakDancerMax failed to map the precise breakpoints of the inversion, whereas DELLY2 succeeded. BreakDancer relies solely on read pair analysis spanning the SV event (Chen *et al.*, 2009), rather than applying a combined approach, such as DELLY2 which analyses discordant reads pairs first and then uses split read analysis to help ascertain the rearrangement (Rausch *et al.*, 2012). A number of false-positive paired-end alignments have been suggested to decrease the accuracy of tools such as BreakDancer (Bartenhagen and Dugas, 2016). SV detection tools usually adopt one of four main approaches, including paired-end analysis, read-depth analysis, split-read analysis or sequence assembly methods. Many algorithms exist that detect SVs and successfully map their breakpoint junctions, such as LUMPY (Layer *et al.*, 2014) which could be tested in the future.

SVs, such as the inversion detected in F1310 can be categorised as “balanced” rearrangements that do not alter changes in DNA dosage, despite altering the orientation of localisation of a genomic segment.

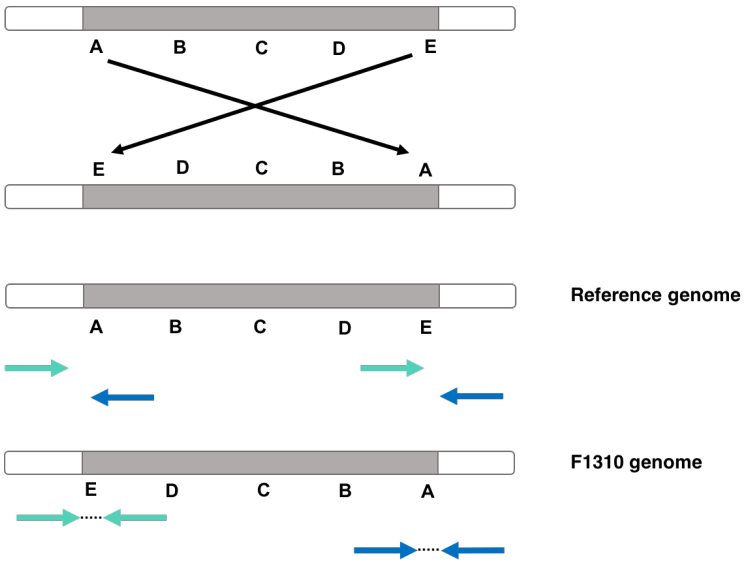


Figure 5-21 Schematic representation of the inverted sequencing reads identified using the IGV in analysis of patient F1310.

The balanced inversion detected in F1310 would not have been identifiable through array-CGH, which solely detects abnormal copy-number, not balanced rearrangements. WGS is a much easier, faster and more efficient approach than older methods, such as vectorette PCR applied in the identification of the complex rearrangements described for the *Factor VIII* gene found in patients with haemophilia A (Naylor *et al.*, 1993).

5.3.2.2 Potential pathogenicity of chromosomal inversions

SVs, such as inversions in the human genome may arise by a variety of mechanisms including non-allelic homologous recombination (NAHR) (Section 5.3.2.3). Interrogation of the genomic sequence flanking the two inversion breakpoints identified in F1310 unveiled a 42 bp palindromic sequence with 100% sequence identity located at the inversion junctions. Inverted repeats or palindromes have been put forward as “hotspots” for both recurrent and non-recurrent chromosomal

rearrangement events (Inagaki *et al.*, 2013). Palindromic sequences have been reported to mediate inter-strand or intra-strand mispairing leading to serial replication slippage (SRS) and fork stalling and template switching (FoSTeS) and microhomology-mediated break-induced replication (MMBIR), with single-stranded hairpin or double-stranded cruciform structures forming on account of intra-strand palindrome annealing (Hastings *et al.*, 2009; Chen *et al.*, 2010).

Research has suggested that most apparent balanced chromosomal inversions are clinically asymptomatic, as the overall amount of genetic material is not altered (Feuk, Carson and Scherer, 2006). The majority of studies detail imbalanced rearrangement CNV, such as deletions and duplications, in human disease. Inversions, however, have been associated with human disorders when breakpoints fall within or near a gene associated with a specific phenotype. They have also been reported to be indirectly causative, and suggested to increase the occurrence of additional rearrangements, such as deletions, that will cause disease. Such examples of recurrent inversions include the disruption of the *idunorate 2-sulphatase* gene in Hunter syndrome (Bondeson *et al.*, 1995) and an inverted repeat-mediated inversion in the region of the *emerin* gene on the X chromosome causing Emery-Dreifuss muscular dystrophy (Small, Iber and Warren, 1997). Balanced chromosomal abnormalities have also been reported to indirectly affect a number of genomic loci, with the breakpoints affecting genes with important roles in neurodevelopment and brain function (Talkowski *et al.*, 2012; Blake *et al.*, 2014). In particular, a 1.8 Mb inversion on chromosome 15 has been reported to predispose individuals to microdeletions in *CHRNA7* (Sharp *et al.*, 2008; Shinawi *et al.*, 2009; Kidd *et al.*, 2008).

5.3.2.3 Using WGS to define the breakpoints of a putative duplication in *SLC38A8*

In Chapter 4, ExomeDepth analysis revealed a potential homozygous duplication, predicted to encompass *SLC38A8* exon 6. In IGV the presence of an exon duplication event was steadfast, as there was a definite increase in the paired-end read depth when analysing the WGS data. The integration site, however, could not be localised. It was hypothesized that the CNV breakpoints could lie outside of the regions targeted by the WES capture reagents.

Studies of reported CNVs in IRDs have highlighted a high correlation between gene size and occurrence of CNVs, with *EYS*, *PCDH15* and *USH2A* being the largest (Van Schil *et al.*, 2017). In comparison to these three genes, sized at approximately 1.98 Mb, 1.82 Mb and 0.8 Mb, respectively, *SLC38A8* is significantly smaller in size at 32.97 kb. The absence of *SLC38A8* from the list of CNV-prone IRD genes generated in the study by Van Schil and colleagues reinforces the rarity of the disorder.

Although a range of algorithms have been developed to detect CNV breakpoints, there is not a “best” method or program. The extraction of discordant read pairs using SAMBLASTER failed to reveal a large cluster or a peak of discordant reads, as interpreted as breakpoints in other studies (Cipriani *et al.*, 2017). Discordant reads were defined as reads that mapped too far apart/close together, were orientated in the wrong direction, or mapped to different chromosomes. Other studies of

duplications have employed the use of CIGAR scores to determine the mapping of split reads (Newman *et al.*, 2015), or a combination of coverage changes, split reads and chimeric reads in order to infer breakpoint junctions (Cipriani *et al.*, 2017). The breakpoints were not defined using the bioinformatics analysis applied in this study. The duplication was also analysed by Dr Chris Watson who was also not able to map the breakpoints. Similar unique read coverage profiles resulting from sequence context have been identified in other studies (Personal Communication, Dr Chris Watson). It is possible that the aligner discards particular reads, or the sequencer fails to read through low complexity bases. A number of read pairs were found to share start sites in the region of the duplication (Figure 5-22). The fact that the common 2 bp deletion identified in F1335 is heterozygous in a homozygous region suggests a duplication is present. As the breakpoints of the duplication have not been captured, this suggests that the region is a CpG island that is difficult to sequence, in a similar vein to the troubles sequencing *FOXD1* (Chapter 4). Long-read sequencing is planned for case F1335, in order to identify the duplication at nucleotide resolution. Workflows such as the Cas9-Assisted Targeting of CHromosome segments (CATCH) method (Gabrieli *et al.*, 2017) and the MinION (Jain *et al.*, 2016; Srivathsan *et al.*, 2018) are being considered. The PacBio system is also being considered, combining the sequence-specific endonuclease activity of the CRISPR/Cas9 system with long-read SMRT sequencing (Tsai *et al.*, 2017).

The data for patient F1335 looks convincing and is supportive of this being a solved case. However, the breakpoints of the proposed duplication need identifying for diagnostic purposes and this will hopefully be facilitated by the long-read sequencing.

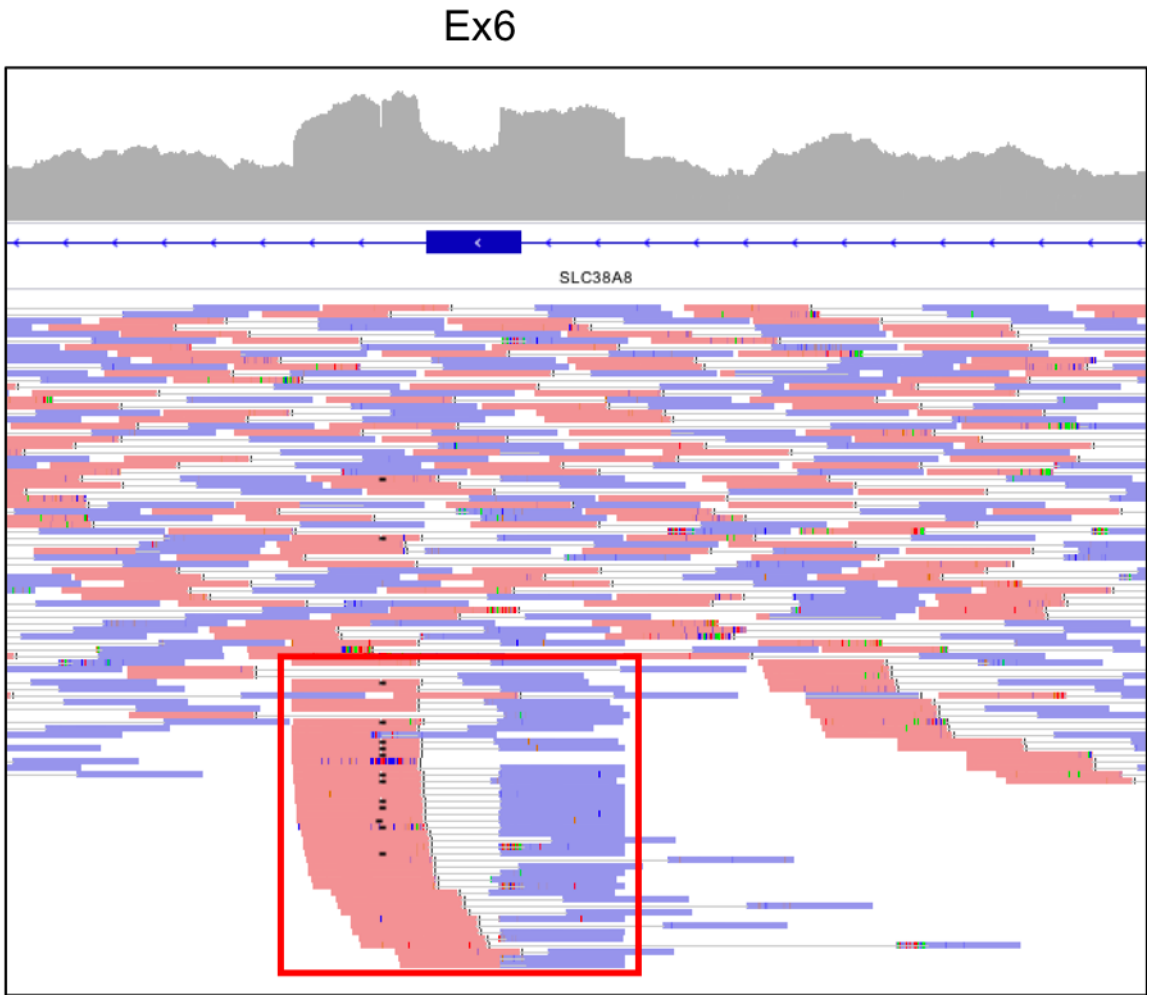


Figure 5-22 Similar start sites of read pairs in the region of the predicted duplication.

A fraction of the WGS read pairs share start sites in the region of the predicted duplication CNV over *SLC38A8* exon 6 (highlighted in the red box) when visualised in IGV.

5.3.2.4 Characteristics of CNVs

Most of the recurrent (common) genomic rearrangements have been proposed to arise through NAHR (Shaw and Lupski, 2004; Stankiewicz and Lupski, 2002). Genomic rearrangements that are recurrent with a common size are characterized by clustered breakpoints. Contributing to their formation are predominantly low copy repeats (LCRs), such as segmental duplications, and retrotransposons, such as long

interspersed elements (LINEs), with the rearrangement breakpoints clustered inside such repeats (Startek *et al.*, 2015; Cardoso *et al.*, 2016; Carvalho and Lupski, 2016). The most frequent microdeletion syndrome, DiGeorge (MIM 188400) is mediated by LCRs on 22q11.2 that are 240 kb and share 99.7% sequence identity (Halford *et al.*, 1993; Edelmann *et al.*, 1999; Edelmann, Pandita and Morrow, 1999).

Genomic rearrangements that are non-recurrent often share a common genomic region of overlap, known as the smallest region of overlap (Gu, Zhang and Lupski, 2008). Such rearrangements are believed to be formed as a result of non-homologous end joining (NHEJ) events occurring at sites of limited homology predisposing more complex rearrangements through mechanisms such as FoSTeS/MMBIR (Lee, Carvalho and Lupski, 2007; Carvalho *et al.*, 2009). Non-recurrent duplications have been observed in the *PLP1* region in Pelizaeus-Merzbacher disease (MIM 312080) (a rare CNS disorder). The breakpoints of the *PLP1* duplications have been reported to not cluster and yield duplicated genomic segments of varying lengths, compared to LCR-mediated NAHR generating recurrent rearrangement (Lee *et al.*, 2006).

The genomic landscape flanking the predicted duplication junctions were investigated to look for repeats using data from the Genome Browsers Ensembl and UCSC, including tracks from RepeatMasker (http://www.repeatmasker.org). In the region of the predicted duplication, repetitive elements, ranging in size from 8-487 bp, were identified. The majority of the repeat elements recorded were high-copy repeats, including LINEs and SINEs (*Alu* elements), endorsing reports that

approximately 44% of the human genome is covered by these common repeats (Chen *et al.*, 2014b) and an LTR endogenous retroviral sequence 1 (ERV1) element spanning the region of the 5' predicted duplication breakpoint. Narrowing the analysis to repeats present in the intronic regions immediately surrounding exon 6 highlighted two SINEs (chr16:84,056,077-84,056,324 and chr16:84,055,813-84,056,041 (intron 6/9)) and one LINE (chr16:84,056,563-84,056,581).

The fact that both the potential duplication and inversion identified in *SLC38A8* affected exons 6, 7 and 8 may indicate that this region is susceptible to recurrent rearrangements as a result of genomic instability of the surrounding genomic landscape (Bailey *et al.*, 2002). Another case with compound heterozygous deletions in *SLC38A8* has been identified affecting exons 2, 7 and 8 (Personal Communication, Dr Kamron Khan).

5.3.3 Investigating "undetectable" mutations in *SLC38A8*

RLM-RACE (Ambion) was employed as an alternative to traditional RACE methods (Maruyama and Sugano, 1994; Schaefer, 1995), in the hope of amplifying only the cDNA from the 5' ends of full length, capped mRNA. The aim was to define the 5' UTR of *SLC38A8* and also screen for potential missing variants/additional exons in the gene in the large cohort of mutation-negative cases. Similar investigation of hidden genetic variation has been performed by Coppieters and colleagues. In their studies of the gene *NMNAT1*, no coding-region mutations were identified in the gene to explain the Leber congenital amaurosis (LCA) congenital retinal dystrophy phenotype (Coppieters *et al.*, 2013; Coppieters *et al.*, 2015). Important in the

modulation of transport of mRNAs out of the nucleus, translation efficiency, mRNA stability and subcellular localisation, UTRs are vital for post-transcriptional regulation of gene expression (van der Velden and Thomas, 1999; Jansen, 2001; Bashirullah, Cooperstock and Lipshitz, 2001). The length of complete 5' UTRs ranges from 18-2,803 bp, with an average of 210bp in humans. They are known to be much shorter than 3' UTRs, which can extend to an average of 1,027 bp in humans (Pesole *et al.*, 2001; Mignone *et al.*, 2002). Remarkably, UTRs comprising of a single base-pair have been found to regulate translation in the unicellular organism *Giardia* using an *in vitro* system (Hughes and Andrews, 1997), highlighting that UTR size doesn't necessarily improve initiation of translation. Application of the Ambion RLM-RACE kit failed to yield a defined 5' product. It was especially difficult to believe any product that appeared from the nested PCR of 5' RACE products, as there were bands present for the minus-TAP control (for which the mRNA had an intact 5' cap structure meaning that adapter could not ligate). However, issues encountered with RLM-RACE troubleshooting were circumvented by designing primers in the presumptive 5' UTR and successfully sequencing ~290 bp of 5' UTR in adult human brain poly(A)+ cDNA. There is the potential that contaminating gDNA was sequenced, however all cDNA samples amplified in this experiment had been checked using a *p53* control and shown not to have contaminating gDNA.

SLC38A8 mRNA was analysed to investigate potential novel *SLC38A8* exons which might harbour mutations. These investigations were prompted by the published findings of the *RPGR* gene, for which one transcript, the *RPGR-ORF15* isoform, is expressed only in the retina. SLC38A8 protein is visible in retinal sections (Poulter *et al.*, 2013), but no *SLC38A8* expression was observed in the commercially available

retina cDNA screened in the current study so there was no opportunity to manually investigate if novel exons were present (Section 5.2.3.4). RNA-seq data was investigated, however this was not ideal as it was from human cornea samples and SHSY-5Y and RPE cell lines. Embryonic retina RNA would be a more suitable tissue to investigate, as the retina is an end-differentiated and non-renewable tissue. It is possible that the SLC38A8 protein is present in the adult tissue, but the mRNA is not.

When looking for splice variants using RT-PCR, the expression of *SLC38A8* in human whole blood was ascertained using a nested PCR with intron-spanning primers. It was anticipated that fresh blood samples would be collected from patients to enable RT-PCR mRNA-based screens to sequence and quantify intronic variants causing potential splicing defects or non-expression, or variants in UTRs. However, blood samples were not available. As only a small section of the gene was amplified, work is ongoing designing a full primer set to amplify the entire gene in blood samples in preparation for patient's samples. This was not trivial as it was important to ensure the PCR was robust, as inclusion of introns can result in the amplification of large products.

WGS was undertaken, focusing on case F1377 with the missing mutation in *SLC38A8*. Only rare, intronic variants were retained after filtering with vcfhacks pipelines. Only intronic variants outside of the canonical splice sites were identified in Case F1377. Such variants may activate cryptic splice acceptor or donor sites, altering the final mRNA transcript and in some cases causing premature termination

of proteins (Liu and Zack, 2013). Disease-causing variants have been found to reside in non-coding regions of many IRD genes, including *PROM1* (Mayer *et al.*, 2015), *ABCA4* (Braun *et al.*, 2013; Bauwens *et al.*, 2015), *USH2A* (Vache *et al.*, 2012; Steele-Stallard *et al.*, 2013) and *CEP290* (den Hollander *et al.*, 2006), suggesting that these non-coding mutations will have been missed in the past mutation screens. Difficulty evaluating the causality and functional impact of such non-coding or synonymous variation, however, is a major issue when investigating these variants (Weisschuh *et al.*, 2016). It is their lower evolutionary constraints that makes determining their magnitude of damage to a genomic region difficult for tools such as SIFT, PolyPhen-2 or CADD, to interpret (Makalowski and Boguski, 1998; Gelfman *et al.*, 2012). *In silico* analysis of the retained *SLC38A8* intronic variants using the HSF tool revealed one main candidate predicted to activate a cryptic intronic splice donor site, causing potential alteration of splicing.

In accordance with this result, it was hypothesised that the aberrant splicing resulted in the activation of a cryptic donor splice site in intron 5 of *SLC38A8*, predicted to extend exon 6 by 289 bp. To confirm this hypothesis, the ideal experiment would be to screen patient-derived RNA. Many studies of mutation effects on splicing have used methodology such as RT-PCR of the mRNA region including the affected intron location and flanking exons. Key examples have been the investigation of potential pathogenicity of deep intronic mutations in *USH2A* using fibroblast cDNA of patients with Usher syndrome (Steele-Stallard *et al.*, 2013) and amplification of illegitimate *COL2A1* transcripts in patient skin fibroblast cell lines in the investigation of retinal detachment (Spickett *et al.*, 2016). Unfortunately, patient RNA was not available so the decision was made to proceed with an *in vitro* splice assay using a sizeable

multi-exon midigene. The use of a BAC construct facilitated the generation of the multi-exon splice construct, whereby the variant of interest was flanked by an upstream and downstream exon of *SLC38A8*. The majority of published IRD minigene experiments have studied variants at intron-exon boundaries (Liu and Zack, 2013). The midigene construct has been reported to supersede the minigene assay, as its larger insert capacity has the ability to include multiple natural exons flanking the variant of interest. This allows *cis*-acting splice factor binding motifs to be retained, yielding a more representative result similar to splicing in patient cells (Sangermano *et al.*, 2017). The study by Sangermano and colleagues used the midigene assay to investigate non-canonical splice site variants in *ABCA4*, mostly situated at positions +3 to +6 downstream of exons and -14 to -3 upstream of exons. This study also suggested the suitability of the assay to test deep intronic variants, such as the variant tested here in patient F1377 (Sangermano *et al.*, 2017). This prompted the experimentation using the midigene construct.

The current study was limited in that there was only time to investigate one intronic variant. Furthermore, the results from the midigene experiments were only performed in HEK293T and COS7 cells and may have yielded more accurate results to reflect the disease phenotype in retina-specific cell lines. Studies have indicated that there are retina-specific splice factors, for example there are splice factors specific to the vertebrate retina, such as Musashi 1 (MSI1), which promotes the splicing of exons in photoreceptors (Murphy *et al.*, 2016). Human cell lines and animal models show differences in splicing, even in the same tissue type (Garanto, Duijkers and Collin, 2015). A key example is the recurrent deep intronic variant c.2991+1655A>G in intron 26 of *CEP290,* whereby the humanized mouse model

fails to represent the splicing defects observed in patients with LCA, making it clear that not all models will recapitulate the human *in vivo* splicing pattern (Garanto, Duijkers and Collin, 2015). This has also been seen in the human iPSC-derived optic cup model of LCA (Parfitt *et al.*, 2016). Other studies, such as the investigation of splicing in *COL2A1* have employed the use of immortalised lens epithelial cells for the expression of minigenes (Spickett *et al.*, 2016). Additionally, there is the possibility that deep intronic variants could affect *cis* or *trans*-acting factors, such as exonic and intronic splice enhancer and inhibitor motifs, that directly affect *SLC38A8* expression (Wang and Cooper, 2007; Sangermano *et al.*, 2017).

The lack of an observable difference between the splicing pattern in WT and the c.691-301delG intronic variant suggests that *in silico* prediction tools cannot be fully relied upon when assessing variants prioritised through WGS. This has also been corroborated by the midigene study by Sangermano et al., 2017. Reported in this study were variants that were not predicted to alter splice sites but did upon assessment, for example c.1937+13T>G, which actually caused a 12bp exon elongation and introduced a premature stop codon in 89% of transcripts (Sangermano *et al.*, 2017). Variants were also reported to have a predicted effect on splicing but did not upon performance of the midigene assay. For example, a rare variant (c.2919-10T>C) was found to cause exon skipping in 39% of the ABCA4 transcripts, while a rare (c.5585-10T>C) variant showed no splice defects. This suggests that a variant's impact depends on the sequence context, such as enhancers and motifs (Sangermano *et al.*, 2017).

For the c.691-301delG variant PCR product amplification was performed across the entire midigene using primers located in *RHO* exons 3 and 5. As discussed by Sangermano and colleagues, the *RHO* exon 3 splice donor site and *RHO* exon 5 splice acceptor sites are relatively strong (Sangermano *et al.*, 2017). The strengths of nearby splice acceptor and donor sites in *SLC38A8* could also overpower the deep intronic variant cryptic donor site activation. Continued investigation to optimise the intrinsic PCRs (using *SLC38A8* primers), is required to conclude the absence of splice defects for the c.691-301delG variant.

In the current study, a limitation of the cell-based assay is the fact that only one variant was analysed at a time. The possibility that other sequence variants, as identified later in the project after application of Alamut Batch, may cause splicing defects, cannot be excluded. It is unfortunate that Alamut Batch only became available at a later stage of the project. It is possible that more than one variant present in a patient may contribute to pathogenic splicing, with perhaps not all variants being rare in this scenario. Now the experiment has been successfully optimised, additional midigene constructs have been created by Phoebe Cambridge (MSc Student, The University of Leeds), ready to assay further variants. However, ideally, a fibroblast sample would be taken from the patient so that iPSC-derived retinal tissue can be generated and analysed for splicing defects and novel exons, for example as investigated in *ABCA4* (Albert *et al.*, 2018; Sangermano *et al.*, 2017). There is also the possibility to use CRISPR in a cell line with known expression of *SLC38A8*. As all splicing constructs are artificial they will not be as representative as the real gene. This methodology could also be used when investigating variants in distal promotors and enhancers.

In the remaining *SLC38A8* mutation-negative cohort, due to patient consent the analysis was restricted to the *SLC38A8* locus only. Further examination of the regions flanking *SLC38A8*, *PAX6, FRMD7* and the albinism genes will be required to search for unidentified SV or intronic variants. In the region of *SLC38A8*, patients F1403 and F1374 shared a rare intronic variant (c.691-1832C>T). A second variant in *SLC38A8*, (p.(Gly53=)) was previously reported for case F1403 in Chapter 3, but no second variant was identified in F1374 that was rare enough to be considered pathogenic. The intronic variant c.633-770A>G, was not reported in publicly available databases and was a key candidate variant warranting further investigation, although no second variant was identified in this patient.

5.3.4 Summary

In this chapter, it is clear that WGS has greater power over WES or targeted gene panels when detecting SVs and variants in non-coding, regulatory regions. Reported for the first time is the use of WGS to explore potential deep intronic causative variants in patients with foveal hypoplasia. At this point of the study evidence for intronic variants as a possible cause underlying foveal hypoplasia has not been concluded. Using WGS, only one definite inversion SV was identified in ten cases analysed. It is highly likely that new genes remain to be identified in the foveal hypoplasia cohort, rather than missed mutations in *SLC38A8*.

Chapter 6 General discussion

6.1 Summary of key findings

This thesis documents the investigation of a cohort with the rare inherited disease foveal hypoplasia, for which the main focus of the research was the identification of novel disease gene(s) and mutations.

In Chapter 3, *SLC38A8* was screened by Sanger sequencing in a novel cohort of 13 patients with a minimum phenotype of foveal hypoplasia and a presumed autosomal recessive inheritance pattern. Only 2/13 (15%) patients had mutations in this gene, leaving a cohort of patients that were analysed using WES in Chapter 4 to investigate potential genetic heterogeneity. This led to the discovery of *PAX6* mutations in two patients, a potential duplication CNV in *SLC38A8* in a third, and the identification of new candidate foveal hypoplasia genes in the remaining cases. In Chapter 5, WGS was utilized to investigate SV and splicing mutations in *SLC38A8*, leading to the identification of a homozygous inversion in one patient.

For the solved *SLC38A8* and *PAX6* cases in this study, the fact that a molecular diagnosis has established an accurate clinical diagnosis. Clinically, this information can be used to inform the patient of the disease prognosis. It can also ensure that accurate genetic counselling is given to patients and family members (Wiggs and Pierce, 2013). This is especially important for *PAX6,* as AD mutations increase the

risk for future pregnancies. Knowing the molecular diagnosis also brings an opportunity to share rare mutations on platforms such as Matchmaker Exchange (https://www.matchmakerexchange.org), and consortia including the European Inherited Retinal Disease Consortium (ERDC) (https://www.erdc.info), the UKIRDC and Genomics England (GeL) (https://www.genomicsengland.co.uk). These platforms are also advantageous as a source of further disease cases when only a single case with a pathogenic variant has been found. In this study, although candidate genes were identified, there is a need to identify additional patients to prove the links to newly implicated genes are real. A success story of such platforms is the identification of *RAX2* variants in the ERDC consortium (Van de Sompele *et al.*, 2018). In the current study, WES data was checked for such genes arising from consortium discussions, but none were found.

6.2 Candidate gene identification

It was anticipated that mutations in the same gene in more than one patient would be found. As the cohort was relatively large for such a rare disease it was logical to hypothesise that this would affect the likelihood of finding more than one mutation in the same gene. However, the data presented in this study suggests that this is not the case. The data points to there being multiple rare novel genes. Research presented in Chapter 4 of this thesis suggests the power with which WES can efficiently identify putative pathogenic variants in candidate disease genes. A homozygous frameshift variant in *LAMP1* was prioritised as a top novel candidate gene in foveal hypoplasia as this variant was unreported in the publicly available databases and the gene is expressed in the early retina. A heterozygous nonsense

variant was identified in *VSX2* in a second patient. This variant was chosen because of the established role of *VSX2* in the developing retina, critical for neural retina progenitor cell proliferation and bipolar cell determination (Liu *et al.*, 1994; Rowan and Cepko, 2004). In a third patient, a heterozygous missense change in *EPHA2* was also prioritised due to the defined role of the ephrins and Eph receptors in proper axon guidance (as introduced in Chapter 1). A final candidate, *CHRNA4,* was prioritised because of the established role of nAChRs in signal transmission at synapses.

The *LAMP1* and *VSX2* variants are not predicted to lead to NMD due to their locations in the end region of the gene. It is predicted that such mutations have an effect because the mutant protein is produced. If the hypothesis is correct, then other mutations elsewhere might either have no effect or be lethal, and so mutations in cases with the same phenotype might cluster near those identified in this study. An example of this are mutations identified in *CTNNB1*, as variants altering the C-terminal domain of the protein cause non-syndromic FEVR (Panagiotou *et al.*, 2017). In the absence of additional families, functional work is required to add further evidence for their causation. As the *SLC38A8* mouse is not reported to have an eye phenotype and mice do not have a fovea, an alternative would be to study zebrafish or a cellular model. A clustered regularly interspaced short palindromic repeat (CRISPR)-cas9 knock-in model would be a clear experiment to specifically investigate if the mutations identified have an effect on transcription. This technology could be used to create a zebrafish or cellular model using homologous recombination or NHEJ. An alternative would be to use cytidine deaminase-based DNA base editors, as the *VSX2* variant p.(Gln298Ter) was created by a C>T

transition (Komor *et al.*, 2016; Shimatani *et al.*, 2017). Although lacking a fovea, a mouse model would be a suitable model for investigating the optic nerve misrouting seen in a number of cases in the cohort.

VSX2 mutations in a microphthalmia patient have been modelled in human induced pluripotent stem cells forming optic vesicle-like structures (hiPSC-OVs) to study retinal development at stages not normally accessible in humans (Phillips *et al.*, 2013; Phillips *et al.*, 2014). It is unfortunate that this functional work is not ideal for a subtle phenotype like foveal hypoplasia, as the optic cup derived from iPSC doesn't have a structured retina and they don't have a fovea. In the future, as this form of technology develops, iPSCs are believed to be the cellular model that all eye researchers will be using (Wu, Doorenbos and Dong Feng, 2016; Chen *et al.*, 2014a). As *VSX2* encodes a transcription factor, it is possible that the potentially truncated protein fails to interact with all of its transcriptional targets. This possibility could be investigated using transcriptomics or ChIPseq.

6.3 The phenotype/genotype spectrum of *SLC38A8* mutations

Two novel mutations were identified in two foveal hypoplasia patients by Sanger sequencing. Screening failed to identify a second variant in *SLC38A8* in one of these cases. CNV analysis of a third patient revealed a potential homozygous duplication. In a fourth patient, an inversion spanning three exons of *SLC38A8* was detected by WGS.

To look for possible phenotype-genotype correlations, the phenotypes of the patients with *SLC38A8* mutations identified in this study have been added to those previously published (Table 6-1). Reviewing this data, all affected patients reported in the literature (n=26), and those included in this study, have foveal hypoplasia, no loss of pigmentation to the hair, skin or eyes and no iris transillumination. Not all patients in the current study had undergone visually evoked potential analysis (n=6 tested) to look for chiasmal misrouting, and of the four patients with variants identified in *SLC38A8*, only F1377 had the defect. Although detection of chiasmal misrouting in the foveal hypoplasia cohort in this study is limited, it is difficult to diagnose in adult patients and in clinics without specialist equipment, meaning that the test is often not performed. Nevertheless, all patients with variants identified in *SLC38A8* in this study, and all patients with reported *SLC38A8* mutations in the literature have "normal" levels of pigmentation, added to no iris transillumination, indicative of a phenotype independent to albinism. In the current study, only one of the patients (F1368) with a variant identified in *SLC38A8* had ASD. ASD remains a variable component in individuals with recessive *SLC38A8* mutations, as reported in the original gene identification study (Poulter *et al.*, 2013). There is no obvious link between the type of mutation or location of the mutation identified in *SLC38A8* and phenotype. Mutations in *SLC38A8* have also been associated with developmental delay and pervasive developmental disorder-like features (Perez *et al.*, 2014), although such features were not reported in the foveal hypoplasia cohort tested this study, or in the original gene identification study. There is the potential for co-existence of phenotypes in patients, such as the Kartagener phenotype with foveal hypoplasia and chiasmal misrouting (van Genderen *et al.*, 2006). To date there have been no neuroimaging studies in patients with *SLC38A8* mutations, but it is unlikely

that abnormal cortical development (as reported by Perez et al) is related to *SLC38A8,* given the lack of findings in other cases. Overall visual acuities were significantly reduced in the foveal hypoplasia cohort, most likely caused by their underdeveloped fovea. Nystagmus was not reported as a feature in a number of cases but may have been present, further adding to reduced visual function. Vision levels in the patients with the novel *SLC38A8* mutations (F1377 and F1368) or SV (F1310 and predicted in F1335) identified in this study were similar to those reported in previous FHONDA patients (clinical correspondence).

Reviewing the consequences of all recessive *SLC38A8* mutations and SVs identified in the literature and this study to date (n=15 including compound het mutations); 9/15 (60%) are substitutions, 1/15 (6.7%) nonsense mutations, 1/15 (6.7%) in-frame deletions, 2/15 (13.3%) frameshift deletions and 2/15 (13.3%) SV.

The findings confirm that there is no genotype/phenotype correlation pattern that is obvious, as the number of samples was limited. It is also not possible to draw firm conclusions from this investigation, as the study was limited by the fact that patients were selectively included in this study if they had a minimum phenotype of foveal hypoplasia and no pigmentation defects.

Family	Ethnicity	Foveal hypoplasia	Optic nerve decussation defects	Nystagmus	Anterior segment abnormalities	Iris transillumination	Other notes on affected individuals
F1 (Poulter et al., 2013) 5 affected c.707T>A p.(Val236Asp)	Pakistani	++	++ (2/5)	++	Posterior embryotoxon in 3/5 and Axenfeld's anomaly present in 2/5 members.	- -	3/5 patients have astigmatism.
F2 (Poulter et al., 2013) 2 affected c.1002delG p.(Ser336Alafs*15)	Afghan	++	++	++	Posterior embryotoxon in both affected members.	- -	-
F3 (Poulter et al., 2013) 1 affected c.1234G>A p.(Gly412Arg)	Northern European	+	+	+	-	-	Kartagener syndrome also present
F4 (Poulter et al., 2013) 1 affected c.1029delG p.(Leu344Cysfs*7)	Pakistani	+	Inconclusive in single affected case	NR	-	-	-
F5 (Poulter et al., 2013) 1 affected c.101T>G p.(Met34Arg)	Turkish	+	+	NR	-	-	-
F6 (Poulter et al., 2013) 3 affected c.697G>A p.(Glu233Lys)	Indian	++	NT	++	-	- -	2/3 have bilateral microphthalmia and the other affected has a unilateral retinochoroidal coloboma (right eye)

F7 (Poulter et al., 2013) 2 affected c.598C>T p.(Gln200Ter) c.845_847delCTG p.(Ala282del)	Northern European	++	++	NR	Axenfeld's anomaly present only in 1/2 members	- -	-
Pedigree 1 (Perez et al., 2014) 3 affected c.95T>G p.(Ile32Ser)	Israeli (Jewish-Indian ancestry)	++	NT	++	NT	- -	All had astigmatism. Mild developmental delay with pervasive developmental disorder-like features were seen in all 3 patients.
Pedigree 2 (Perez et al., 2014) 2 affected c.95T>G p.(Ile32Ser)	Israeli (Jewish-Indian ancestry)	++	NT	++	NT	- -	All had astigmatism.
Pedigree 3 (Perez et al., 2014) 4 affected c.95T>G p.(Ile32Ser)	Israeli (Jewish-Indian ancestry)	++	NT	++	NT	- -	All had astigmatism. Diabetes mellitus and hypertension was seen in one affected patient
Pedigree 1 (Toral *et al.*, 2017) 2 affected c.848A>C p.(Asp283Ala)	Ashkenazi Jewish	++	NT	++	NT	- -	Both sibs had moderate hyperopia and astigmatism
F1368 c.848A>C p.(Asp283Ala)	Caucasian	+	NT	+	+	-	-
F1377 c.2T>C p.(Met1?)	Caucasian	+	+	+	NT	-	-

F1335 duplication *SLC38A8* exon 6	South Asian	+	NT	NR	-	-	-
F1310 inversion *SLC38A8*	Iraqi	+	NT	+	-	-	-

Table 6-1 Summary of clinical phenotypes of families with *SLC38A8* mutations and SVs.

Phenotypes of previously published foveal hypoplasia patients with mutations (Poulter *et al.*, 2013; Perez *et al.*, 2014) alongside phenotypes of patients with identified variants in *SLC38A8* from the current study. NR not reported; NT not tested; + present in single affected case; ++ present in all affected; - absent in single affected case; - - absent in all affected. The paper published by Perez et al., does not detail which affected individuals had full-field electroretinography and VEP analysis, but states that 6 patients appeared normal for these tests. They also state that colour vision, as assessed by the Ishihara colour vision test, was normal in 4 patients.

Regarding the “undetectable” second mutation in F1377, the analysis to date focused on the *SLC38A8* locus, with a second mutation thought to arise outside of the screened coding sequence. It is possible, however, that variants could be further afield than hypothesised in this study, as proper gene regulation depends not only on the necessary transcription factors, but also on the integrity, chromatin conformation and nuclear positioning of the gene (Kleinjan and Coutinho, 2009). Such position effect can be seen when the level of gene expression is impeded by chromosomal alterations or through disruption of control regions. For instance, enhancers for *SOX9* disrupted in Pierre Robin Sequence are located at 1.3Mb from the promoter (Benko *et al.*, 2009). More often long-range enhancers are seen in the introns of neighbouring genes, as seen with the *PAX6 cis*-regulatory elements in patients with aniridia (Bhatia *et al.*, 2014; Kleinjan *et al.*, 2001). Therefore, there is a need to examine genomic regions of *SLC38A8’s* neighbouring genes, *NECAB2* and *MBTPS1*. Mapping of topologically associated domains (TADs) using chromosome conformation capture methodologies (3C, 4C, 5C, Hi-C technologies) will help to further annotate the non-coding genome (Dekker *et al.*, 2002; Dostie *et al.*, 2006; van Berkum *et al.*, 2010; Zhao *et al.*, 2006; Matharu and Ahanger, 2015).

6.4 Unsolved cases

In 7/12 (58%) cases no clear pathogenic candidate gene was identified in the analysis of WES data. It is possible that WES-related technical issues hampered the identification of disease genes, for example difficulty aligning highly repetitive regions. Such technical issues became evident in the failure to identify the *FOXD1* variant due to poor probe coverage in Chapter 4.

This study focused on the entire *SLC38A8* locus when hunting for mutations not identifiable using the conventional methods. However, it is possible that causal mutations could also lie in the non-coding or regulatory regions of other known foveal hypoplasia genes, including *PAX6* and *FRMD7,* or in the albinism genes. Such genomic regions would not have been captured by WES. However, as WGS has now been performed in a number of these cases, there is the potential to investigate additional genes for SV or non-coding variants. There are now numerous tools for the annotation of non-coding regions of the genome, such as the *in silico* tool FATHMM applied in Chapter 5, although work is still ongoing to characterise these parts of the genome (Shihab *et al.*, 2013; Roadmap Epigenomics *et al.*, 2015; Ma *et al.*, 2018).

6.5 The future of genetic research

Differentiating disease-causing from non-pathogenic variants was a running theme of this study, particularly in the large volume of data generated by WGS and the intronic variants identified in case F1377 (with the undetectable second mutation in *SLC38A8)*. As researchers shift to studying genomes it is going to be more difficult to spot the pathogenic variant, as more variants will be listed. Large population datasets are useful, but it could be that these datasets include carriers of recessive alleles and possibly even foveal hypoplasia patients, as gnomAD contains *SLC38A8* homozygotes. Studies have reported the importance of gathering evidence for causation from several sources, as each has less power when used independently (MacArthur *et al.*, 2014). Therefore, in this study, parameters for investigating pathogenicity included reviewing a variant's frequency in the control population,

conservation of the amino acid residue, *in silico* pathogenicity prediction scores, segregation in the family where available and a genotype and phenotype consistent with those published previously. It became evident in this study that pathogenicity prediction tools give inconsistent results. This demonstrates that caution must be taken when reporting variant pathogenicity using such tools in the absence of functional studies. However, it is the time, cost, availability of reagents and relevant specimens (for example from the patient), that are important factors for consideration when planning such functional investigations.

6.6 Future directions

It is often questioned whether foveal hypoplasia with optic nerve misrouting is actually mild albinism, since in both albinos and FHONDA cases, foveal hypoplasia and chiasmal misrouting are strongly correlated (Dorey *et al.*, 2003). The degree of misrouting has been suggested to correlate with the loss of pigmentation (Dorey *et al.*, 2003; Von Dem Hagen *et al.*, 2007). However, the role of melanin in the development of the foveal pit has yet to be elucidated. There is no correlation between levels of pigmentation and reduced GCL density, which is often proposed to be a critical factor for normal foveal development (Leventhal *et al.*, 1989; Donatien, Aigner and Jeffery, 2002). The fact that optic chiasm formation takes place prior to the development of the fovea, suggests that the disease gene and encoded protein in patients with foveal hypoplasia and optic nerve decussation defects must be involved in both optic axon guidance and the lateral displacement of the inner retinal layers. This could lead to the centrifugal migratory defect of inner retinal neurons causing foveal hypoplasia and developmental abnormalities of the optic nerve head.

This is consistent with the expression pattern of *SLC38A8* observed in retinal tissues (Poulter *et al.*, 2013), but there is a need to observe the expression in the developing optic nerve.

This study has shown the importance of analysing whole-exome data with a hypothesis-free approach, as it is clear that researchers cannot be certain of the clinical diagnosis or form of inheritance reported from the clinicians, especially after the discoveries of the *PAX6* mutations in the cohort. The recruitment of patients with overlapping phenotypes has helped determine the phenotypic spectrum of *SLC38A8* mutations, but it has furthermore highlighted the need to screen *SLC38A8* in cases of isolated foveal hypoplasia (Curran and Robb, 1976; Giocanti-Auregan *et al.*, 2014; Karaca *et al.*, 2014; Mota *et al.*, 2012; Oliver *et al.*, 1987) and especially in patients with mutation-negative albinism. Mutation screening of known OA genes in a Danish cohort only accounted for 10/21 (48%) of the patients with AR OA, suggesting that *SLC38A8* and other candidate foveal hypoplasia genes require screening (Grønskov *et al.*, 2009). Further to screening *SLC38A8* in cases of albinism, CRISPR-cas9-generated *SLC38A8* iPSC mutants could be differentiated into RPE to see if they form pigmented RPE, like WT. Albinism variants could be added as a positive control for this assay.

Treatment of IRDs is an area of extensive study, with technologies including gene therapy for *RPE65* associated with LCA or RP, which is now an approved treatment in the US (Russell *et al.*, 2017). Gene augmentation is a sub-type of, and is the commonest form of gene therapy, such as the targeting of photoreceptors and the

RPE that has been successfully performed in a Phase 1/2 trial in choroideremia patients (MacLaren *et al.*, 2014; Edwards *et al.*, 2016). There are reports of studies investigating foveal hypoplasia therapies (Summers *et al.*, 2014). Studies of foveal morphology using OCT have shown that retinal development is not arrested in foveal hypoplasia, but is ongoing at a reduced rate and magnitude in comparison with unaffected individuals post-birth and even past the age of 5 (Gottlob, 2017). This indicates that there is postnatal plasticity with the potential for administration of therapy in early infancy or childhood to improve retinal development and vision (Gottlob, 2017).

In conclusion, even though further research is warranted, the work in this thesis helps towards a better understanding of foveal hypoplasia through the identification of novel candidate genes and the mutations in *SLC38A8* not detectable using the conventional methods. This study highlights the power of using NGS, particularly WGS in the identification of SV, but also the complications that arise when interpreting the data, for example the failure to map the breakpoints of the duplication CNV.

List of References

Abad-Navarro, F., de la Morena-Barrio, M. E., Fernández-Breis, J. T. and Corral, J. (2018) 'Lost in translation: bioinformatic analysis of variations affecting the translation initiation codon in the human genome', *Bioinformatics*, pp. 453-453.

Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., Handsaker, R. E., Kang, H. M., Marth, G. T. and McVean, G. A. (2012) 'An integrated map of genetic variation from 1,092 human genomes', *Nature,* 491(7422), pp. 56-65.

Abel, H. J. and Duncavage, E. (2013) 'Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches', *Cancer Genetics,* 206(12), pp. 432-440.

Acevedo, A. C., Poulter, J. A., Alves, P. G., de Lima, C. L., Castro, L. C., Yamaguti, P. M., Paula, L. M., Parry, D. A., Logan, C. V., Smith, C. E. L., Johnson, C. A., Inglehearn, C. F. and Mighell, A. J. (2015) 'Variability of systemic and oro-dental phenotype in two families with non-lethal Raine syndrome with FAM20C mutations', *BMC Medical Genetics,* 16, pp. 8-8.

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

Ahmed, S. S., Lott, M. N. and Marcus, D. M. (2005) 'The macular xanthophylls.', *Survey of Ophthalmology,* 50(2), pp. 183-193.

Al-Araimi, M., Pal, B., Poulter, J. A., van Genderen, M. M., Carr, I., Cudrnak, T., Brown, L., Sheridan, E., Mohamed, M. D., Bradbury, J., Ali, M., Inglehearn, C. F. and Toomes, C. (2013) 'A new recessively inherited disorder composed of foveal hypoplasia, optic nerve decussation defects and anterior segment dysgenesis maps to chromosome 16q23.3-24.1', *Molecular Vision,* 19, pp. 2165-2172.

Al-Saleh, A. A., Hellani, A. and Abu-Amero, K. K. (2011) 'Isolated foveal hypoplasia: report of a new case and detailed genetic investigation.', *International Ophthalmology,* 31(2), pp. 117-120.

Albert, S., Garanto, A., Sangermano, R., Khan, M., Bax, N. M., Hoyng, C. B., Zernant, J., Lee, W., Allikmets, R., Collin, R. W. J. and Cremers, F. P. M. (2018) 'Identification and Rescue of Splice Defects Caused by Two Neighboring Deep-Intronic ABCA4 Mutations Underlying Stargardt Disease', *The American Journal of Human Genetics,* 102(4), pp. 517-527.

Ali, M., McKibbin, M., Booth, A., Parry, D. A., Jain, P., Riazuddin, S. A., Hejtmancik, J. F., Khan, S. N., Firasat, S., Shires, M., Gilmour, D. F., Towns, K., Murphy, A.-L., Azmanov, D., Tournev, I., Cherninkova, S., Jafri, H., Raashid, Y., Toomes, C., Craig, J., Mackey, D. A., Kalaydjieva, L., Riazuddin, S. and Inglehearn, C. F. (2009) 'Null Mutations in LTBP2 Cause Primary Congenital Glaucoma', *The American Journal of Human Genetics,* 84(5), pp. 664-671.

Alkuraya, F. S. (2013) 'The application of next-generation sequencing in the autozygosity mapping of human recessive diseases', *Human Genetics,* 132(11), pp. 1197-1211.

Allikmets, R., Singh, N., Sun, H., Shroyer, N. F., Hutchinson, A., Chidambaram, A., Gerrard, B., Baird, L., Stauffer, D., Peiffer, A., Rattner, A., Smallwood, P., Li, Y., Anderson, K. L., Lewis, R. A., Nathans, J., Leppert, M., Dean, M. and Lupski, J. R. (1997) 'A photoreceptor cell-specific ATP-binding transporter

gene (ABCR) is mutated in recessive Starqardt macular dystrophy', *Nature Genetics,* 15, pp. 236.

Amini, R., Rocha-Martins, M. and Norden, C. (2018) 'Neuronal Migration and Lamination in the Vertebrate Retina', *Frontiers in neuroscience,* 11, pp. 742-742.

Anderson, D. M. G., Ablonczy, Z., Koutalos, Y., Hanneken, A. M., Spraggins, J. M., Calcutt, M. W., Crouch, R. K., Caprioli, R. M. and Schey, K. L. (2017) 'Bis(monoacylglycero)phosphate lipids in the retinal pigment epithelium implicate lysosomal/endosomal dysfunction in a model of Stargardt disease and human retinas', *Scientific Reports,* 7(1), pp. 17352-17352.

Anderson, D. R. (1969) 'Ultrastructure of human and monkey lamina cribrosa and optic nerve head', *Archives of Ophthalmology,* 82(6), pp. 800-814.

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. and Young, I. G. (1981) 'Sequence and organization of the human mitochondrial genome', *Nature,* 290(5806), pp. 457-465.

Andrejewski, N., Punnonen, E.-L., Guhde, G., Tanaka, Y., Lüllmann-Rauch, R., Hartmann, D., von Figura, K. and Saftig, P. (1999) 'Normal Lysosomal Morphology and Function in LAMP-1-deficient Mice', *Journal of Biological Chemistry,* 274(18), pp. 12692-12701.

Andrews, N. W. (1995) 'Lysosome recruitment during host cell invasion by Trypanosoma cruzi', *Trends in Cell Biology,* 5(3), pp. 133-137.

Andrews, N. W. (2000) 'Regulated secretion of conventional lysosomes', *Trends in Cell Biology,* 10(8), pp. 316-321.

Anikster, Y., Huizing, M., White, J., Shevchenko, Y. O., Fitzpatrick, D. L., Touchman, J. W., Compton, J. G., Bale, S. J., Swank, R. T., Gahl, W. A. and Toro, J. R. (2001) 'Mutation of a new gene causes a unique form of Hermansky–Pudlak syndrome in a genetic isolate of central Puerto Rico', *Nature Genetics,* 28, pp. 376.

Apkarian, P. (1996) 'Chiasmal crossing defects in disorders of binocular vision', *Eye,* 10(2), pp. 222-232.

Appelmans, F., Wattiaux, R. and De Duve, C. (1955) 'Tissue fractionation studies. 5. The association of acid phosphatase with a special class of cytoplasmic granules in rat liver', *The Biochemical Journal,* 59(3), pp. 438-445.

Arno, G., Holder, G. E., Chakarova, C. and et al. (2016) 'Recessive retinopathy consequent on mutant g-protein β subunit 3 (gnb3)', *JAMA Ophthalmology,* 134(8), pp. 924-927.

Astuti, G. D. N., van den Born, L. I., Khan, M. I., Hamel, C. P., Bocquet, B., Manes, G., Quinodoz, M., Ali, M., Toomes, C., McKibbin, M., El-Asrag, M. E., Haer-Wigman, L., Inglehearn, C. F., Black, G. C. M., Hoyng, C. B., Cremers, F. P. M. and Roosing, S. (2018) 'Identification of Inherited Retinal Disease-Associated Genetic Variants in 11 Candidate Genes', *Genes,* 9(1), pp. 21.

Audo, I., Bujakowska, K. M., Léveillard, T., Mohand-Saïd, S., Lancelot, M.-E., Germain, A., Antonio, A., Michiels, C., Saraiva, J.-P., Letexier, M., Sahel, J.-A., Bhattacharya, S. S. and Zeitz, C. (2012) 'Development and application of a next-generation-sequencing (NGS) approach to detect known and novel gene defects underlying retinal diseases', *Orphanet Journal of Rare Diseases,* 7, pp. 8-8.

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D. and Thibault, J. (2013) 'From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline', *Current Protocols in Bioinformatics*, pp. 11-10.

Azuma, N., Nishina, S., Yanagisawa, H., Okuyama, T. and Yamada, M. (1996) 'PAX6 missense mutation in isolated foveal hypoplasia', *Nature Genetics,* 13(2), pp. 141-142.

Azuma, N., Tadokoro, K., Asaka, A., Yamada, M., Yamaguchi, Y., Handa, H., Matsushima, S., Watanabe, T., Kohsaka, S., Kida, Y., Shiraishi, T., Ogura, T., Shimamura, K. and Nakafuku, M. (2005) 'The Pax6 isoform bearing an alternative spliced exon promotes the development of the neural retinal structure', *Human Molecular Genetics,* 14(6), pp. 735-745.

Azuma, N., Yamaguchi, Y., Handa, H., Hayakawa, M., Kanai, A. and Yamada, M. (1999) 'Missense Mutation in the Alternative Splice Region of the PAX6 Gene in Eye Anomalies', *The American Journal of Human Genetics,* 65(3), pp. 656-663.

Azuma, N., Yamaguchi, Y., Handa, H., Tadokoro, K., Asaka, A., Kawase, E. and Yamada, M. (2003) 'Mutations of the PAX6 Gene Detected in Patients with a Variety of Optic-Nerve Malformations', *The American Journal of Human Genetics,* 72(6), pp. 1565-1570.

Azzopardi, P. and Cowey, A. (1993) 'Preferential representation of the fovea in the primary visual cortex', *Nature,* 361, pp. 719.

Baere, E. D., Lefever, S., Meire, F., Schil, K. V., Devlieghere, E., Leroy, B. P. and Coppieters, F. (2011) 'Molecular Diagnosis Of Autosomal Recessive Retinal Dystrophies By Homozygosity Mapping With SNP Arrays', *Investigative Ophthalmology & Visual Science,* 52(14), pp. 5377-5377.

Bailey, J. A., Gu, Z., Clark, R. A., Reinert, K., Samonte, R. V., Schwartz, S., Adams, M. D., Myers, E. W., Li, P. W. and Eichler, E. E. (2002) 'Recent segmental duplications in the human genome', *Science,* 297.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. and McCammon, J. A. (2001) 'Electrostatics of nanosystems: Application to microtubules and the ribosome', *Proceedings of the National Academy of Sciences of the United States of America,* 98(18), pp. 10037-10041.

Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Edmond, M. J., Nickerson, D. A. and Shendure, J. (2011) 'Exome sequencing as a tool for Mendelian disease gene discovery.', *Nature Reviews Genetics,* 12, pp. 745-755.

Bar-Yosef, U., Abuelaish, I., Harel, T., Hendler, N., Ofir, R. and Birk, O. S. (2004) 'CHX10 mutations cause non-syndromic microphthalmia/anophthalmia in Arab and Jewish kindreds.', *Human Genetics,* 115(4), pp. 302-309.

Barber, A. N. (1955) *Embryology of the human eye.* Henry Kimpton.

Barbieri, A. M., Lupo, G., Bulfone, A., Andreazzoli, M., Mariani, M., Fougerousse, F., Consalez, G. G., Borsani, G., Beckmann, J. S., Barsacchi, G., Ballabio, A. and Banfi, S. (1999) 'A homeobox gene, vax2, controls the patterning of the eye dorsoventral axis', *Proceedings of the National Academy of Sciences,* 96(19), pp. 10729.

Barbosa, M. D. F. S., Barrat, F. J., Tchernev, V. T., Nguyen, Q. A., Mishra, V. S., Colman, S. D., Pastural, E., Dufourcq-Lagelouse, R., Fischer, A., Holcombe, R. F., Wallace, M. R., Brandt, S. J., De Saint Basile, G. and Kingsmore, S. F. (1997) 'Identification of mutations in two major mRNA isoforms of the Chediak-Higashi syndrome gene in human and mouse', *Human Molecular Genetics,* 6.

Bartenhagen, C. and Dugas, M. (2016) 'Robust and exact structural variation detection with paired-end and soft-clipped alignments: SoftSV compared with eight algorithms', *Briefings in Bioinformatics,* 17(1), pp. 51-62.

Bashirullah, A., Cooperstock, R. L. and Lipshitz, H. D. (2001) 'Spatial and temporal control of RNA stability', *Proceedings of the National Academy of Sciences of the United States of America,* 98(13), pp. 7025-7028.

Bassi, M. T., Schiaffino, M. V., Renieri, A., Nigris, F. D., Galli, L., Bruttini, M., Gebbia, M., Bergen, A. A. B., Lewis, R. A. and Ballabio, A. (1995) 'Cloning of the gene for ocular albinism type 1 from the distal short arm of the X chromosome', *Nature Genetics,* 10, pp. 13.

Bauwens, M., De Zaeytijd, J., Weisschuh, N., Kohl, S., Meire, F., Dahan, K., Depasse, F., De Jaegere, S., De Ravel, T., De Rademaeker, M., Loeys, B., Coppieters, F., Leroy, B. P. and De Baere, E. (2015) 'An Augmented ABCA4 Screen Targeting Noncoding Regions Reveals a Deep Intronic Founder Variant in Belgian Stargardt Patients', *Human Mutation,* 36(1), pp. 39--42.

Baylor, D. A. (1987) 'Photoreceptor signals and vision. Proctor lecture', *Investigative Ophthalmology & Visual Science,* 28(1), pp. 34-49.

Belkadi, A., Bolze, A., Itan, Y., Cobat, A., Vincent, Q., 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 of the United States of America,* 112(17), pp. 5473-5478.

Benedek, G. B. (1971) 'Theory of transparency of the eye.', *Applied Optics,* 10(3), pp. 459-473.

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C. T., McBride, D., Golzio, C., Fisher, M., Perry, P., Abadie, V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M. M., Picard, A., Temple, I. K., Thomas, P., Vazquez, M.-P., Vekemans, M., Crollius, H. R., Hastie, N. D., Munnich, A., Etchevers, H. C., Pelet, A., Farlie, P. G., FitzPatrick, D. R. and Lyonnet, S. (2009) 'Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence', *Nature Genetics,* 41(3), pp. 359-364.

Bennett, T. M., M'Hamdi, O., Hejtmancik, J. F. and Shiels, A. (2017) 'Germ-line and somatic EPHA2 coding variants in lens aging and cataract', *PLoS One,* 12(12), pp. e0189881-e0189881.

Bharti, K., Gasper, M., Ou, J., Brucato, M., Clore-Gronenborn, K., Pickel, J. and Arnheiter, H. (2012) 'A Regulatory Loop Involving PAX6, MITF, and WNT Signaling Controls Retinal Pigment Epithelium Development', *PLoS Genetics,* 8(7), pp. e1002757.

Bhatia, S., Bengani, H., Fish, M., Brown, A., Divizia, Maria T., de Marco, R., Damante, G., Grainger, R., van Heyningen, V. and Kleinjan, Dirk A. (2013) 'Disruption of Autoregulatory Feedback by a Mutation in a Remote, Ultraconserved PAX6 Enhancer Causes Aniridia', *The American Journal of Human Genetics,* 93(6), pp. 1126-1134.

Bhatia, S., Monahan, J., Ravi, V., Gautier, P., Murdoch, E., Brenner, S., van Heyningen, V., Venkatesh, B. and Kleinjan, D. A. (2014) 'A survey of ancient conserved non-coding elements in the PAX6 locus reveals a landscape of interdigitated cis-regulatory archipelagos', *Developmental Biology,* 387(2), pp. 214-228.

Bhattacharya, S., Yin, J., Winborn, C. S., Zhang, Q., Yue, J. and Chaum, E. (2017) 'Prominin-1 Is a Novel Regulator of Autophagy in the Human Retinal Pigment Epithelium', *Investigative Ophthalmology & Visual Science,* 58(4), pp. 2366-2387.

Birney, E. and Stamatoyannopoulos, J. A. and Dutta, A. and Guigó, R. and Gingeras, T. R. and Margulies, E. H. and Weng, Z. and Snyder, M. and Dermitzakis, E. T. and Thurman, R. E. and Kuehn, M. S. and Taylor, C. M. and Neph, S. and Koch, C. M. and Asthana, S. and Malhotra, A. and Adzhubei, I. and Greenbaum, J.

A. and Andrews, R. M. and Flicek, P. and Boyle, P. J. and Cao, H. and Carter, N. P. and Clelland, G. K. and Davis, S. and Day, N. and Dhami, P. and Dillon, S. C. and Dorschner, M. O. and Fiegler, H. and Giresi, P. G. and Goldy, J. and Hawrylycz, M. and Haydock, A. and Humbert, R. and James, K. D. and Johnson, B. E. and Johnson, E. M. and Frum, T. T. and Rosenzweig, E. R. and Karnani, N. and Lee, K. and Lefebvre, G. C. and Navas, P. A. and Neri, F. and Parker, S. C. and Sabo, P. J. and Sandstrom, R. and Shafer, A. and Vetrie, D. and Weaver, M. and Wilcox, S. and Yu, M. and Collins, F. S. and Dekker, J. and Lieb, J. D. and Tullius, T. D. and Crawford, G. E. and Sunyaev, S. and Noble, W. S. and Dunham, I. and Denoeud, F. and Reymond, A. and Kapranov, P. and Rozowsky, J. and Zheng, D. and Castelo, R. and Frankish, A. and Harrow, J. and Ghosh, S. and Sandelin, A. and Hofacker, I. L. and Baertsch, R. and Keefe, D. and Dike, S. and Cheng, J. and Hirsch, H. A. and Sekinger, E. A. and Lagarde, J. and Abril, J. F. and Shahab, A. and Flamm, C. and Fried, C. and Hackermüller, J. and Hertel, J. and Lindemeyer, M. and Missal, K. and Tanzer, A. and Washietl, S. and Korbel, J. and Emanuelsson, O. and Pedersen, J. S. and Holroyd, N. and Taylor, R. and Swarbreck, D. and Matthews, N. and Dickson, M. C. and Thomas, D. J. and Weirauch, M. T. and Gilbert, J. and Drenkow, J. and Bell, I. and Zhao, X. and Srinivasan, K. G. and Sung, W. K. and Ooi, H. S. and Chiu, K. P. and Foissac, S. and Alioto, T. and Brent, M. and Pachter, L. and Tress, M. L. and Valencia, A. and Choo, S. W. and Choo, C. Y. and Ucla, C. and Manzano, C. and Wyss, C. and Cheung, E. and Clark, T. G. and Brown, J. B. and Ganesh, M. and Patel, S. and Tammana, H. and Chrast, J. and Henrichsen, C. N. and Kai, C. and Kawai, J. and Nagalakshmi, U. and Wu, J. and Lian, Z. and Lian, J. and Newburger, P. and Zhang, X. and Bickel, P. and Mattick, J. S. and Carninci, P. and Hayashizaki, Y. and Weissman, S. and Hubbard, T. and Myers, R. M. and Rogers, J. and Stadler, P. F. and Lowe, T. M. and Wei, C. L. and Ruan, Y. and Struhl, K. and Gerstein, M. and Antonarakis, S. E. and Fu, Y. and Green, E. D. and Karaöz, U. and Siepel, A. and Taylor, J. and Liefer, L. A. and Wetterstrand, K. A. and Good, P. J. and Feingold, E. A. and Guyer, M. S. and Cooper, G. M. and Asimenos, G. and Dewey, C. N. and Hou, M. and Nikolaev, S. and Montoya-Burgos, J. I. and Löytynoja, A. and Whelan, S. and Pardi, F. and Massingham, T. and Huang, H. and Zhang, N. R. and Holmes, I. and Mullikin, J. C. and Ureta-Vidal, A. and Paten, B. and Seringhaus, M. and Church, D. and Rosenbloom, K. and Kent, W. J. and Stone, E. A. and Batzoglou, S. and Goldman, N. and Hardison, R. C. and Haussler, D. and Miller, W. and Sidow, A. and Trinklein, N. D. and Zhang, Z. D. and Barrera, L. and Stuart, R. and King, D. C. and Ameur, A. and Enroth, S. and Bieda, M. C. and Kim, J. and Bhinge, A. A. and Jiang, N. and Liu, J. and Yao, F. and Vega, V. B. and Lee, C. W. and Ng, P. and Shahab, A. and Yang, A. and Moqtaderi, Z. and Zhu, Z. and Xu, X. and Squazzo, S. and Oberley, M. J. and Inman, D. and Singer, M. A. and Richmond, T. A. and Munn, K. J. and Rada-Iglesias, A. and Wallerman, O. and Komorowski, J. and Fowler, J. C. and Couttet, P. and Bruce, A. W. and Dovey, O. M. and Ellis, P. D. and Langford, C. F. and Nix, D. A. and Euskirchen, G. and Hartman, S. and Urban, A. E. and Kraus, P. and Van Calcar, S. and Heintzman, N. and Kim, T. H. and Wang, K. and Qu, C. and Hon, G. and Luna, R. and Glass, C. K. and Rosenfeld, M. G. and Aldred, S. F. and Cooper, S. J. and Halees, A. and Lin, J. M. and Shulha, H. P. and Zhang, X. and Xu, M. and Haidar, J. N. and Yu, Y. and Ruan, Y. and Iyer, V. R. and Green, R. D. and Wadelius, C. and Farnham, P. J. and Ren, B. and Harte, R. A. and Hinrichs, A. S. and Trumbower, H. and Clawson, H. and Hillman-Jackson, J. and Zweig, A. S. and Smith, K. and Thakkapallayil, A. and Barber, G. and Kuhn, R. M. and Karolchik, D. and Armengol, L. and Bird, C. P. and de Bakker, P. I. and Kern, A.

D. and Lopez-Bigas, N. and Martin, J. D. and Stranger, B. E. and Woodroffe, A. and Davydov, E. and Dimas, A. and Eyras, E. and Hallgrímsdóttir, I. B. and Huppert, J. and Zody, M. C. and Abecasis, G. R. and Estivill, X. and Bouffard, G. G. and Guan, X. and Hansen, N. F. and Idol, J. R. and Maduro, V. V. and Maskeri, B. and McDowell, J. C. and Park, M. and Thomas, P. J. and Young, A. C. and Blakesley, R. W. and Muzny, D. M. and Sodergren, E. and Wheeler, D. A. and K.C, W. and Jiang, H. and Weinstock, G. M. and Gibbs, R. A. and Graves, T. and Fulton, R. and Mardis, E. R. and Wilson, R. K. and Clamp, M. and Cuff, J. and Gnerre, S. and Jaffe, D. B. and Chang, J. L. and Lindblad-Toh, K. and Lander, E. S. and Koriabine, M. and Nefedov, M. and Osoegawa, K. and Yoshinaga, Y. and Zhu, B. and de Jong, P. J. (2012) 'An Integrated Encyclopedia of DNA Elements in the Human Genome', *Nature,* 489(7414), pp. 57-74.

Biswas, S. and Lloyd, I. C. (1999) 'Oculocutaneous albinism', *Archives of Disease in Childhood,* 80(6), pp. 565-569.

Blake, J., Riddell, A., Theiss, S., Gonzalez, A. P., Haase, B., Jauch, A., Janssen, J. W. G., Ibberson, D., Pavlinic, D., Moog, U., Benes, V. and Runz, H. (2014) 'Sequencing of a Patient with Balanced Chromosome Abnormalities and Neurodevelopmental Disease Identifies Disruption of Multiple High Risk Loci by Structural Variation', *PLoS One,* 9(3), pp. e90894.

Bok, D. (1993) 'The retinal pigment epithelium: a versatile partner in vision', *Journal of Cell Science,* 1993(Supplement 17), pp. 189-195.

Bonaventure, J., Domingues, M. J. and Large, L. (2013) 'Cellular and molecular mechanisms controlling the migration of melanocytes and melanoma cells.', *Pigment Cell & Melanoma Research,* 26(3), pp. 316-325.

Bondeson, M. L., Dahl, N., Malmgren, H., Kleijer, W. J., Tonnesen, T., Carlberg, B. M. and Pettersson, U. (1995) 'Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome', *Human Molecular Genetics,* 4.

Bossi, G. and Griffiths, G. M. (2005) 'CTL secretory lysosomes: biogenesis and secretion of a harmful organelle', *Seminars in Immunology,* 17(1), pp. 87-94.

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

Bourguignon, C., Li, J. and Papalopulu, N. (1998) 'XBF-1, a winged helix transcription factor with dual activity, has a role in positioning neurogenesis in Xenopus competent ectoderm', *Development,* 125(24), pp. 4889.

Boynton, R. M. and Whitten, D. N. (1970) 'Visual Adaptation in Monkey Cones: Recordings of Late Receptor Potentials', *Science,* 170(3965), pp. 1423.
Brantley-Sieders, D. M., Caughron, J., Hicks, D., Pozzi, A., Ruiz, J. C. and Chen, J. (2004) 'EphA2 receptor tyrosine kinase regulates endothelial cell migration and vascular assembly through phosphoinositide 3-kinase-mediated Rac1 GTPase activation', *Journal of Cell Science,* 117(10), pp. 2037.

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.

Bujakowska, K. M., Fernandez-Godino, R., Place, E., Consugar, M., Navarro-Gomez, D., White, J., Bedoukian, E. C., Zhu, X., Xie, H. M., Gai, X., Leroy, B. P. and

Pierce, E. A. (2016) 'Copy-number variation is an important contributor to the genetic causality of inherited retinal degenerations', *Genetics In Medicine,* 19, pp. 643.

Bäumer, N., Marquardt, T., Stoykova, A., Ashery-Padan, R., Chowdhury, K. and Gruss, P. (2002) 'Pax6 is required for establishing naso-temporal and dorsal characteristics of the optic vesicle', *Development,* 129(19), pp. 4535.

Calvo, S. E., Pagliarini, D. J. and Mootha, V. K. (2009) 'Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans', *Proceedings of the National Academy of Sciences,* 106(18), pp. 7507-7512.

Carden, S. M., Boissy, R. E., Schoettker, P. J. and Good, W. V. (1998) 'Albinism: modern molecular diagnosis', *British journal of ophthalmology,* 82(2), pp. 189-195.

Cardoso, A. R., Oliveira, M., Amorim, A. and Azevedo, L. (2016) 'Major influence of repetitive elements on disease-associated copy number variants (CNVs)', *Human Genomics,* 10, pp. 30.

Carr, I. M., Bhaskar, S., O'Sullivan, J., Aldahmesh, M. A., Shamseldin, H. E., Markham, A. F., Bonthron, D. T., Black, G. and Alkuraya, F. S. (2013) 'Autozygosity mapping with exome sequence data', *Human mutation,* 34(1), pp. 50-56.

Carss, K. J., Arno, G., Erwood, M., Stephens, J., Sanchis-Juan, A., Hull, S., Megy, K., Grozeva, D., Dewhurst, E., Malka, S., Plagnol, V., Penkett, C., Stirrups, K., Rizzo, R., Wright, G., Josifova, D., Bitner-Glindzicz, M., Scott, R. H., Clement, E., Allen, L., Armstrong, R., Brady, A. F., Carmichael, J., Chitre, M., Henderson, R. H. H., Hurst, J., MacLaren, R. E., Murphy, E., Paterson, J., Rosser, E., Thompson, D. A., Wakeling, E., Ouwehand, W. H., Michaelides, M., Moore, A. T., Webster, A. R. and Raymond, F. L. (2017) 'Comprehensive rare variant analysis via Whole-Genome sequencing to determine the molecular pathology of inherited retinal disease.', *American Journal of Human Genetics,* 100(1), pp. 75-90.

Carvalho, C. M. B. and Lupski, J. R. (2016) 'Mechanisms underlying structural variant formation in genomic disorders', *Nature Reviews Genetics,* 17(4), pp. 224-238.

Carvalho, C. M. B., Zhang, F., Liu, P., Patel, A., Sahoo, T., Bacino, C. A., Shaw, C., Peacock, S., Pursley, A., Tavyev, Y. J., Ramocki, M. B., Nawara, M., Obersztyn, E., Vianna-Morgante, A. M., Stankiewicz, P., Zoghbi, H. Y., Cheung, S. W. and Lupski, J. R. (2009) 'Complex rearrangements in patients with duplications of MECP2 can occur by fork stalling and template switching', *Human Molecular Genetics,* 18(12), pp. 2188-2203.

Cavener, D. R. and Ray, S. C. (1991) 'Eukaryotic start and stop translation sites', *Nucleic Acids Research,* 19(12), pp. 3185-3192.

Chakravarthy, U., Biundo, E., Saka, R. O., Fasser, C., Bourne, R. and Little, J.-A. (2017) 'The Economic Impact of Blindness in Europe', *Ophthalmic Epidemiology,* 24(4), pp. 239-247.

Chan-Ling, T. (1994) 'Glial, neuronal and vascular interactions in the mammalian retina', *Progress in Retinal and Eye Research,* 13(1), pp. 357-389.

Chen, F. K., McLenachan, S., Edel, M., Da Cruz, L., Coffey, P. J. and Mackey, D. A. (2014a) 'iPS Cells for Modelling and Treatment of Retinal Diseases', *Journal of Clinical Medicine,* 3(4), pp. 1511-1541.

Chen, J.-M., Cooper, D. N., Férec, C., Kehrer-Sawatzki, H. and Patrinos, G. P. (2010) 'Genomic rearrangements in inherited disease and cancer', *Seminars in Cancer Biology,* 20(4), pp. 222-233.

Chen, J. W., Murphy, T. L., Willingham, M. C., Pastan, I. and August, J. T. (1985) 'Identification of two lysosomal membrane glycoproteins', *The Journal of Cell Biology,* 101(1), pp. 85-95.

Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L. and Mardis, E. R. (2009) 'BreakDancer: an algorithm for high-resolution mapping of genomic structural variation', *Nature Methods,* 6(9), pp. 677-681.

Chen, L., Zhou, W., Zhang, L. and Zhang, F. (2014b) 'Genome Architecture and Its Roles in Human Copy Number Variation', *Genomics & Informatics,* 12(4), pp. 136-144.

Choi, J.-H., Shin, J.-H., Seo, J. H., Jung, J.-H. and Choi, K.-D. (2015) 'A start codon mutation of the FRMD7 gene in two Korean families with idiopathic infantile nystagmus', *Scientific Reports,* 5, pp. 13003.

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), pp. e46688.

Chow, R. L. and Lang, R. A. (2001) 'Early Eye Development in Vertebrates', *Annual Review of Cell and Developmental Biology,* 17(1), pp. 255-296.

Chrástek, R., Wolf, M., Donath, K., Niemann, H., Paulus, D., Hothorn, T., Lausen, B., Lämmer, R., Mardin, C. Y. and Michelson, G. (2005) 'Automated segmentation of the optic nerve head for diagnosis of glaucoma', *Medical Image Analysis,* 9(4), pp. 297-314.

Chung, K. Y., Leung, K. M., Lin, C. C., Tam, K. C., Has, Y. L., Taylor, J. S. and Chan, S. O. (2004) 'Regionally specific expression of L1 and sialylated NCAM in the retinofugal pathway of mouse embryos.', *The Journal of Comparative Neurology,* 471(4), pp. 482-498.

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, pp. 7512.

Clark, K. L., Halay, E. D., Lai, E. and Burley, S. K. (1993) 'Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5', *Nature,* 364, pp. 412.

Clark, M. J., Chen, R., Lam, H. Y. K., Karczewski, K. J., Chen, R., Euskirchen, G., Butte, A. J. and Snyder, M. (2011) 'Performance comparison of exome DNA sequencing technologies', *Nature biotechnology,* 29(10), pp. 908-914.

Cogan, D. G. (1967) 'Congenital nystagmus.', *Canadian Journal of Ophthalmology,* 2(1), pp. 4-10.

Collin, R. W. J., van den Born, L. I., Klevering, B. J., de Castro-Miró, M., Littink, K. W., Arimadyo, K., Azam, M., Yazar, V., Zonneveld, M. N., Paun, C. C., Siemiatkowska, A. M., Strom, T. M., Hehir-Kwa, J. Y., Kroes, H. Y., de Faber, J.-T. H. N., van Schooneveld, M. J., Heckenlively, J. R., Hoyng, C. B., den Hollander, A. I. and Cremers, F. P. M. (2011) 'High-Resolution Homozygosity Mapping Is a Powerful Tool to Detect Novel Mutations Causative of Autosomal Recessive RP in the Dutch Population', *Investigative Ophthalmology & Visual Science,* 52(5), pp. 2227-2239.

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., Wellcome Trust Case Control, C., 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.

Cooper, G. M. and Shendure, J. (2011) 'Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data.', *Nature Reviews Genetics,* 12(9), pp. 628-640.

Coppieters, F., Baert, A., Todeschini, A. L., Van Cauwenbergh, C., Bauwens, M., De Jaegere, S., de Ravel de l'Argentière, T., Meire, F., Abdelmoula Bouayed, N. and Florentin-Arar, L. (2013) 'Regulatory Mutations in the 5'UTR of NMNAT1, encoding the Nuclear Isoform of Nicotinamide Nucleotide Adenylyltransferase 1, cause Leber Congenital Amaurosis', *European Journal of Human Genetics,* 21(suppl. 2), pp. 128-128.

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.

Costin, G.-E., Valencia, J. C., Vieira, W. D., Lamoreux, M. L. and Hearing, V. J. (2003) 'Tyrosinase processing and intracellular trafficking is disrupted in mouse primary melanocytes carrying the underwhite (uw) mutation. A model for oculocutaneous albinism (OCA) type 4', *Journal of Cell Science,* 116(15), pp. 3203.

Cowey, A. (1964) 'Projection of the retina on to striate and prestriate cortex in the squirrel monkey, Saimiri sciureus. ', *Journal of Neurophysiology, 2*7(3), pp. 366-393.

Creel, D. J., Summers, C. G. and King, R. A. (1990) 'Visual anomalies associated with albinism', *Ophthalmic paediatrics and genetics,* 11(3), pp. 193-200.

Cremers, F. P. M., Boon, C. J. F., Bujakowska, K. and Zeitz, C. (2018) 'Special Issue Introduction: Inherited Retinal Disease: Novel Candidate Genes, Genotype–Phenotype Correlations, and Inheritance Models', *Genes,* 9(4), pp. 215.

Cullinane, A. R., Vilboux, T., O'Brien, K., Curry, J. A., Maynard, D. M., Carlson-Donohoe, H., Ciccone, C., Program, N. C. S., Markello, T. C., Gunay-Aygun, M., Huizing, M. and Gahl, W. A. (2011) 'Homozygosity Mapping and Whole Exome Sequencing to Detect SLC45A2 and G6PC3 Mutations in a Single Patient with Oculocutaneous Albinism and Neutropenia', *The Journal of Investigative Dermatology,* 131(10), pp. 2017-2025.

Curcio, C. A. and Sloan, K. R. (1992) 'Packing geometry of human cone photoreceptors: variations with eccentricity and evidence for local anisotropy.', *Visual Neuroscience,* 9(2), pp. 169-180.

Curran, R. E. and Robb, R. M. (1976) 'Isolated foveal hypoplasia', *Archives of Ophthalmology,* 94(1), pp. 48-50.

Czerny, T., Schaffner, G. and Busslinger, M. (1993) 'DNA sequence recognition by Pax proteins: bipartite structure of the paired domain and its binding site', *Genes & Development,* 7(10), pp. 2048-2061.

Dai, J., Buhusi, M., Demyanenko, G. P., Brennaman, L. H., Hruska, M., Dalva, M. B. and Maness, P. F. (2013) 'Neuron Glia-Related Cell Adhesion Molecule (NrCAM) Promotes Topographic Retinocollicular Mapping', *PLoS One,* 8(9), pp. e73000.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R.

and Genomes Project Analysis, G. (2011) 'The variant call format and VCFtools', *Bioinformatics,* 27(15), pp. 2156-2158.

Daniel, P. M. and Whitteridge, D. (1961) 'The representation of the visual field on the cerebral cortex in monkeys', *The Journal of Physiology,* 159(2), pp. 203-221.

Dave, A., Laurie, K., Staffieri, S. E., Taranath, D., Mackey, D. A., Mitchell, P., Wang, J. J., Craig, J. E., Burdon, K. P. and Sharma, S. (2013) 'Mutations in the EPHA2 Gene Are a Major Contributor to Inherited Cataracts in South-Eastern Australia.', *PLoS One,* 8(8), pp. e72518.

Dawson, D. W., Volpert, O. V., Gillis, P., Crawford, S. E., Xu, H. J., Benedict, W. and Bouck, N. P. (1999) 'Pigment Epithelium-Derived Factor: A Potent Inhibitor of Angiogenesis', *Science,* 285(5425), pp. 245.

de Ligt, J., Boone, P. M., Pfundt, R., Vissers, L. E., Richmond, T., Geoghegan, J., O'Moore, K., de Leeuw, N., Shaw, C., Brunner, H. G., Lupski, J. R., Veltman, J. A. and Hehir-Kwa, J. Y. (2013) 'Detection of clinically relevant copy number variants with whole-exome sequencing.', *Human Mutation,* 34(10), pp. 1439-1448.

de Melo Reis, R. A., Ventura, A. L. M., Schitine, C. S., de Mello, M. C. F. and de Mello, F. G. (2008) 'Müller Glia as an Active Compartment Modulating Nervous Activity in the Vertebrate Retina: Neurotransmitters and Trophic Factors', *Neurochemical Research,* 33(8), pp. 1466-1474.

Deiner, M. S., Kennedy, T. E., Fazeli, A., Serafini, T., Tessier-Lavigne, M. and Sretavan, D. W. (1997) 'Netrin-1 and DCC Mediate Axon Guidance Locally at the Optic Disc: Loss of Function Leads to Optic Nerve Hypoplasia', *Neuron,* 19(3), pp. 575-589.

Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002) 'Capturing Chromosome Conformation', *Science,* 295(5558), pp. 1306.

del Marmol, V. and Beermann, F. (1996) 'Tyrosinase and related proteins in mammalian pigmentation', *FEBS Letters,* 381(3), pp. 165-168.

Delamere, N. A. (2005) 'Ciliary Body and Ciliary Epithelium', *Advances in Organ Biology,* 10, pp. 127-148.

Dell'Angelica, E. C., Shotelersuk, V., Aguilar, R. C., Gahl, W. A. and Bonifacino, J. S. (1999) 'Altered Trafficking of Lysosomal Proteins in Hermansky-Pudlak Syndrome Due to Mutations in the beta-3A Subunit of the AP-3 Adaptor.', *Molecular Cell,* 3(1), pp. 11-21.

Demyanenko, G. P. and Maness, P. F. (2003) 'The L1 Cell Adhesion Molecule Is Essential for Topographic Mapping of Retinal Axons', *The Journal of Neuroscience,* 23(2), pp. 530.

den Hollander, A. I., Koenekoop, R. K., Mohamed, M. D., Arts, H. H., Boldt, K., Towns, K. V., Sedmak, T., Beer, M., Nagel-Wolfrum, K., McKibbin, M., Dharmaraj, S., Lopez, I., Ivings, L., Williams, G. A., Springell, K., Woods, C. G., Jafri, H., Rashid, Y., Strom, T. M., van der Zwaag, B., Gosens, I., Kersten, F. F. J., van Wijk, E., Veltman, J. A., Zonneveld, M. N., van Beersum, S. E. C., Maumenee, I. H., Wolfrum, U., Cheetham, M. E., Ueffing, M., Cremers, F. P. M., Inglehearn, C. F. and Roepman, R. (2007a) 'Mutations in LCA5, encoding the ciliary protein lebercilin, cause Leber congenital amaurosis', *Nature Genetics,* 39(7), pp. 889-895.

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.

den Hollander, A. I., Lopez, I., Yzer, S., Zonneveld, M. N., Janssen, I. M., Strom, T. M., Hehir-Kwa, J. Y., Veltman, J. A., Arends, M. L., Meitinger, T., Musarella, M. A., van den Born, L. I., Fishman, G. A., Maumenee, I. H., Rohrschneider, K., Cremers, F. P. M. and Koenekoop, R. K. (2007b) 'Identification of Novel Mutations in Patients with Leber Congenital Amaurosis and Juvenile RP by Genome-wide Homozygosity Mapping with SNP Microarrays', *Investigative Ophthalmology & Visual Science,* 48(12), pp. 5690-5698.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A. and Hanna, M. (2011) 'A framework for variation discovery and genotyping using next-generation DNA sequencing data', *Nature Genetics,* 43(5), pp. 491-498.

Diaz de Arce, A. J., Noderer, W. L. and Wang, C. L. (2018) 'Complete motif analysis of sequence requirements for translation initiation at non-AUG start codons', *Nucleic Acids Research,* 46(2), pp. 985-994.

Dickson, B. J. (2002) 'Molecular Mechanisms of Axon Guidance', *Science,* 298(5600), pp. 1959.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., Sorek, R. and Rechavi, G. (2012) 'Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq', *Nature,* 485, pp. 201.

Donatien, P., Aigner, B. and Jeffery, G. (2002) 'Variations in cell density in the ganglion cell layer of the retina as a function of ocular pigmentation', *European Journal of Neuroscience,* 15(10), pp. 1597-1602.

Dorey, S. E., Neveu, M. M., Burton, L. C., Sloper, J. J. and Holder, G. E. (2003) 'The clinical features of albinism and their correlation with visual evoked potentials', *The British Journal of Ophthalmology,* 87(6), pp. 767-772.

Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. D., Krumm, A., Lamb, J., Nusbaum, C., Green, R. D. and Dekker, J. (2006) 'Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements', *Genome research,* 16(10), pp. 1299-1309.

Dyer, M. A. and Cepko, C. L. (2001) 'Regulating proliferation during retinal development', *Nature Reviews Neuroscience,* 2, pp. 333.

Edelmann, L., Pandita, R. K. and Morrow, B. E. (1999) 'Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome', *American Journal of Human Genetics,* 64(4), pp. 1076-1086.

Edelmann, L., Pandita, R. K., Spiteri, E., Funke, B., Goldberg, R., Palanisamy, N., Chaganti, R. S., Magenis, E., Shprintzen, R. J. and Morrow, B. E. (1999) 'A common molecular basis for rearrangement disorders on chromosome 22q11.', *Human Molecular Genetics,* 7, pp. 1157-1167.

Edwards, T. L., Jolly, J. K., Groppe, M., Barnard, A. R., Cottriall, C. L., Tolmachova, T., Black, G. C., Webster, A. R., Lotery, A. J., Holder, G. E., Xue, K., Downes, S. M., Simunovic, M. P., Seabra, M. C. and MacLaren, R. E. (2016) 'Visual Acuity after Retinal Gene Therapy for Choroideremia', *The New England Journal of Medicine,* 374(20), pp. 1996-1998.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., deWinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park,

I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J. and Turner, S. (2009) 'Real-Time DNA Sequencing from Single Polymerase Molecules', *Science,* 323(5910), pp. 133.

Eiken, H. G., Knappskog, P. M., Apold, J., Skjelkvåle, L. and Boman, H. (1992) 'A de novo phenylketonuria mutation: ATG (met) to ATA (ile) in the start codon of the phenylalanine hydroxylase gene', *Human Mutation,* 1(5), pp. 388-391.

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), pp. e78496.

El-Asrag, Mohammed E., Sergouniotis, Panagiotis I., McKibbin, M., Plagnol, V., Sheridan, E., Waseem, N., Abdelhamed, Z., McKeefry, D., Van Schil, K., Poulter, James A., Johnson, C. A., Carr, I. M., Leroy, B. P., De Baere, E., Inglehearn, C. F., Webster, A. R., Toomes, C. and Ali, M. (2015) 'Biallelic Mutations in the Autophagy Regulator DRAM2 Cause Retinal Dystrophy with Early Macular Involvement', *The 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', *The European Journal of Human Genetics*.

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. (2018) '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.

Engerman, R. L. (1976) 'Development of the macular circulation.', *Investigative Ophthalmology,* 15(10), pp. 835-840.

Epstein, J., Cai, J., Glaser, T., Jepeal, L. and Maas, R. (1994a) 'Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes', *Journal of Biological Chemistry,* 269(11), pp. 8355-8361.

Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S. and Maas, R. L. (1994b) 'Two independent and interactive DNA-binding subdomains of the Pax6 paired domain are regulated by alternative splicing', *Genes & Development,* 8(17), pp. 2022-2034.

Erskine, L. and Herrera, E. (2007) 'The retinal ganglion cell axon's journey: Insights into molecular mechanisms of axon guidance', *Developmental Biology,* 308(1), pp. 1-14.

Eskelinen, E.-L. (2006) 'Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy', *Molecular Aspects of Medicine,* 27(5), pp. 495-502.

Eskelinen, E.-L., Schmidt, C. K., Neu, S., Willenborg, M., Fuertes, G., Salvador, N., Tanaka, Y., Lüllmann-Rauch, R., Hartmann, D., Heeren, J., von Figura, K., Knecht, E. and Saftig, P. (2004) 'Disturbed cholesterol traffic but normal proteolytic function in LAMP-1/LAMP-2 double-deficient fibroblasts', *Molecular Biology of the Cell,* 15(7), pp. 3132-3145.

Euler, T., Haverkamp, S., Schubert, T. and Baden, T. (2014) 'Retinal bipolar cells: elementary building blocks of vision', *Nature Reviews Neuroscience,* 15, pp. 507.

Evans, T. A. and Bashaw, G. J. (2010) 'Axon Guidance at the Midline: Of Mice and Flies', *Current Opinion in Neurobiology,* 20(1), pp. 79-85.

Faiyaz-Ul-Haque, M., Zaidi, S. H. E., Al-Mureikhi, M. S., Peltekova, I., Tsui, L. C. and Teebi, A. S. (2007) 'Mutations in the CHX10 gene in non-syndromic microphthalmia/anophthalmia patients from Qatar', *Clinical Genetics,* 72(2), pp. 164-166.

Farrell, R. A., McCally, R. L. and Tatham, P. E. (1973) 'Wave-length dependencies of light scattering in normal and cold swollen rabbit corneas and their structural implications.', *The Journal of Physiology,* 233(3), pp. 589-612.

Feng, Y., Niu, T., Xing, H., Xu, X., Chen, C., Peng, S., Wang, L., Laird, N. and Xu, X. (2004) 'A Common Haplotype of the Nicotine Acetylcholine Receptor α4 Subunit Gene Is Associated with Vulnerability to Nicotine Addiction in Men', *The American Journal of Human Genetics,* 75(1), pp. 112-121.

Ferda Percin, E., Ploder, L. A., Yu, J. J., Arici, K., Horsford, J. D., Rutherford, A., Bapat, B., Cox, D. W., Duncan, A. M. V., Kalnins, V. I., Kocak-Altintas, A., Sowden, J. C., Traboulsi, E., Sarfarazi, M. and McInnes, R. R. (2000) 'Human microphthalmia associated with mutations in the retinal homeobox gene CHX10', *Nature Genetics,* 25, pp. 397.

Ferrington, D. A., Sinha, D. and Kaarniranta, K. (2016) 'Defects in retinal pigment epithelial cell proteolysis and the pathology associated with age-related macular degeneration', *Progress in retinal and eye research,* 51, pp. 69-89.

Feuk, L., Carson, A. R. and Scherer, S. W. (2006) 'Structural variation in the human genome', *Nature Reviews Genetics,* 7.

Fiorentino, A., Fujinami, K., Arno, G., Robson, A. G., Pontikos, N., Arasanz Armengol, M., Plagnol, V., Hayashi, T., Iwata, T., Parker, M., Fowler, T., Rendon, A., Gardner, J. C., Henderson, R. H., Cheetham, M. E., Webster, A. R., Michaelides, M. and Hardcastle, A. J. (2017) 'Missense variants in the X-linked gene PRPS1 cause retinal degeneration in females', *Human Mutation,* 39(1), pp. 80-91.

Fiorentino, A., Yu, J., Arno, G., Pontikos, N., Halford, S., Broadgate, S., Michaelides, M., Carss, K. J., Raymond, F. L., Cheetham, M. E., Webster, A. R., Downes, S. M., Hardcastle, A. J., Consortium, N. I.-B. R. D. and Consortium, U. K. I. R. D. (2018) 'Novel homozygous splicing mutations in ARL2BP cause autosomal recessive retinitis pigmentosa', *Molecular Vision,* 24, pp. 603-612.

Flanagan, J. G. and Van Vactor, D. (1998) 'Through the Looking Glass: Axon Guidance at the Midline Choice Point', *Cell,* 92(4), pp. 429-432.

Forsius, H. and Eriksson, A. W. (1964) 'Ein neues Augensyndrom mit X-chromosomaler Transmission: eine Sippe mit Fundusalbinismus, Foveahypoplasie,

Nystagmus, Myopie, Astigmatismus und Dyschromatopsie.', *Klinische Monatsblätter für Augenheilkunde,* 144, pp. 447-457.

Francis, P. J. (2006) 'Genetics of inherited retinal disease', *Journal of the Royal Society of Medicine,* 99(4), pp. 189-191.

Frank, H. R., Landers, M. B. r., Williams, R. J. and Sidbury, J. B. (1974) 'A new dominant progressive foveal dystrophy.', *The American Journal of Ophthalmology,* 78(6), pp. 903-916.

Frebourg, T. (2014) 'The Challenge for the Next Generation of Medical Geneticists', *Human Mutation,* 35(8), pp. 909-911.

Fromer, M., Moran, Jennifer L., Chambert, K., Banks, E., Bergen, Sarah E., Ruderfer, Douglas M., Handsaker, Robert E., McCarroll, Steven A., O'Donovan, Michael C., Owen, Michael J., Kirov, G., Sullivan, Patrick F., Hultman, Christina M., Sklar, P. and Purcell, Shaun M. (2012) 'Discovery and Statistical Genotyping of Copy-Number Variation from Whole-Exome Sequencing Depth', *The American Journal of Human Genetics,* 91(4), pp. 597-607.

Fruttiger, M. (2007) 'Development of the retinal vasculature.', *Angiogenesis,* 10(2), pp. 77-88.

Fuhrmann, S. (2010) 'Eye Morphogenesis and Patterning of the Optic Vesicle', *Current Topics in Developmental Biology,* 93, pp. 61-84.

Gabrieli, T., Sharim, H., Michaeli, Y. and Ebenstein, Y. (2017) 'Cas9-Assisted Targeting of CHromosome segments (CATCH) for targeted nanopore sequencing and optical genome mapping', *bioRxiv*.

Gao, Z., Waggoner, D., Stephens, M., Ober, C. and Przeworski , C. (2015) 'An estimate of the average number of recessive lethal mutations carried by humans.', *Genetics,* 199(4), pp. 1243-1254.

Garanto, A., Duijkers, L. and Collin, R. W. J. (2015) 'Species-Dependent Splice Recognition of a Cryptic Exon Resulting from a Recurrent Intronic CEP290 Mutation that Causes Congenital Blindness', *International Journal of Molecular Sciences,* 16(3), pp. 5285-5298.

Gargiulo, A., Testa, F., Rossi, S., Di Iorio, V., Fecarotta, S., de Berardinis, T., Iovine, A., Magli, A., Signorini, S., Fazzi, E., Galantuomo, M. S., Fossarello, M., Montefusco, S., Ciccodicola, A., Neri, A., Macaluso, C., Simonelli, F. and Surace, E. M. (2011) 'Molecular and Clinical Characterization of Albinism in a Large Cohort of Italian Patients', *Investigative Ophthalmology & Visual Science,* 52(3), pp. 1281-1289.

Gariano, R. F., Iruela-Arispe, M. L. and Hendrickson, A. E. (1994) 'Vascular development in primate retina: comparison of laminar plexus formation in monkey and human', *Investigative Ophthalmology & Visual Science,* 35(9), pp. 3442-3455.

Gelfman, S., Burstein, D., Penn, O., Savchenko, A., Amit, M., Schwartz, S., Pupko, T. and Ast, G. (2012) 'Changes in exon–intron structure during vertebrate evolution affect the splicing pattern of exons', *Genome Research,* 22(1), pp. 35-50.

George, A., Zand, D. J., Hufnagel, R. B., Sharma, R., Sergeev, Y. V., Legare, J. M., Rice, G. M., Scott Schwoerer, J. A., Rius, M., Tetri, L., Gamm, D. M., Bharti, K. and Brooks, B. P. (2016) 'Biallelic Mutations in MITF Cause Coloboma, Osteopetrosis, Microphthalmia, Macrocephaly, Albinism, and Deafness', *The American Journal of Human Genetics,* 99(6), pp. 1388-1394.

Georges, P., Madigan, M. C. and Provis, J. M. (1999) 'Apoptosis during development of the human retina: relationship to foveal development and retinal synaptogenesis.', *The Journal of Comparative Neurology,* 413(2), pp. 198-208.

Gerhardt, H., Golding, M., Fruttiger, M., Ruhrberg, C., Lundkvist, A., Abramsson, A., Jeltsch, M., Mitchell, C., Alitalo, K. and Shima, D. (2003) 'VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia', *The Journal of Cell Biology,* 161(6), pp. 1163-1177.

Ghiasvand, N. M., Rudolph, D. D., Mashayekhi, M., Brzezinski Iv, J. A., Goldman, D. and Glaser, T. (2011) 'Deletion of a remote enhancer near ATOH7 disrupts retinal neurogenesis, causing NCRNA disease', *Nature Neuroscience,* 14(5), pp. 578-586.

Gilissen, C., Hehir-Kwa, J. Y., Thung, D. T., van de Vorst, M., van Bon, B. W. M., Willemsen, M. H., Kwint, M., Janssen, I. M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H. G., de Vries, B. B. A., Kleefstra, T., Brunner, H. G., Vissers, L. E. L. M. and Veltman, J. A. (2014) 'Genome sequencing identifies major causes of severe intellectual disability', *Nature,* 511, pp. 344.

Giocanti-Auregan, A., Witmer, M. T., Radcliffe, N. M. and D'Amico, D. J. (2014) 'Isolated foveal hypoplasia without nystagmus.', *International Ophthalmology,* 34, pp. 877.

Glaser, T., Jepeal, L., Edwards, J. G., Young, R. S., Favor, J. and Maas, R. L. (1994) 'PAX6 gene dosage effect in a family with congenital cataracts, aniridia, anophthalmia and central nervous system defects.', *Nature Genetics,* 7(4), pp. 463-471.

Glaser, T., Walton, D. S. and Maas, R. L. (1992) 'Genomic structure, evolutionary conservation and aniridia mutations in the human PAX6 gene', *Nature Genetics,* 2(3), pp. 232-239.

Gottlob, I. 'Abnormal foveal development.'. *ARVO*, Baltimore, MD, USA.

Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001) 'Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure', *Journal of Molecular Biology,* 313 (4), pp. 903-919.

Green, J. S., Parfey, P. S., Harnett, J. D., Farid, N. R., Cramer, B. C., Johnson, G., Heath, O., McManamon, P. J., O'Leary, E. and Pryse-Phillips, W. (1989) 'The Cardinal Manifestations of Bardet-Biedl Syndrome, a form of Laurence-Moon-Biedl Syndrome.', *The New England Journal of Medicine,* 321, pp. 1002-1009.

Griffin, C., Kleinjan, D. A., Doe, B. and van Heyningen, V. (2002) 'New 3′ elements control Pax6 expression in the developing pretectum, neural retina and olfactory region', *Mechanisms of Development,* 112(1), pp. 89-100.

Grillo, G., Turi, A., Licciulli, F., Mignone, F., Liuni, S., Banfi, S., Gennarino, V. A., Horner, D. S., Pavesi, G., Picardi, E. and Pesole, G. (2010) 'UTRdb and UTRsite (RELEASE 2010): a collection of sequences and regulatory motifs of the untranslated regions of eukaryotic mRNAs', *Nucleic Acids Research,* 38(Database issue), pp. D75-D80.

Griscelli, C., Durandy, A., Guy-Grand, D., Daguillard, F., Herzog, C. and Prunieras, M. (1978) 'A syndrome associating partial albinism and immunodeficiency', *The American Journal of Medicine,* 65(4), pp. 691-702.

Grønskov, K., Dooley, C. M., Østergaard, E., Kelsh, R. N., Hansen, L., Levesque, M. P., Vilhelmsen, K., Møllgård, K., Stemple, D. L. and Rosenberg, T. (2013) 'Mutations in c10orf11, a melanocyte-differentiation gene, cause autosomal-recessive albinism', *The American Journal of Human Genetics,* 92(3), pp. 415-421.

Grønskov, K., Ek, J. and Brondum-Nielsen, K. (2007) 'Oculocutaneous albinism', *Orphanet Journal of Rare Diseases,* 2, pp. 43-43.

Grønskov, K., Ek, J., Sand, A., Scheller, R., Bygum, A., Brixen, K., Brondum-Nielsen, K. and Rosenberg, T. (2009) 'Birth Prevalence and Mutation Spectrum in Danish Patients with Autosomal Recessive Albinism', *Investigative Ophthalmology & Visual Science,* 50(3), pp. 1058-1064.

Grønskov, K., Rosenberg, T., Sand, A. and Brøndum-Nielsen, K. (1999) 'Mutational analysis of PAX6: 16 novel mutations including 5 missense mutations with a mild aniridia phenotype', *The European Journal of Human Genetics,* 7(3), pp. 274-286.

Gu, W., Zhang, F. and Lupski, J. R. (2008) 'Mechanisms for human genomic rearrangements', *PathoGenetics,* 1, pp. 4-4.

Gu, X., Neric, N. J., Crabb, J. S., Crabb, J. W., Bhattacharya, S. K., Rayborn, M. E., Hollyfield, J. G. and Bonilha, V. L. (2012) 'Age-related changes in the retinal pigment epithelium (RPE)', *PLoS One,* 7(6), pp. e38673.

Guo, T., Tan, Z.-P., Chen, H.-M., Zheng, D.-y., liu, L., Huang, X.-G., Chen, P., Luo, H. and Yang, Y.-F. (2017) 'An effective combination of whole-exome sequencing and runs of homozygosity for the diagnosis of primary ciliary dyskinesia in consanguineous families', *Scientific Reports,* 7, pp. 7905.

Guo, Y., Sheng, Q., Samuels, D. C., Lehmann, B., Bauer, J. A., Pietenpol, J. and Shyr, Y. (2013) 'Exome copy number variation estimation tools using array comparative genomic hybridisation as control.', *BioMed Research International,* 2013.

Halder, G., Callaerts, P. and Gehring, W. J. (1995) 'Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila', *Science,* 267(5205), pp. 1788.

Halford, S., Lindsay, E., Nayudu, M., Carey, A. H., Baldini, A. and Scambler, P. J. (1993) 'Low-copy-number repeat sequences flank the DiGeorge/velo-cardio-facial syndrome loci at 22q11.', *Human Molecular Genetics,* 2, pp. 191-196.

Halfter, W., Dong, S., Schurer, B., Ring, C., Cole, G. J. and Eller, A. (2005) 'Embryonic Synthesis of the Inner Limiting Membrane and Vitreous Body', *Investigative Ophthalmology & Visual Science,* 46(6), pp. 2202-2209.

Hamel, C. (2006) 'Retinitis pigmentosa', *Orphanet Journal of Rare Diseases,* 1, pp. 40-40.

Hansen, M. J., Dallal, G. E. and Flanagan, J. G. (2004) 'Retinal Axon Response to Ephrin-As Shows a Graded, Concentration-Dependent Transition from Growth Promotion to Inhibition', *Neuron,* 42(5), pp. 717-730.

Hanson, I., Churchill, A., Love, J., Axton, R., Moore, T., Clarke, M., Meire, F. and van Heyningen, V. (1999) 'Missense Mutations in the Most Ancient Residues of the PAX6 Paired Domain Underlie a Spectrum of Human Congenital Eye Malformations', *Human Molecular Genetics,* 8(2), pp. 165-172.

Hardcastle, Z. and Papalopulu, N. (2000) 'Distinct effects of XBF-1 in regulating the cell cycle inhibitor p27(XIC1) and imparting a neural fate', *Development,* 127(6), pp. 1303.

Harman, A. M. and Jeffery, G. (1992) 'Distinctive pattern of organisation in the retinofugal pathway of a marsupial: I. Retina and optic nerve', *Journal of Comparative Neurology,* 325(1), pp. 47-56.

Harnett, J. D., Green, J. S., Cramer, B. C., Johnson, G., Chafe, L., McManamon, P., Farid, N. R., Pryse-Phillips, W. and Parfey, P. S. (1988) 'The Spectrum of Renal Disease in Laurence-Mood-Biedl Syndrome.', *The New England Journal of Medicine,* 319, pp. 615-618.

Harper, A. R. and Summers, J. A. (2015) 'The Dynamic Sclera: Extracellular Matrix Remodeling in Normal Ocular Growth and Myopia Development', *Experimental Eye Research,* 133, pp. 100-111.

Hassell, J. R. and Birk, D. E. (2010) 'The Molecular Basis of Corneal Transparency', *Experimental eye research,* 91(3), pp. 326-335.

Hastings, P. J., Lupski, J. R., Rosenberg, S. M. and Ira, G. (2009) 'Mechanisms of change in gene copy number', *Nature reviews. Genetics,* 10(8), pp. 551-564.

Hatini, V., Tao, W. and Lai, E. (1994) 'Expression of winged helix genes, BF-1 and BF-2, define adjacent domains within the developing forebrain and retina.', *Journal of Neurobiology,* 10, pp. 1293-1309.

Hayashi, M. and Suzuki, T. 2005. Oculocutaneous Albinism Type 4. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018.

Hayden, E. C. (2014) 'Technology: The $1000 genome.', *Nature,* 507, pp. 294-295.

Heavner, W. and Pevny, L. (2012) 'Eye development and retinogenesis', *Cold Spring Harbor Perspectives in Biology,* 4(12), pp. 008391.

Hendrickson, A. (2016) 'Development of Retinal Layers in Prenatal Human Retina', *The American Journal of Ophthalmology,* 161, pp. 29-35.e1.

Hendrickson, A., Possin, D., Vajzovic, L. and Toth, C. A. (2012) 'Histologic development of the human fovea from midgestation to maturity', *The American Journal of Ophthalmology,* 154(5), pp. 767-778.e2.

Hendrickson, A., Troll, D., Possin, D. and Springer, A. (2006) 'Development of the neural retina and its vasculature in the marmoset Callithrix jacchus', *The Journal of Comparative Neurology,* 497(2), pp. 270-286.

Hendrickson, A. E. and Yuodelis, C. (1984) 'The morphological development of the human fovea', *Ophthalmology,* 91(6), pp. 603-612.

Henikoff, S. and Henikoff, J. G. (1993) 'Performance evaluation of amino acid substitution matrices', *Proteins: Structure, Function, and Bioinformatics,* 17(1), pp. 49-61.

Hernandez, M. R., Igoe, F. and Neufeld, A. H. (1986) 'Extracellular Matrix of the Human Optic Nerve Head', *The American Journal of Ophthalmology,* 102(2), pp. 139-148.

Hernandez, M. R., Luo, X. X., Igoe, F. and Neufeld, A. H. (1987) 'Extracellular Matrix of the Human Lamina Cribrosa', *The American Journal of Ophthalmology,* 104(6), pp. 567-576.

Herrera, E., Marcus, R., Li, S., Williams, S. E., Erskine, L., Lai, E. and Mason, C. (2004) 'Foxd1 is required for proper formation of the optic chiasm', *Development,* 131(22), pp. 5727.

Hever, A., KA, W. and V., v. H. (2006) 'Developmental malformations of the eye: the role of PAX6, SOX2 and OTX2.', *Clinical Genetics,* 69(6), pp. 459-470.

Highnam, G., Wang, J. J., Kusler, D., Zook, J., Vijayan, V., Leibovich, N. and Mittelman, D. (2015) 'An analytical framework for optimizing variant discovery from personal genomes', *Nature Communications,* 6, pp. 6275.

Hildebrand, G. and R. Fielder, A. (2011) *Anatomy and Physiology of the Retina.*

Hill, R. E., Favor, J., Hogan, B. L. M., Ton, C. C. T., Saunders, G. F., Hanson, I. M., Prosser, J., Jordan, T., Hastie, N. D. and Heyningen, V. v. (1991) 'Mouse Small

eye results from mutations in a paired-like homeobox-containing gene', *Nature,* 354(6354), pp. 522-525.

Hingorani, M., Williamson, K. A., Moore, A. T. and van Heyningen, V. (2009) 'Detailed Ophthalmologic Evaluation of 43 Individuals with PAX6 Mutations', *Investigative Ophthalmology & Visual Science,* 50(6), pp. 2581-2590.

Hopker, V. H., Shewan, D., Tessier-Lavigne, M., Poo, M.-m. and Holt, C. (1999) 'Growth-cone attraction to netrin-1 is converted to repulsion by laminin-1', *Nature,* 401(6748), pp. 69-73.

Horton, J. C. (1997) 'Wilbrand's knee of the primate optic chiasm is an artefact of monocular enucleation', *Transactions of the American Ophthalmological Society,* 95, pp. 579-609.

Horton, P. and Nakai, K. (1997) 'Better prediction of protein cellular localization sites with the k nearest neighbors classifier.', *Proceedings. International Conference on Intelligent Systems for Molecular Biology,* 5, pp. 147-152.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W. and Zimmermann, P. (2008) 'Genevestigator V3: A Reference Expression Database for the Meta-Analysis of Transcriptomes', *Advances in Bioinformatics,* 2008, pp. 420747.

Huang, L., Kuo, Y.-M. and Gitschier, J. (1999) 'The pallid gene encodes a novel, syntaxin 13-interacting protein involved in platelet storage pool deficiency', *Nature Genetics,* 23, pp. 329.

Huang, Q., Wang, S., Sorenson, C. M. and Sheibani, N. (2008) 'PEDF-Deficient Mice Exhibit an Enhanced Rate of Retinal Vascular Expansion and are More Sensitive to Hyperoxia-Mediated Vessel Obliteration', *Experimental Eye Research,* 87(3), pp. 226-241.

Hughes, M. J. G. and Andrews, D. W. (1997) 'A single nucleotide is a sufficient 5′ untranslated region for translation in an eukaryotic in vitro system.', *FEBS Letters,* 414(1), pp. 19-22.

Hughes, S., Yang, H. and Chan-Ling, T. (2000) 'Vascularization of the human fetal retina: roles of vasculogenesis and angiogenesis', *Investigative Ophthalmology & Visual Science,* 41(5), pp. 1217-1228.

Huh, S., Hatini, V., Marcus, R. C., Li, S. C. and Lai, E. (1999) 'Dorsal–Ventral Patterning Defects in the Eye of BF-1-Deficient Mice Associated with a Restricted Loss of shh Expression', *Developmental Biology,* 211(1), pp. 53-63.

Hull, S., Arno, G., Plagnol, V., Chamney, S., Russell-Eggitt, I., Thompson, D., Ramsden, S. C., Black, G. C. M., Robson, A., Holder, G. E., Moore, A. T. and Webster, A. R. (2014) 'The Phenotypic Variability of Retinal Dystrophies Associated With Mutations in CRX, With Report of a Novel Macular Dystrophy Phenotype', *Investigative Ophthalmology & Visual Science,* 55(10), pp. 6934-6944.

Hunter, S. G., Zhuang, G., Brantley-Sieders, D., Swat, W., Cowan, C. W. and Chen, J. (2006) 'Essential Role of Vav Family Guanine Nucleotide Exchange Factors in EphA Receptor-Mediated Angiogenesis', *Molecular and Cellular Biology,* 26(13), pp. 4830-4842.

Hutton, S. M. and Spritz, R. A. (2008a) 'A Comprehensive Genetic Study of Autosomal Recessive Ocular Albinism in Caucasian Patients', *Investigative Ophthalmology & Visual Science,* 49(3), pp. 868-872.

Hutton, S. M. and Spritz, R. A. (2008b) 'Comprehensive Analysis of Oculocutaneous Albinism among Non-Hispanic Caucasians Shows that OCA1 Is the Most Prevalent OCA Type', *The Journal of Investigative Dermatology,* 128(10), pp. 2442-2450.

Huynh, K. K., Eskelinen, E.-L., Scott, C. C., Malevanets, A., Saftig, P. and Grinstein, S. (2007) 'LAMP proteins are required for fusion of lysosomes with phagosomes', *The EMBO Journal,* 26(2), pp. 313-324.

Hägglund, M. G. A., Hellsten, S. V., Bagchi, S., Philippot, G., Löfqvist, E., Nilsson, V. C. O., Almkvist, I., Karlsson, E., Sreedharan, S., Tafreshiha, A. and Fredriksson, R. (2015) 'Transport of l-Glutamine, l-Alanine, l-Arginine and l-Histidine by the Neuron-Specific Slc38a8 (SNAT8) in CNS', *Journal of Molecular Biology,* 427(6, Part B), pp. 1495-1512.

Ibrahim, M. T., Alarcon-Martinez, T., Lopez, I., Fajardo, N., Chiang, J. and Koenekoop, R. K. (2018) 'A complete, homozygous CRX deletion causing nullizygosity is a new genetic mechanism for Leber congenital amaurosis', *Scientific Reports,* 8, pp. 5034.

Idrees, F., Vaindeanu, D., Fraser, S. G., Sowden, J. C. and Khaw, P. T. (2006) 'A review of anterior segment dysgeneses.', *Survey of Ophthalmology,* 51(3), pp. 213-231.

Inagaki, H., Ohye, T., Kogo, H., Tsutsumi, M., Kato, T., Tong, M., Emanuel, B. S. and Kurahashi, H. (2013) 'Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations', *Nature Communications,* 4, pp. 1592.

Ionides, A. C., Berry, V., Mackay, D. S., Moore, A. T., Bhattacharya, S. S. and Shiels, A. (1997) 'A locus for autosomal dominant posterior polar cataract on chromosome 1p.', *Human Molecular Genetics,* 6(1), pp. 47-51.

Ionita-Laza, I., McCallum, K., Xu, B. and Buxbaum, J. D. (2016) 'A spectral approach integrating functional genomic annotations for coding and noncoding variants', *Nature Genetics,* 48, pp. 214-220.

Iseri, S. U., Wyatt, A. W., Nurnberg, G., Kluck, C., Nurnburg, P., Holder, G. E., Blair, E., Salt, A. and Ragge, N. K. (2010) 'Use of genome-wide SNP homozygosity mapping in small pedigrees to identify new mutations in VSX2 causing recessive microphthalmia and a semidominant inner retinal dystrophy.', *Human Genetics,* 128(1), pp. 51-60.

Izquierdo, N. J., Townsend, W. and Hussels, I. E. (1995) 'Ocular findings in the Hermansky-Pudlak syndrome', *Transactions of the American Ophthalmological Society,* 93, pp. 191-202.

Jackson, B. C., Carpenter, C., Nebert, D. W. and Vasiliou, V. (2010) 'Update of human and mouse forkhead box (FOX) gene families', *Human Genomics,* 4(5), pp. 345-352.

Jain, M., Olsen, H. E., Paten, B. and Akeson, M. (2016) 'The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community', *Genome Biology,* 17, pp. 239.

Jansen, R.-P. (2001) 'mRNA localization: message on the move', *Nature Reviews Molecular Cell Biology,* 2, pp. 247.

Jaworski, A., Tom, I., Tong, R. K., Gildea, H. K., Koch, A. W., Gonzalez, L. C. and Tessier-Lavigne, M. (2015) 'Operational redundancy in axon guidance through the multifunctional receptor Robo3 and its ligand NELL2', *Science,* 350(6263), pp. 961.

Jeffery, G. and Erskine, L. (2005) 'Variations in the architecture and development of the vertebrate optic chiasm', *Progress in Retinal and Eye Research,* 24(6), pp. 721-753.

Jeffery, G. and Harman, A. M. (1992) 'Distinctive pattern of organisation in the retinofugal pathway of a marsupial: II. Optic chiasm', *Journal of Comparative Neurology,* 325(1), pp. 57-67.

Jordan, T., Hanson, I., Zaletayev, D., Hodgson, S., Prosser, J., Seawright, A., Hastie, N. and van Heyningen, V. (1992) 'The human PAX6 gene is mutated in two patients with aniridia.', *Nature Genetics,* 1, pp. 328-332.

Joubert, M., Eisenring, J. J., Robb, J. P. and Andermann, F. (1969) 'Familial agenesis of the cerebellar vermis. A syndrome of episodic hyperpnea, abnormal eye movements, ataxia, and retardation.', *Neurology,* 19(9), pp. 813-825.

Jun, G., Guo, H., Klein, B. E. K., Klein, R., Wang, J. J., Mitchell, P., Miao, H., Lee, K. E., Joshi, T., Buck, M., Chugha, P., Bardenstein, D., Klein, A. P., Bailey-Wilson, J. E., Gong, X., Spector, T. D., Andrew, T., Hammond, C. J., Elston, R. C., Iyengar, S. K. and Wang, B. (2009) 'EPHA2 Is Associated with Age-Related Cortical Cataract in Mice and Humans', *PLoS Genetics,* 5(7), pp. e1000584.

Kadalayil, L., Rafiq, S., Rose-Zerilli, M. J. J., Pengelly, R. J., Parker, H., Oscier, D., Strefford, J. C., Tapper, W. J., Gibson, J., Ennis, S. and Collins, A. (2015) 'Exome sequence read depth methods for identifying copy number changes', *Briefings in Bioinformatics,* 16(3), pp. 380-392.

Kamaraj, B. and Purohit, R. (2014) 'Mutational Analysis of Oculocutaneous Albinism: A Compact Review', *BioMed Research International,* 2014, pp. 905472.

Kammandel, B., Chowdhury, K., Stoykova, A., Aparicio, S., Brenner, S. and Gruss, P. (1999) 'Distinctcis-Essential Modules Direct the Time–Space Pattern of thePax6Gene Activity', *Developmental Biology,* 205(1), pp. 79-97.

Kanda, A., Friedman James, S., Nishiguchi Koji, M. and Swaroop, A. (2007) 'Retinopathy mutations in the bZIP protein NRL alter phosphorylation and transcriptional activity', *Human Mutation,* 28(6), pp. 589-598.

Kaplan, J., De Domenico, I. and Ward, D. M. (2008) 'Chediak-Higashi syndrome', *Currrent Opinions in Hematology,* 15.

Karaca, E. E., Çubuk, M. Ö., Ekici, F., Akçam, H. T., Waisbourd, M. and Hasanreisoglu, M. (2014) 'Isolated Foveal Hypoplasia: Clinical Presentation and Imaging Findings', *Optometry and Vision Science,* 91(4).

Karakousis, P. C., John, S. J., Behling, K. C., Surface, E. M., Smith, J. E., Hendrickson, A., Tang, W. X., Bennett, J. and Milam, A. H. (2001) 'Localisation of pigment epithelium derived factor (PEDF) in developing and adult human ocular tissues.', *Molecular Vision,* 7, pp. 154-163.

Katsanis, S. H. and Katsanis, N. (2013) 'Molecular genetic testing and the future of clinical genomics', *Nature Reviews Genetics,* 14(6), pp. 415-426.

Kausar, T., Bhatti, M. A., Ali, M., Sheikh, R. S. and Ahmed, Z. M. (2013) 'OCA5, a novel locus for non-syndromic oculocutaneous albinism, maps to chromosome 4q24.', *Clinical Genetics,* 84(1), pp. 91-93.

Kearse, M. G. and Wilusz, J. E. (2017) 'Non-AUG translation: a new start for protein synthesis in eukaryotes', *Genes & Development,* 31(17), pp. 1717-1731.

Kefalov, V. J. (2012) 'Rod and Cone Visual Pigments and Phototransduction through Pharmacological, Genetic, and Physiological Approaches', *Journal of Biological Chemistry,* 287(3), pp. 1635-1641.

Khan, A. O., Aldahmesh, M. A., Noor, J., Salem, A. and Alkuraya, F. S. (2015) 'Lens Subluxation and Retinal Dysfunction in a Girl with Homozygous VSX2 Mutation', *Ophthalmic Genetics,* 36(1), pp. 8-13.

Kidd, J. M., Cooper, G. M., Donahue, W. F., Hayden, H. S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., Haugen, E., Zerr, T.,

Yamada, N. A., Tsang, P., Newman, T. L., Tüzün, E., Cheng, Z., Ebling, H. M., Tusneem, N., David, R., Gillett, W., Phelps, K. A., Weaver, M., Saranga, D., Brand, A., Tao, W., Gustafson, E., McKernan, K., Chen, L., Malig, M., Smith, J. D., Korn, J. M., McCarroll, S. A., Altshuler, D. A., Peiffer, D. A., Dorschner, M., Stamatoyannopoulos, J., Schwartz, D., Nickerson, D. A., Mullikin, J. C., Wilson, R. K., Bruhn, L., Olson, M. V., Kaul, R., Smith, D. R. and Eichler, E. E. (2008) 'Mapping and sequencing of structural variation from eight human genomes', *Nature,* 453(7191), pp. 56-64.

Kima, P. E., Burleigh, B. and Andrews, N. W. (2001) 'Surface-targeted lysosomal membrane glycoprotein-1 (Lamp-1) enhances lysosome exocytosis and cell invasion by Trypanosoma cruzi', *Cellular Microbiology,* 2(6), pp. 477-486.

Kinnear, P. E., Jay, B. and Witkop, C. J. J. (1985) 'Albinism', *Survey of Ophthalmology,* 30(2), pp. 75-101.

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M. and Shendure, J. (2014) 'A general framework for estimating the relative pathogenicity of human genetic variants', *Nature Genetics,* 46(3), pp. 310-315.

Kirschner, R., Rosenberg, T., Schultz-Heienbrok, R., Lenzner, S., Feil, S., Roepman, R., Cremers, F. P., Ropers, H. H. and Berger, W. (1999) 'RPGR transcription studies in mouse and human tissues reveal a retina-specific isoform that is disrupted in a patient with X-linked retinitis pigmentosa.', *Human Molecular Genetics,* 8(8), pp. 1571-1578.

Klein, M. L., Schultz, D. W., Edwards, A. and et al. (1998) 'Age-related macular degeneration: Clinical features in a large family and linkage to chromosome 1q', *Archives of Ophthalmology,* 116(8), pp. 1082-1088.

Kleinjan, D. A., Schedl, A., Quinlan, R. A., Danes, S. and van Heyningen, V. (2001) 'Aniridia-associated translocations, DNase hypersensitivity, sequence comparison and transgenic analysis redefine the functional domain of PAX6', *Human Molecular Genetics,* 10, pp. 2049-2059.

Kleinjan, D. A., Seawright, A., Childs, A. J. and van Heyningen, V. (2004) 'Conserved elements in Pax6 intron 7 involved in (auto)regulation and alternative transcription', *Developmental Biology,* 265(2), pp. 462-477.

Kleinjan, D. A., Seawright, A., Mella, S., Carr, C. B., Tyas, D. A., Simpson, T. I., Mason, J. O., Price, D. J. and van Heyningen, V. (2006) 'Long-range downstream enhancers are essential for Pax6 expression', *Developmental Biology,* 299(2), pp. 563-581.

Kleinjan, D. J. and Coutinho, P. (2009) 'Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease.', *Briefings in Functional Genomics and Proteomics,* 8(4), pp. 317-332.

Koboldt, Daniel C., Larson, David E., Sullivan, Lori S., Bowne, Sara J., Steinberg, Karyn M., Churchill, Jennifer D., Buhr, Aimee C., Nutter, N., Pierce, Eric A., Blanton, Susan H., Weinstock, George M., Wilson, Richard K. and Daiger, Stephen P. (2014) 'Exome-Based Mapping and Variant Prioritization for Inherited Mendelian Disorders', *The American Journal of Human Genetics,* 94(3), pp. 373-384.

Kochetov, A. V. (2008) 'Alternative translation start sites and hidden coding potential of eukaryotic mRNAs.', *Bioessays,* 30(7), pp. 683-691.

Kochetov, A. V., Sarai, A., Rogozin, I. B., Shumny, V. K. and Kolchanov, N. A. (2005) 'The role of alternative translation start sites in the generation of human protein diversity.', *Molecular Genetics and Genomics,* 273(6), pp. 491-496.

Koga, M., Matsuda, M., Kawamura, T., Sogo, T., Shigeno, A., Nishida, E. and Ebisuya, M. (2014) 'Foxd1 is a mediator and indicator of the cell reprogramming process', *Nature Communications,* 5, pp. 3197.

Kohl, S., Marx, T., Giddings, I., Jägle, H., Jacobson, S. G., Apfelstedt-Sylla, E., Zrenner, E., Sharpe, L. T. and Wissinger, B. (1998) 'Total colourblindness is caused by mutations in the gene encoding the α-subunit of the cone photoreceptor cGMP-gated cation channel', *Nature Genetics,* 19, pp. 257.

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. and Liu, D. R. (2016) 'Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage', *Nature,* 533(7603), pp. 420-424.

Kong, A., Gudbjartsson, D. F., Sainz, J., Jonsdottir, G. M., Gudjonsson, S. A., Richardsson, B., Sigurdardottir, S., Barnard, J., Hallbeck, B. and Masson, G. (2002) 'A high-resolution recombination map of the human genome', *Nature Genetics,* 31(3), pp. 241-247.

Kornfeld, S. and Mellman, I. (1989) 'The Biogenesis of Lysosomes', *Annual Review of Cell Biology,* 5(1), pp. 483-525.

Kozak, M. (1986) 'Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes', *Cell,* 44(2), pp. 283-292.

Kozak, M. (1987) 'An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs', *Nucleic Acids Research,* 15(20), pp. 8125-8148.

Kozak, M. (1991) 'Structural features in eukaryotic mRNAs that modulate the initiation of translation.', *Journal of Biological Chemistry,* 266(30), pp. 19867-19870.

Kozak, M. (1995) 'Adherence to the first-AUG rule when a second AUG codon follows closely upon the first.', *Proceedings of the National Academy of Science,* 92, pp. 2662-2666.

Kozak, M. (1997) 'Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6', *The EMBO Journal,* 16(9), pp. 2482-2492.

Kozak, M. (2002) 'Pushing the limits of the scanning mechanism for initiation of translation', *Gene,* 299(1), pp. 1-34.

Kozmik, Z., Czerny, T. and Busslinger, M. (1997) 'Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of Pax6 and Pax8', *The EMBO Journal,* 16(22), pp. 6793-6803.

Kozulin, P., Natoli, R., Madigan, M. C., O'Brien, K. M. B. and Provis, J. M. (2009a) 'Gradients of Eph-A6 expression in primate retina suggest roles in both vascular and axon guidance', *Molecular Vision,* 15, pp. 2649-2662.

Kozulin, P., Natoli, R., O'Brien, K. M. B., Madigan, M. C. and Provis, J. M. (2009b) 'Differential expression of anti-angiogenic factors and guidance genes in the developing macula', *Molecular Vision,* 15, pp. 45-59.

Kozulin, P. and Provis, J. M. (2009) 'Differential gene expression in the developing human macula: microarray analysis using rare tissue samples', *Journal of Ocular Biology, Diseases, and Informatics,* 2(4), pp. 176-189.

Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. (2001) 'Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes .', *Journal of Molecular Biology,* 305, pp. 567-580.

Krumm, N., Sudmant, P. H., Ko, A., O'Roak, B. J., Malig, M., Coe, B. P., Project, N. E. S., Quinlan, A. R., Nickerson, D. A. and Eichler, E. E. (2012) 'Copy number variation detection and genotyping from exome sequence data', *Genome Research,* 22(8), pp. 1525-1532.

Krumpaszky, H. G., Lüdtke, R., Mickler, A., Klauss, V. and Selbmann, H. K. (1999) 'Blindness Incidence in Germany', *Ophthalmologica,* 213(3), pp. 176-182.

Kumar, P., Henikoff, S. and Ng, P. C. (2009) 'Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm', *Nature Protocols,* 4(8), pp. 1073-1081.

Kundra, R. and Kornfeld, S. (1999) 'Asparagine-linked Oligosaccharides Protect Lamp-1 and Lamp-2 from Intracellular Proteolysis', *Journal of Biological Chemistry,* 274(43), pp. 31039-31046.

Kupfer, C., Chumbley, L. and Downer, J. C. (1967) 'Quantitative histology of optic nerve, optic tract and lateral geniculate nucleus of man', *Journal of Anatomy,* 101(Pt 3), pp. 393-401.

Kurugol, Z., Ozkinay, F., Vardar, F., Karacali, S., Kutukculer, N. and Deveci, R. (2001) 'Griscelli syndrome: Report of a case and review of the literature', *Pediatrics International,* 43, pp. 298-301.

Käll, L., Krogh, A. and Sonnhammer, E. L. (2004) 'A combined transmembrane topology and signal peptide prediction method.', *Journal of Molecular Biology,* 338(5), pp. 1027-1036.

Käsmann-Kellner, B. and Seitz, B. (2007) 'Phenotype of the visual system in oculocutaneous and ocular albinism', *Der Ophthalmologe: Zeitschrift der Deutschen Ophthalmologischen Gesellschaft,* 104(8), pp. 648-661.

Köberlein, J., Beifus, K., Schaffert, C. and Finger, R. P. (2013) 'The economic burden of visual impairment and blindness: a systematic review', *BMJ Open,* 3(11), pp. e003471.

La Vail, M. M., Rapaport, D. H. and Rakic, P. (1991) 'Cytogenesis in the monkey retina', *The Journal of Comparative Neurology,* 309(1), pp. 86-114.

Lai, E., Clark, K. L., Burley, S. K. and Darnell, J. E. (1993) 'Hepatocyte nuclear factor 3/fork head or "winged helix" proteins: a family of transcription factors of diverse biologic function', *Proceedings of the National Academy of Sciences,* 90(22), pp. 10421-10423.

Laissue, P., Lakhal, B., Vatin, M., Batista, F., Burgio, G., Mercier, E., Santos, E. D., Buffat, C., Sierra-Diaz, D. C., Renault, G., Montagutelli, X., Salmon, J., Monget, P., Veitia, R. A., Méhats, C., Fellous, M., Gris, J.-C., Cocquet, J. and Vaiman, D. (2016) 'Association of FOXD1 variants with adverse pregnancy outcomes in mice and humans', *Open Biology,* 6(10), pp. 160109.

Lamb, T. D., Collin, S. P. and Pugh, E. N. (2007) 'Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup', *Nature Reviews Neuroscience,* 8(12), pp. 960-976.

Lambot, M.-A., Depasse, F., Noel, J.-C. and Vanderhaeghen, P. (2005) 'Mapping Labels in the Human Developing Visual System and the Evolution of Binocular Vision', *The Journal of Neuroscience,* 25(31), pp. 7232.

Lander, E. S. and Linton, L. M. and Birren, B. and Nusbaum, C. and Zody, M. C. and Baldwin, J. and Devon, K. and Dewar, K. and Doyle, M. and Fitzhugh, W. and Funke, R. and Gage, D. and Harris, K. and Heaford, A. and Howland, J. and Kann, L. and Lehoczky, J. and LeVine, R. and McEwan, P. and McKernan, K. and Meldrim, J. and Mesirov, J. P. and Miranda, C. and Morris, W. and Naylor, J. and Raymond, C. and Rosetti, M. and Santos, R. and Sheridan, A. and Sougnez, C. and Stange-Thomann, N. and Stojanovic, N. and Subramanian, A. and Wyman, D. and Rogers, J. and Sulston, J. and Ainscough, R. and Beck, S. and Bentley, D. and Burton, J. and Clee, C. and Carter, N. and Coulson, A. and Deadman, R. and Deloukas, P. and Dunham, A. and Dunham, I. and Durbin, R. and French, L. and Grafham, D. and

Gregory, S. and Hubbard, T. and Humphray, S. and Hunt, A. and Jones, M. and Lloyd, C. and McMurray, A. and Matthews, L. and Mercer, S. and Milne, S. and Mullikin, J. C. and Mungall, A. and Plumb and Ross, M. and Shownkeen, R. and Sims, S. and Waterston, R. H. and Wilson, R. K. and Hillier, L. W. and McPherson, J. D. and Marra, M. A. and Mardis, E. R. and Fulton, L. A. and Chinwalla, A. T. and Pepin, K. H. and Gish, W. R. and Chissoe, S. L. and Wendl, M. C. and Delehaunty, K. D. and Miner, T. L. and Delehaunty, A. and Kramer, J. B. and Cook, L. L. and Fulton, R. S. and Johnson, D. L. and Minx, P. J. and Clifton, S. W. and Hawkins, T. and Branscomb, E. and Predki, P. and Richardson, P. and Wenning, S. and Slezak, T. and Doggett, N. and Cheng, J.-F. and Olsen, A. and Lucas, S. and Elkin, C. and Uberbacher, E. and Frazier, M. and Gibbs, R. A. and Muzny, D. M. and Scherer, S. E. and Bouck, J. B. and Sodergren, E. J. and Worley, K. C. and Rives, C. M. and Gorrell, J. H. and Metzker, M. L. and Naylor, S. L. and Kucherlapati, R. S. and Nelson, D. L. and Weinstock, G. M. and Sakaki, Y. and Fujiyama, A. and Hattori, M. and Yada, T. and Toyoda, A. and Itoh, T. and Kawagoe, C. and Watanabe, H. and Totoki, Y. and Taylor, T. and Weissenbach, J. and Heilig, R. and Saurin, W. and Artiguenave, F. and Brottier, P. and Bruls, T. and Pelletier, E. and Robert, C. and Wincker, P. and Rosenthal, A. and Platzer, M. and Nyakatura, G. and Taudien, S. and Rump, A. and Smith, D. R. and Doucette-Stamm, L. and Rubenfield, M. and Weinstock, K. and Lee, H. M. and Dubois, J. and Yang, H. and Yu, J. and Wang, J. and Huang, G. and Gu, Y. and Hood, L. and Rowen, L. and Madan, A. and Qin, S. and Davis, R. W. and Federspiel, N. A. and Abola, A. P. and Proctor, M. J. and Roe, B. A. and Chen, F. and Pan, H. and Ramser, J. and Lehrach, H. and Reinhardt, R. and McCombie, W. R. and de la Bastide, M. and Dedhia, N. and Blöcker, H. and Hornischer, K. and Nordsiek, G. and Agarwala, R. and Aravind, L. and Bailey, J. A. and Bateman, A. and Batzoglou, S. and Birney, E. and Bork, P. and Brown, D. G. and Burge, C. B. and Cerutti, L. and Chen, H.-C. and Church, D. and Clamp, M. and Copley, R. R. and Doerks, T. and Eddy, S. R. and Eichler, E. E. and Furey, T. S. and Galagan, J. and Gilbert, J. G. R. and Harmon, C. and Hayashizaki, Y. and Haussler, D. and Hermjakob, H. and Hokamp, K. and Jang, W. and Johnson, L. S. and Jones, T. A. and Kasif, S. and Kaspryzk, A. and Kennedy, S. and Kent, W. J. and Kitts, P. and Koonin, E. V. and Korf, I. and Kulp, D. and Lancet, D. and Lowe, T. M. and McLysaght, A. and Mikkelsen, T. and Moran, J. V. and Mulder, N. and Pollara, V. J. and Ponting, C. P. and Schuler, G. and Schultz, J. and Slater, G. and Smit, A. F. A. and Stupka, E. and Szustakowki, J. and Thierry-Mieg, D. and Thierry-Mieg, J. and Wagner, L. and Wallis, J. and Wheeler, R. and Williams, A. and Wolf, Y. I. and Wolfe, K. H. and Yang, S.-P. and Yeh, R.-F. and Collins, F. and Guyer, M. S. and Peterson, J. and Felsenfeld, A. and Wetterstrand, K. A. and Myers, R. M. and Schmutz, J. and Dickson, M. and Grimwood, J. and Cox, D. R. and Olson, M. V. and Kaul, R. and Raymond, C. and Shimizu, N. and Kawasaki, K. and Minoshima, S. (2001) 'Initial sequencing and analysis of the human genome', *Nature,* 409(6822), pp. 860-921.

Landrum, M. J., Lee, J. M., Benson, M., Brown, G., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Hoover, J., Jang, W., Katz, K., Ovetsky, M., Riley, G., Sethi, A., Tully, R., Villamarin-Salomon, R., Rubinstein, W. and Maglott, D. R. (2016) 'ClinVar: public archive of interpretations of clinically relevant variants', *Nucleic Acids Research,* 44(Database issue), pp. D862-D868.

Lassen, N., Pappa, A., Black, W. J., Jester, J. V., Day, B. J., Min, E. and Vasiliou, V. (2006) 'Antioxidant function of corneal ALDH3A1 in cultured stromal fibroblasts.', *Free Radical Biology & Medicine,* 41(9), pp. 1459-1469.

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

Leber, T. (1869) 'Ueber Retinitis pigmentosa und angeborene Amaurose', *Archiv für Ophthalmologie,* 15(3), pp. 1-25.

Lee, J. A., Carvalho, C. M. B. and Lupski, J. R. (2007) 'A DNA Replication Mechanism for Generating Nonrecurrent Rearrangements Associated with Genomic Disorders', *Cell,* 131(7), pp. 1235-1247.

Lee, J. A., Inoue, K., Cheung, S. W., Shaw, C. A., Stankiewicz, P. and Lupski, J. R. (2006) 'Role of genomic architecture in PLP1 duplication causing Pelizaeus-Merzbacher disease', *Human Molecular Genetics,* 15(14), pp. 2250-2265.

Lee, K., Berg, J. S., Milko, L., Crooks, K., Lu, M., Bizon, C., Owen, P., Wilhelmsen, K. C., Weck, K. E., Evans, J. P. and Garg, S. (2015) 'High Diagnostic Yield of Whole Exome Sequencing in Participants with Retinal Dystrophies in a Clinical Ophthalmology Setting', *American Journal of Ophthalmology,* 160(2), pp. 354-363.e9.

Lefler, W. H., Wadsworth, J. A. and Sidbury, J. B. J. (1971) 'Hereditary macular degeneration and amino-aciduria.', *American Journal of Ophthalmology,* (71), pp. 224-230.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., MacArthur, D. G. and Exome Aggregation, C. (2016) 'Analysis of protein-coding genetic variation in 60,706 humans', *Nature,* 536(7616), pp. 285-291.

Lettice, L. A., Hill, A. E., Devenney, P. S. and Hill, R. E. (2008) 'Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly', *Human Molecular Genetics,* 17(7), pp. 978-985.

Lettice, L. A., Horikoshi, T., Heaney, S. J. H., van Baren, M. J., van der Linde, H. C., Breedveld, G. J., Joosse, M., Akarsu, N., Oostra, B. A., Endo, N., Shibata, M., Suzuki, M., Takahashi, E., Shinka, T., Nakahori, Y., Ayusawa, D., Nakabayashi, K., Scherer, S. W., Heutink, P., Hill, R. E. and Noji, S. (2002) 'Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly', *Proceedings of the National Academy of Sciences of the United States of America,* 99(11), pp. 7548-7553.

Leventhal, A. G., J., A. S., J, V. D. and Tiande, S. (1989) 'Extrinsic determinants of retinal ganglion cell development in primates', *The Journal of Comparative Neurology,* 286(2), pp. 170--189.

Lewis, V., Green, S. A., Marsh, M., Vihko, P., Helenius, A. and Mellman, I. (1985) 'Glycoproteins of the lysosomal membrane', *The Journal of Cell Biology,* 100(6), pp. 1839-1847.

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. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.', *ARXIV*, pp. 1-3.

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

Li, M. D., Beuten, J., Ma, J. Z., Payne, T. J., Lou, X.-Y., Garcia, V., Duenes, A. S., Crews, K. M. and Elston, R. C. (2005) 'Ethnic- and gender-specific association of the nicotinic acetylcholine receptor α4 subunit gene (CHRNA4) with nicotine dependence', *Human Molecular Genetics,* 14(9), pp. 1211-1219.

Li, W., Zhang, Q., Oiso, N., Novak, E. K., Gautam, R., O'Brien, E. P., Tinsley, C. L., Blake, D. J., Spritz, R. A., Copeland, N. G., Jenkins, N. A., Amato, D., Roe, B. A., Starcevic, M., Dell'Angelica, E. C., Elliott, R. W., Mishra, V., Kingsmore, S. F., Paylor, R. E. and Swank, R. T. (2003) 'Hermansky-Pudlak syndrome type 7 (HPS-7) results from mutant dysbindin, a member of the biogenesis of lysosome-related organelles complex 1 (BLOC-1)', *Nature Genetics,* 35(1), pp. 84-89.

Li, Z., Weng, H., Su, R., Weng, X., Zuo, Z., Li, C., Huang, H., Nachtergaele, S., Dong, L., Hu, C., Qin, X., Tang, L., Wang, Y., Hong, G.-M., Huang, H., Wang, X., Chen, P., Gurbuxani, S., Arnovitz, S., Li, Y., Li, S., Strong, J., Neilly, M. B., Larson, R. A., Jiang, X., Zhang, P., Jin, J., He, C. and Chen, J. (2017) 'FTO Plays an Oncogenic Role in Acute Myeloid Leukemia as a N6-Methyladenosine RNA Demethylase', *Cancer Cell,* 31(1), pp. 127-141.

Liew, G., Michaelides, M. and Bunce, C. (2014) 'A comparison of the causes of blindness certifications in England and Wales in working age adults (16–64 years), 1999–2000 with 2009–2010', *BMJ Open,* 4(2), pp. e004015.

Lin, S., Choe, J., Du, P., Triboulet, R. and Gregory, R. I. (2016) 'METTL3 promotes translation in human cancer cells', *Molecular Cell,* 62(3), pp. 335-345.

Lippincott-Schwartz, J. and Fambrough, D. M. (1986) 'Lysosomal membrane dynamics: structure and interorganellar movement of a major lysosomal membrane glycoprotein', *The Journal of Cell Biology,* 102(5), pp. 1593-1605.

Lisabeth, E. M., Falivelli, G. and Pasquale, E. B. (2013) 'Eph Receptor Signaling and Ephrins.', *Cold Spring Harbour Perspectives in Biology,* 5.

Littink, K. W., den Hollander, A. I., Cremers, F. P. M. and Collin, R. W. J. (2012) 'The Power of Homozygosity Mapping: Discovery of New Genetic Defects in Patients with Retinal Dystrophy.', *In: LaVail M., Ash J., Anderson R., Hollyfield J., Grimm C. (eds) Retinal Degenerative Diseases. Advances in Experimental Medicine and Biology*. Boston, MA: Springer.

Liu, I. S. C., Chen, J., Ploder, L., Vidgen, D., van der Kooy, D., Kalnins, V. I. and McInnes, R. R. (1994) 'Developmental expression of a novel murine homeobox gene (Chx10): Evidence for roles in determination of the neuroretina and inner nuclear layer.', *Neuron,* 13, pp. 377-393.

Liu, L., Dilworth, D., Gao, L., Monzon, J., Summers, A., Lassam, N. and Hogg, D. (1999) 'Mutation of the CDKN2A 5' UTR creates an aberrant initiation codon and predisposes to melanoma', *Nature Genetics,* 21, pp. 128.

Liu, M. M. and Zack, D. J. (2013) 'Alternative splicing and retinal degeneration', *Clinical Genetics,* 84(2), pp. 142-149.

Liu, X., Han, S., Wang, Z., Gelernter, J. and Yang, B.-Z. (2013) 'Variant callers for next-generation sequencing data: a comparison study', *PLoS One,* 8(9), pp. e75619-e75619.

Logan, C. V., Cossins, J., Rodríguez Cruz, P. M., Parry, D. A., Maxwell, S., Martínez-Martínez, P., Riepsaame, J., Abdelhamed, Z. A., Lake, A. V. R., Moran, M., Robb, S., Chow, G., Sewry, C., Hopkins, P. M., Sheridan, E., Jayawant, S., Palace, J., Johnson, C. A. and Beeson, D. (2015) 'Congenital Myasthenic Syndrome Type 19 Is Caused by Mutations in COL13A1, Encoding the Atypical Non-fibrillar Collagen Type XIII α1 Chain', *The American Journal of Human Genetics,* 97(6), pp. 878-885.

Loges, N. T., Olbrich, H., Becker-Heck, A., Häffner, K., Heer, A., Reinhard, C., Schmidts, M., Kispert, A., Zariwala, M. A., Leigh, M. W., Knowles, M. R., Zentgraf, H., Seithe, H., Nürnberg, G., Nürnberg, P., Reinhardt, R. and Omran, H. (2009) 'Deletions and Point Mutations of LRRC50 Cause Primary Ciliary Dyskinesia Due to Dynein Arm Defects', *The American Journal of Human Genetics,* 85(6), pp. 883-889.
Lowings, P., Yavuzer, U. and Goding, C. R. (1992) 'Positive and negative elements regulate a melanocyte-specific promoter', *Molecular and Cellular Biology,* 12(8), pp. 3653-3662.

Lozano, M. L., Rivera, J., Sánchez-Guiu, I. and Vicente, V. (2014) 'Towards the targeted management of Chediak-Higashi syndrome', *Orphanet Journal of Rare Diseases,* 9(1), pp. 132.

Luo, D., Chen, H. and Jimbow, K. (1994) 'Cotransfection of Genes Encoding Human Tyrosinase and Tyrosinase-Related Protein-1 Prevents Melanocyte Death and Enhances Melanin Pigmentation and Gene Expression of Lamp-1', *Experimental Cell Research,* 213(1), pp. 231-241.

Ma, W., Ay, F., Lee, C., Gulsoy, G., Deng, X., Cook, S., Hesson, J., Cavanaugh, C., Ware, C. B., Krumm, A., Shendure, J., Blau, C. A., Disteche, C. M., Noble, W. S. and Duan, Z. (2018) 'Using DNase Hi-C techniques to map global and local three-dimensional genome architecture at high resolution', *Methods,* 142, pp. 59-73.

MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., Abecasis, G. R., Adams, D. R., Altman, R. B., Antonarakis, S. E., Ashley, E. A., Barrett, J. C., Biesecker, L. G., Conrad, D. F., Cooper, G. M., Cox, N. J., Daly, M. J., Gerstein, M. B., Goldstein, D. B., Hirschhorn, J. N., Leal, S. M., Pennacchio, L. A., Stamatoyannopoulos, J. A., Sunyaev, S. R., Valle, D., Voight, B. F., Winckler, W. and Gunter, C. (2014) 'Guidelines for investigating causality of sequence variants in human disease', *Nature,* 508, pp. 469.

Mackay, D. S., Bennett, T. M., Culican, S. M. and Shiels, A. (2014) 'Exome sequencing identifies novel and recurrent mutations in GJA8 and CRYGD associated with inherited cataract', *Human Genomics,* 8(1), pp. 19.

Mackenzie, B. and Erickson, J. (2004) 'Sodium-coupled neutral amino acid (System N/A) transporters of the SLC38 gene family', *Pflügers Archiv,* 447(5), pp. 784-795.

MacLaren, R. E., Groppe, M., Barnard, A. R., Cottriall, C. L., Tolmachova, T., Seymour, L., Clark, K. R., During, M. J., Cremers, F. P. M., Black, G. C. M., Lotery, A. J., Downes, S. M., Webster, A. R. and Seabra, M. C. (2014) 'Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial', *Lancet,* 383(9923), pp. 1129-1137.

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

Makalowski, W. and Boguski, M. S. (1998) 'Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences', *Proceedings of the National Academy of Sciences,* 95(16), pp. 9407-9412.

Malhotra, D. and Sebat, J. (2012) 'CNVs: Harbinger of a Rare Variant Revolution in Psychiatric Genetics', *Cell,* 148(6), pp. 1223-1241.

Malicdan, M. C., Noguchi, S., Nonaka, I., Saftig, P. and Nishino, I. (2008) 'Lysosomal myopathies: An excessive build-up in autophagosomes is too much to handle', *Neuromuscular Disorders,* 18(7), pp. 521-529.

Mann, I. (1969) *The Development of the Human Eye.* Grune and Stratton.

Manzer, R., Pappa, A., Estey, T., Sladek, N., Carpenter, J. F. and Vasiliou, V. (2003) 'Ultraviolet radiation decreases expression and induces aggregation of corneal ALDH3A1.', *Chemico-biological Interactions,* 143-144, pp. 45-53.

Marcus, R. C., Shimamura, K., Sretavan, D., Lai, E., Rubenstein, J. L. R. and Mason, C. A. (1999) 'Domains of regulatory gene expression and the developing optic chiasm: Correspondence with retinal axon paths and candidate signalling cells.', *The Journal of Comparative Neurology,* 403(3), pp. 346-358.

Marks, M. S. and Seabra, M. C. (2001) 'The melanosome: membrane dynamics in black and white', *Nature Reviews Molecular Cell Biology,* 2, pp. 738.

Martínez, J. J., Toledo, J. F., Fernández, E. and Ferrández, J. M. (2008) 'A retinomorphic architecture based on discrete-time cellular neural networks using reconfigurable computing', *Neurocomputing,* 71(4), pp. 766-775.

Maruyama, K. and Sugano, S. (1994) 'Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides', *Gene,* 138(1), pp. 171-174.

Massey, S. C. (1990) 'Cell types using glutamate as a neurotransmitter in the vertebrate retina', *Progress in Retinal Research,* 9, pp. 399-425.

Matharu, N. K. and Ahanger, S. H. (2015) 'Chromatin Insulators and Topological Domains: Adding New Dimensions to 3D Genome Architecture', *Genes,* 6(3), pp. 790-811.

Matthijs, G., Souche, E., Alders, M., Corveleyn, A., Eck, S., Feenstra, I., Race, V., Sistermans, E., Sturm, M., Weiss, M., Yntema, H., Bakker, E., Scheffer, H. and Bauer, P. (2016) 'Guidelines for diagnostic next-generation sequencing', *The European Journal of Human Genetics,* 24(1), pp. 2-5.

Maurice, D. M. (1957) 'The structure and transparency of the cornea.', *The Journal of Physiology,* 136(2), pp. 263-286.

Mayer, A. K., Rohrschneider, K., Strom, T. M., Glöckle, N., Kohl, S., Wissinger, B. and Weisschuh, N. (2015) 'Homozygosity mapping and whole-genome sequencing reveals a deep intronic PROM1 mutation causing cone-rod dystrophy by pseudoexon activation', *The European Journal of Human Genetics*.

McBride, D. J., Buckle, A., van Heyningen, V. and Kleinjan, D. A. (2011) 'DNaseI Hypersensitivity and Ultraconservation Reveal Novel, Interdependent Long-Range Enhancers at the Complex Pax6 Cis-Regulatory Region', *PLoS One,* 6(12), pp. e28616.

McKay, J. D., Patterson, B., Craig, J. E., Russell-Eggitt, I. M., Wirth, M. G., Burdon, K. P., Hewitt, A. W., Cohn, A. C., Kerdraon, Y. and Mackey, D. A. (2005) 'The telomere of human chromosome 1p contains at least two independent autosomal dominant congenital cataract genes', *The British Journal of Ophthalmology,* 89(7), pp. 831-834.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S. and Daly, M. (2010) 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data', *Genome Research,* 20(9), pp. 1297-1303.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P. and Cunningham, F. (2010) 'Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor', *Bioinformatics,* 26(16), pp. 2069-2070.

McPherson, J. D. and Marra, M. and Hillier, L. and Waterston, R. H. and Chinwalla, A. and Wallis, J. and Sekhon, M. and Wylie, K. and Mardis, E. R. and Wilson, R. K. and Fulton, R. and Kucaba, T. A. and Wagner-McPherson, C. and Barbazuk, W. B. and Gregory, S. G. and Humphray, S. J. and French, L. and Evans, R. S. and Bethel, G. and Whittaker, A. and Holden, J. L. and McCann, O. T. and Dunham, A. and Soderlund, C. and Scott, C. E. and Bentley, D. R. and Schuler, G. and Chen, H. C. and Jang, W. and Green, E. D. and Idol, J. R. and Maduro, V. V. and Montgomery, K. T. and Lee, E. and Miller, A. and Emerling, S. and Kucherlapati and Gibbs, R. and Scherer, S. and Gorrell, J. H. and Sodergren, E. and Clerc-Blankenburg, K. and Tabor, P. and Naylor, S. and Garcia, D. and de Jong, P. J. and Catanese, J. J. and Nowak, N. and Osoegawa, K. and Qin, S. and Rowen, L. and Madan, A. and Dors, M. and Hood, L. and Trask, B. and Friedman, C. and Massa, H. and Cheung, V. G. and Kirsch, I. R. and Reid, T. and Yonescu, R. and Weissenbach, J. and Bruls, T. and Heilig, R. and Branscomb, E. and Olsen, A. and Doggett, N. and Cheng, J. F. and Hawkins, T. and Myers, R. M. and Shang, J. and Ramirez, L. and Schmutz, J. and Velasquez, O. and Dixon, K. and Stone, N. E. and Cox, D. R. and Haussler, D. and Kent, W. J. and Furey, T. and Rogic, S. and Kennedy, S. and Jones, S. and Rosenthal, A. and Wen, G. and Schilhabel, M. and Gloeckner, G. and Nyakatura, G. and Siebert, R. and Schlegelberger, B. and Korenberg, J. and Chen, X. N. and Fujiyama, A. and Hattori, M. and Toyoda, A. and Yada, T. and Park, H. S. and Sakaki, Y. and Shimizu, N. and Asakawa, S. and Kawasaki, K. and Sasaki, T. and Shintani, A. and Shimizu, A. and Shibuya, K. and Kudoh, J. and Minoshima, S. and Ramser, J. and Seranski, P. and Hoff, C. and Poustka, A. and Reinhardt, R. and Lehrach, H. and Consortium, I. H. G. M. (2001) 'A physical map of the human genome', *Nature,* 409(6822), pp. 934-941.

Mellott, M. L., Brown, J., Jr, Fingert, J. H. and et al. (1999) 'Clinical characterization and linkage analysis of a family with congenital x-linked nystagmus and deuteranomaly', *Archives of Ophthalmology,* 117(12), pp. 1630-1633.

Merin, S. and Auerbach, R. (1976) 'Retinitis pigmentosa', *Survey of Ophthalmology,* 20, pp. 303-346.

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E. and Jaffrey, S. R. (2012) 'Comprehensive Analysis of mRNA Methylation Reveals Enrichment in 3' UTRs and Near Stop Codons', *Cell,* 149(7), pp. 1635-1646.

Meynert, A. M., Ansari, M., FitzPatrick, D. R. and Taylor, M. S. (2014) 'Variant detection sensitivity and biases in whole genome and exome sequencing.', *BMC Bioinformatics,* 15(247).

Michaelides, M., Jeffery, G. and Moore, A. T. (2012) 'Developmental macular disorders: phenotypes and underlying molecular genetic basis', *British Journal of Ophthalmology,* 96(7), pp. 917-924.

Mignone, F., Gissi, C., Liuni, S. and Pesole, G. (2002) 'Untranslated regions of mRNAs', *Genome Biology,* 3(3), pp. reviews0004.1-reviews0004.10.

Montoliu, T., Gronskov, K., Wei, A.-H., Martinez-Garcia, M., Fernandez, A., Arveiler, B., Morice-Picard, F., Riazuddin, S., Suzuki, T., Ahmed, Z. M., Rosenberg, T. and Li,

W. (2013) 'Increasing the complexity: new genes and new types of albinism.', *Pigment Cell & Melanoma Research,* 27, pp. 11-18.

Morimura, H., Fishman, G. A., Grover, S. A., Fulton, A. B., Berson, E. L. and Dryja, T. P. (1998) 'Mutations in the RPE65 gene in patients with autosomal recessive retinitis pigmentosa or Leber congenital amaurosis', *Proceedings of the National Academy of Sciences,* 95(6), pp. 3088-3093.

Mota, A., Fonseca, S., Carneiro, A., Magalhaes, A., Brandao, E. and Falcao-Reis, F. (2012) 'Isolated Foveal Hypoplasia: Tomographic, Angiographic and Autofluorescence Patterns', *Case Reports in Ophthalmological Medicine,* 2012, pp. 3.

Murphy, D., Cieply, B., Carstens, R., Ramamurthy, V. and Stoilov, P. (2016) 'The Musashi 1 Controls the Splicing of Photoreceptor-Specific Exons in the Vertebrate Retina', *PLoS Genetics,* 12(8), pp. e1006256.

Ménasché, G., Pastural, E., Feldmann, J., Certain, S., Ersoy, F., Dupuis, S., Wulffraat, N., Bianchi, D., Fischer, A., Le Deist, F. and de Saint Basile, G. (2000) 'Mutations in RAB27A cause Griscelli syndrome associated with haemophagocytic syndrome', *Nature Genetics,* 25, pp. 173.

Nakagawa, S., Brennan, C., Johnson, K. G., Shewan, D., Harris, W. A. and Holt, C. E. (2000) 'Ephrin-B Regulates the Ipsilateral Routing of Retinal Axons at the Optic Chiasm', *Neuron,* 25(3), pp. 599-610.

Nakagawa, S., Niimura, Y., Gojobori, T., Tanaka, H. and Miura, K.-i. (2008) 'Diversity of preferred nucleotide sequences around the translation initiation codon in eukaryote genomes', *Nucleic Acids Research,* 36(3), pp. 861-871.

Naylor, J., Brinke, A., Hassock, S., Green, P. M. and Giannelli, F. (1993) 'Characteristic mRNA abnormality found in half the patients with severe haemophilia A is due to large DNA inversions.', *Human Molecular Genetics,* 11, pp. 1773-1778.

Neveling, K., Collin, R. W. J., Gilissen, C., van Huet, R. A. C., Visser, L., Kwint, M. P., Gijsen, S. J., Zonneveld, M. N., Wieskamp, N., de Ligt, J., Siemiatkowska, A. M., Hoefsloot, L. H., Buckley, M. F., Kellner, U., Branham, K. E., den Hollander, A. I., Hoischen, A., Hoyng, C., Klevering, B. J., van den Born, L. I., Veltman, J. A., Cremers, F. P. M. and Scheffer, H. (2012) 'Next-generation genetic testing for retinitis pigmentosa', *Human Mutation,* 33(6), pp. 963-972.

Neveu, M. M. and Jeffery, G. (2007) 'Chiasm formation in man is fundamentally different from that in the mouse', *Eye,* 21, pp. 1264.

Newman, S., Hermetz, Karen E., Weckselblatt, B. and Rudd, M K. (2015) 'Next-Generation Sequencing of Duplication CNVs Reveals that Most Are Tandem and Some Create Fusion Genes at Breakpoints', *The American Journal of Human Genetics,* 96(2), pp. 208-220.

Ng, P. C. and Henikoff, S. (2003) 'SIFT: predicting amino acid changes that affect protein function', *Nucleic Acids Research,* 31(13), pp. 3812-3814.

Ng, S. B., Buckingham, K. J., Lee, C., Bigham, A. W., Tabor, H. K., Dent, K. M., Huff, C. D., Shannon, P. T., Jabs, E. W., Nickerson, D. A., Shendure, J. and Bamshad, M. J. (2010) 'Exome sequencing identifies the cause of a Mendelian disorder', *Nature genetics,* 42(1), pp. 30-35.

Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A. and Shendure, J. (2009) 'Targeted Capture and Massively Parallel Sequencing of Twelve Human Exomes', *Nature,* 461(7261), pp. 272-276.

Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) 'Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.', *Protein Engineering,* 10(1), pp. 1-6.

Nishiguchi, K. M., Tearle, R. G., Liu, Y. P., Oh, E. C., Miyake, N., Benaglio, P., Harper, S., Koskiniemi-Kuendig, H., Venturini, G., Sharon, D., Koenekoop, R. K., Nakamura, M., Kondo, M., Ueno, S., Yasuma, T. R., Beckmann, J. S., Ikegawa, S., Matsumoto, N., Terasaki, H., Berson, E. L., Katsanis, N. and Rivolta, C. (2013) 'Whole genome sequencing in patients with retinitis pigmentosa reveals pathogenic DNA structural changes and NEK2 as a new disease gene', *Proceedings of the National Academy of Sciences,* 110(40), pp. 16139-16144.

Noderer, W. L., Flockhart, R. J., Bhaduri, A., Diaz de Arce, A. J., Zhang, J., Khavari, P. A. and Wang, C. L. (2014) 'Quantitative analysis of mammalian translation initiation sites by FACS-seq', *Molecular Systems Biology,* 10(8), pp. 748.
O'Donnell, F. E. and Pappas, H. R. (1982) 'Autosomal dominant foveal hypoplasia and presenile cataracts. A new syndrome.', *Archives of Ophthalmology,* 100(2), pp. 279-281.

Oetting, W. S. (2006) *Albinism: genetics.*: eLS.

Oetting, W. S. and King, R. A. (1999) 'Molecular basis of albinism: mutations and polymorphisms of pigmentation genes associated with albinism', *Human Mutation,* 13(2), pp. 99.

Oetting, W. S., Summers, C. G. and King, R. A. (1994) 'Albinism and the associated ocular defects.', *Metabolic, Paediatric and Systemic Ophthalmology (1985),* 17(1-4), pp. 5-9.

Ogawa, K., Pasqualini, R., Lindberg, R. A., Kain, R., Freeman, A. L. and Pasquale, E. B. (2000) 'The ephrin-A1 ligand and its receptor, EphA2, are expressed during tumor neovascularization', *Oncogene,* 19, pp. 6043.

Oh, J., Bailin, T., Fukai, K., Feng, G. H., Ho, L., Mao, J.-i., Frenk, E., Tamura, N. and Spritz, R. A. (1996) 'Positional cloning of a gene for Hermansky–Pudlak syndrome, a disorder of cytoplasmic organelles', *Nature Genetics,* 14, pp. 300.

Oliver, M. D., Dotan, S. A., Chemke, J. and Abraham, F. A. (1987) 'Isolated foveal hypoplasia', *The British Journal of Ophthalmology,* 71(12), pp. 926-930.

Omasits, U., Ahrens, C. H., Müller, S. and Wollscheid, B. (2014) 'Protter: interactive protein feature visualization and integration with experimental proteomic data.', *Bioinformatics,* 30(6), pp. 884-886.

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M. R., Zschocke, J. and Trajanoski, Z. (2014) 'A survey of tools for variant analysis of next-generation genome sequencing data.', *Briefings in Bioinformatics,* 15(2), pp. 256-278.

Pal, B., Mohamed, M., Keen, T., Williams, G., Bradbury, J., Sheridan, E. and Inglehearn, C. (2004) 'A new phenotype of recessively inherited foveal hypoplasia and anterior segment dysgenesis maps to a locus on chromosome 16q23.2–24.2', *Journal of Medical Genetics,* 41(10), pp. 772-777.

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', *The American Journal of Human Genetics,* 100(6), pp. 960-968.

Pappa, A., Brown, D., Koutalos, Y., DeGregori, J., White, C. and Vasiliou, V. (2005) 'Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes

survival of human corneal epithelial cells.', *The Journal of Biological Chemistry,* 280(30), pp. 27998-28006.

Parfitt, David A., Lane, A., Ramsden, Conor M., Carr, A.-Jayne F., Munro, Peter M., Jovanovic, K., Schwarz, N., Kanuga, N., Muthiah, Manickam N., Hull, S., Gallo, J.-M., da Cruz, L., Moore, Anthony T., Hardcastle, Alison J., Coffey, Peter J. and Cheetham, Michael E. (2016) 'Identification and Correction of Mechanisms Underlying Inherited Blindness in Human iPSC-Derived Optic Cups', *Cell Stem Cell,* 18(6), pp. 769-781.

Parry, D. A., Brookes, S. J., Logan, C. V., Poulter, J. A., El-Sayed, W., Al-Bahlani, S., Al Harasi, S., Sayed, J., Raïf, E. M., Shore, R. C., Dashash, M., Barron, M., Morgan, J. E., Carr, I. M., Taylor, G. R., Johnson, C. A., Aldred, M. J., Dixon, M. J., Wright, J. T., Kirkham, J., Inglehearn, C. F. and Mighell, A. J. (2012) 'Mutations in C4orf26, encoding a peptide with in vitro hydroxyapatite crystal nucleation and growth activity, cause amelogenesis imperfecta', *The American Journal of Human Genetics,* 91(3), pp. 565-571.

Parsons, M. T., Whiley, P. J., Beesley, J., Drost, M., de Wind, N., Thompson, B. A., Marquart, L., Hopper, J. L., Jenkins, M. A., Australasian Colorectal Cancer Family, R., Brown, M. A., Tucker, K., Warwick, L., Buchanan, D. D. and Spurdle, A. B. (2015) 'Consequences of germline variation disrupting the constitutional translational initiation codon start sites of MLH1 and BRCA2: use of potential alternative start sites and implications for predicting variant pathogenicity', *Molecular Carcinogenesis,* 54(7), pp. 513-522.

Pasquale, E. B. (2005) 'Eph receptor signalling casts a wide net on cell behaviour. ', *Nature Reviews. Molecular Cell Biology,* 6(6), pp. 462-475.

Pastural, E., Barrat, F. J., Dufourcq-Lagelouse, R., Certain, S., Sanal, O., Jabado, N., Seger, R., Griscelli, C., Fischer, A. and Basile, G. d. S. (1997) 'Griscelli disease maps to chromosome 15q21 and is associated with mutations in the Myosin-Va gene', *Nature Genetics,* 16, pp. 289.

Patel-Hett, S. and D'Amore, P. A. (2011) 'Signal Transduction in Vasculogenesis and Developmental Angiogenesis', *The International Journal of Developmental Biology,* 55(0), pp. 353-363.

Peng, J., Fabre, P. J., Dolique, T., Swikert, S. M., Kermasson, L., Shimogori, T. and Charron, F. (2018) 'Sonic Hedgehog Is a Remotely Produced Cue that Controls Axon Guidance Trans-axonally at a Midline Choice Point', *Neuron,* 97(2), pp. 326-340.e4.

Perez, Y., Gradstein, L., Flusser, H., Markus, B., Cohen, I., Langer, Y., Marcus, M., Lifshitz, T., Kadir, R. and Birk, O. S. (2014) 'Isolated foveal hypoplasia with secondary nystagmus and low vision is associated with a homozygous SLC38A8 mutation', *The European Journal of Human Genetics,* 22(5), pp. 703-706.

Perou, C. M., Moore, K. J., Nagle, D. L., Misumi, D. J., Woolf, E. A., McGrail, S. H., Holmgren, L., Brody, T. H., Dussault, B. J., Monroe, C. A., Duyk, G. M., Pryor, R. J., Li, L., Justice, M. J. and Kaplan, J. (1996) 'Identification of the murine beige gene by YAC complementation and positional cloning', *Nature Genetics,* 13.

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F. and Liuni, S. (2001) 'Structural and functional features of eukaryotic mRNA untranslated regions', *Gene,* 276(1), pp. 73-81.

Petros, T. J., Rebsam, A. and Mason, C. A. (2008) 'Retinal Axon Growth at the Optic Chiasm: To Cross or Not to Cross', *Annual Review of Neuroscience,* 31(1), pp. 295-315.

Phillips, J., Perez, E., Wallace, K., Martin, J., Singh, R., Capowski, E., Wright, L., Clark, E., Percin, E. and Gamm, D. (2013) 'Investigating the role of VSX2 (CHX10) in human retinogenesis using iPS cells', *Investigative Ophthalmology & Visual Science,* 54(15), pp. 4052-4052.

Phillips, M. J., Perez, E. T., Martin, J. M., Reshel, S. T., Wallace, K. A., Capowski, E. E., Singh, R., Wright, L. S., Clark, E. M., Barney, P. M., Stewart, R., Dickerson, S. J., Miller, M. J., Percin, E. F., Thomson, J. A. and Gamm, D. M. (2014) 'Modeling human retinal development with patient-specific iPS cells reveals multiple roles for VSX2', *Stem Cells (Dayton, Ohio),* 32(6), pp. 1480-1492.

Picker, A., Cavodeassi, F., Machate, A., Bernauer, S., Hans, S., Abe, G., Kawakami, K., Wilson, S. W. and Brand, M. (2009) 'Dynamic Coupling of Pattern Formation and Morphogenesis in the Developing Vertebrate Retina', *PLoS Biology,* 7(10), pp. e1000214.

Pierrache, L. H. M., Kimchi, A., Ratnapriya, R., Roberts, L., Astuti, G. D. N., Obolensky, A., Beryozkin, A., Tjon-Fo-Sang, M. J. H., Schuil, J., Klaver, C. C. W., Bongers, E. M. H. F., Haer-Wigman, L., Schalij, N., Breuning, M. H., Fischer, G. M., Banin, E., Ramesar, R. S., Swaroop, A., van den Born, L. I., Sharon, D. and Cremers, F. P. M. (2017) 'Whole-Exome Sequencing Identifies Biallelic IDH3A Variants as a Cause of Retinitis Pigmentosa Accompanied by Pseudocoloboma.', *Ophthalmology,* 124(7), pp. 992-1003.

Pirooznia, M., Goes, F. S. and Zandi, P. P. (2015) 'Whole-genome CNV analysis: advances in computational approaches', *Frontiers in Genetics,* 6, pp. 138.
Plachez, C., Andrews, W., Liapi, A., Knoell, B., Drescher, U., Mankoo, B., Zhe, L., Mambetisaeva, E., Annan, A., Bannister, L., Parnavelas, J. G., Richards, L. J. and Sundaresan, V. (2008) 'Robos are required for the correct targeting of retinal ganglion cell axons in the visual pathway of the brain', *Molecular and Cellular Neuroscience,* 37(4), pp. 719-730.

Plagnol, V., Curtis, J., Epstein, M., Mok, K. Y., Stebbings, E., Grigoriadou, S., Wood, N. W., Hambleton, S., Burns, S. O. and Thrasher, A. J. (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.

Plaza, S., Dozier, C., Langlois, M. C. and Saule, S. (1995) 'Identification and characterization of a neuroretina-specific enhancer element in the quail Pax-6 (Pax-QNR) gene', *Molecular and Cellular Biology,* 15(2), pp. 892-903.

Poulter, James A., Al-Araimi, M., Conte, I., van Genderen, Maria M., Sheridan, E., Carr, Ian M., Parry, David A., Shires, M., Carrella, S., Bradbury, J., Khan, K., Lakeman, P., Sergouniotis, Panagiotis I., Webster, Andrew R., Moore, Anthony T., Pal, B., Mohamed, Moin D., Venkataramana, A., Ramprasad, V., Shetty, R., Saktivel, M., Kumaramanickavel, G., Tan, A., Mackey, David A., Hewitt, Alex W., Banfi, S., Ali, M., Inglehearn, Chris F. and Toomes, C. (2013) 'Recessive Mutations in SLC38A8 Cause Foveal Hypoplasia and Optic Nerve Misrouting without Albinism', *The American Journal of Human Genetics,* 93(6), pp. 1143-1150.

Pras, E., Kristal, D., Shoshany, N., Volodarsky, D., Vulih, I., Celniker, G., Isakov, O., Shomron, N. and Pras, E. (2015) 'Rare genetic variants in Tunisian Jewish patients suffering from age-related macular degeneration', *Journal of Medical Genetics,* 52(7), pp. 484.

Provis, J. M. (2001) 'Development of the primate retinal vasculature', *Progress in Retinal and Eye Research,* 20(6), pp. 799-821.

Provis, J. M., Dubis, A. M., Maddess, T. and Carroll, J. (2013) 'Adaptation of the Central Retina for High Acuity Vision: Cones, the Fovea and the Avascular Zone', *Progress in Retinal and Eye Research,* 35, pp. 63-81.

Provis, J. M., Penfold, P. L., Cornish, E. E., Sandercoe, T. M. and Madigan, M. C. (2010) 'Anatomy and development of the macula: specialisation and the vulnerability to macular degeneration', *Clinical and Experimental Optometry,* 88(5), pp. 269-281.

Provis, J. M., Sandercoe, T. and Hendrickson, A. E. (2000) 'Astrocytes and Blood Vessels Define the Foveal Rim during Primate Retinal Development', *Investigative Ophthalmology & Visual Science,* 41(10), pp. 2827-2836.

Quiring, R., Walldorf, U., Kloter, U. and Gehring, W. J. (1994) 'Homology of the eyeless gene of Drosophila to the Small eye gene in mice and Aniridia in humans', *Science,* 265(5173), pp. 785.

Raine, C. S. (1894) *Morphology of myelin and myelination. Myelin* 2 edn. New York: Plenum, p. 1-41.

Rakoczy, P. E., Sarks, S. H., Daw, N. and Constable, I. J. (1999) 'Distribution of Cathepsin D in Human Eyes with or without Age-related Maculopathy', *Experimental Eye Research,* 69(4), pp. 367-374.

Rattner, A., Sun, H. and Nathans, J. (1999) 'Molecular Genetics of Human Retinal Disease', *Annual Review of Genetics,* 33(1), pp. 89-131.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V. and Korbel, J. O. (2012) 'DELLY: structural variant discovery by integrated paired-end and split-read analysis', *Bioinformatics,* 28(18), pp. i333-i339.

Ravi, V., Bhatia, S., Gautier, P., Loosli, F., Tay, B.-H., Tay, A., Murdoch, E., Coutinho, P., van Heyningen, V., Brenner, S., Venkatesh, B. and Kleinjan, D. A. (2013) 'Sequencing of Pax6 Loci from the Elephant Shark Reveals a Family of Pax6 Genes in Vertebrate Genomes, Forged by Ancient Duplications and Divergences', *PLoS Genetics,* 9(1), pp. e1003177.

Raviv, S., Bharti, K., Rencus-Lazar, S., Cohen-Tayar, Y., Schyr, R., Evantal, N., Meshorer, E., Zilberberg, A., Idelson, M., Reubinoff, B., Grebe, R., Rosin-Arbesfeld, R., Lauderdale, J., Lutty, G., Arnheiter, H. and Ashery-Padan, R. (2014) 'PAX6 regulates melanogenesis in the retinal pigmented epithelium through feed-forward regulatory interactions with MITF', *PLoS Genetics,* 10(5), pp. e1004360-e1004360.

Recchia, F. M., Carvalho-Recchia, C. A. and Trese, M. T. (2002) 'Optical coherence tomography in the diagnosis of foveal hypoplasia', *Archives of Ophthalmology,* 120(11), pp. 1587-1588.

Redon, R., Ishikawa, S., Fitch, K. R., Feuk, L., Perry, G. H., Andrews, T. D., Fiegler, H., Shapero, M. H., Carson, A. R., Chen, W., Cho, E. K., Dallaire, S., Freeman, J. L., Gonzalez, J. R., Gratacos, M., Huang, J., Kalaitzopoulos, D., Komura, D., MacDonald, J. R., Marshall, C. R., Mei, R., Montgomery, L., Nishimura, K., Okamura, K., Shen, F., Somerville, M. J., Tchinda, J., Valsesia, A., Woodwark, C., Yang, F., Zhang, J., Zerjal, T., Zhang, J., Armengol, L., Conrad, D. F., Estivill, X., Tyler-Smith, C., Carter, N. P., Aburatani, H., Lee, C., Jones, K. W., Scherer, S. W. and Hurles, M. E. (2006) 'Global variation in copy number in the human genome', *Nature,* 444(7118), pp. 444-454.

Reinecke, R. D. (1997) 'Idiopathic infantile nystagmus: Diagnosis and treatment.', *Journal of American Association for Pediatric Ophthalmology and Strabismus,* 1(2), pp. 67-82.

RetNet (2018) *http://www.sph.uth.tmc.edu/RetNet/*.

Roadmap Epigenomics, C. and Kundaje, A. and Meuleman, W. and Ernst, J. and Bilenky, M. and Yen, A. and Heravi-Moussavi, A. and Kheradpour, P. and Zhang, Z. and Wang, J. and Ziller, M. J. and Amin, V. and Whitaker, J. W. and Schultz, M. D. and Ward, L. D. and Sarkar, A. and Quon, G. and Sandstrom, R. S. and Eaton, M. L. and Wu, Y.-C. and Pfenning, A. and Wang, X. and ClaussnitzerYaping Liu, M. and Coarfa, C. and Alan Harris, R. and Shoresh, N. and Epstein, C. B. and Gjoneska, E. and Leung, D. and Xie, W. and David Hawkins, R. and Lister, R. and Hong, C. and Gascard, P. and Mungall, A. J. and Moore, R. and Chuah, E. and Tam, A. and Canfield, T. K. and Scott Hansen, R. and Kaul, R. and Sabo, P. J. and Bansal, M. S. and Carles, A. and Dixon, J. R. and Farh, K.-H. and Feizi, S. and Karlic, R. and Kim, A.-R. and Kulkarni, A. and Li, D. and Lowdon, R. and Elliott, G. and Mercer, T. R. and Neph, S. J. and Onuchic, V. and Polak, P. and Rajagopal, N. and Ray, P. and Sallari, R. C. and Siebenthall, K. T. and Sinnott-Armstrong, N. A. and Stevens, M. and Thurman, R. E. and Wu, J. and Zhang, B. and Zhou, X. and Abdennur, N. and Adli, M. and Akerman, M. and Barrera, L. and Antosiewicz-Bourget, J. and Ballinger, T. and Barnes, M. J. and Bates, D. and Bell, R. J. A. and Bennett, D. A. and Bianco, K. and Bock, C. and Boyle, P. and Brinchmann, J. and Caballero-Campo, P. and Camahort, R. and Carrasco-Alfonso, M. J. and Charnecki, T. and Chen, H. and Chen, Z. and Cheng, J. B. and Cho, S. and Chu, A. and Chung, W.-Y. and Cowan, C. and Athena Deng, Q. and Deshpande, V. and Diegel, M. and Ding, B. and Durham, T. and Echipare, L. and Edsall, L. and Flowers, D. and Genbacev-Krtolica, O. and Gifford, C. and Gillespie, S. and Giste, E. and Glass, I. A. and Gnirke, A. and Gormley, M. and Gu, H. and Gu, J. and Hafler, D. A. and Hangauer, M. J. and Hariharan, M. and Hatan, M. and Haugen, E. and He, Y. and Heimfeld, S. and Herlofsen, S. and Hou, Z. and Humbert, R. and Issner, R. and Jackson, A. R. and Jia, H. and Jiang, P. and Johnson, A. K. and Kadlecek, T. and Kamoh, B. and Kapidzic, M. and Kent, J. and Kim, A. and Kleinewietfeld, M. and Klugman, S. and Krishnan, J. and Kuan, S. and Kutyavin, T. and Lee, A.-Y. and Lee, K. and Li, J. and Li, N. and Li, Y. and Ligon, K. L. and Lin, S. and Lin, Y. and Liu, J. and Liu, Y. and Luckey, C. J. and Ma, Y. P. and Maire, C. and Marson, A. and Mattick, J. S. and Mayo, M. and McMaster, M. and Metsky, H. and Mikkelsen, T. and Miller, D. and Miri, M. and Mukame, E. and Nagarajan, R. P. and Neri, F. and Nery, J. and Nguyen, T. and O'Geen, H. and Paithankar, S. and Papayannopoulou, T. and Pelizzola, M. and Plettner, P. and Propson, N. E. and Raghuraman, S. and Raney, B. J. and Raubitschek, A. and Reynolds, A. P. and Richards, H. and Riehle, K. and Rinaudo, P. and Robinson, J. F. and Rockweiler, N. B. and Rosen, E. and Rynes, E. and Schein, J. and Sears, R. and Sejnowski, T. and Shafer, A. and Shen, L. and Shoemaker, R. and Sigaroudinia, M. and Slukvin, I. and Stehling-Sun, S. and Stewart, R. and Subramanian, S. L. and Suknuntha, K. and Swanson, S. and Tian, S. and Tilden, H. and Tsai, L. and Urich, M. and Vaughn, I. and Vierstra, J. and Vong, S. and Wagner, U. and Wang, H. and Wang, T. and Wang, Y. and Weiss, A. and Whitton, H. and Wildberg, A. and Witt, H. and Won, K.-J. and Xie, M. and Xing, X. and Xu, I. and Xuan, Z. and Ye, Z. and Yen, C.-a. and Yu, P. and Zhang, X. and Zhang, X. and Zhao, J. and Zhou, Y. and Zhu, J. and Zhu, Y. and Ziegler, S. and Beaudet, A. E. and Boyer, L. A. and De Jager, P. L. and Farnham, P. J. and Fisher, S. J. and Haussler, D. and Jones, S. J. M. and Li, W. and Marra, M. A. and McManus, M. T. and Sunyaev, S. and Thomson, J. A. and Tlsty, T. D. and Tsai, L.-H. and Wang, W. and Waterland, R. A. and Zhang, M. Q. and Chadwick, L. H. and Bernstein, B. E. and Costello, J. F. and Ecker, J. R. and Hirst, M. and Meissner, A. and Milosavljevic, A. and Ren, B. and Stamatoyannopoulos, J.

A. and Wang, T. and Kellis, M. and Pfenning, A. R. and Claussnitzer, M. and Liu, Y. and Harris, R. A. and Hawkins, R. D. and Hansen, R. S. (2015) 'Integrative analysis of 111 reference human epigenomes', *Nature,* 518, pp. 317.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G. and Mesirov, J. P. (2011) 'Integrative genomics viewer', *Nature Biotechnology,* 29(1), pp. 24-26.

Rooryck, C., Morice-Picard, F., Elcioglu, N. H., Lacombe, D., Taieb, A. and Arveiller, B. (2008) 'Molecular diagnosis of oculocutaneous albinism: new mutations in the OCA1-4 genes and practical aspects.', *Pigment Cell & Melanoma Research,* 21(5), pp. 583-587.

Rowan, S. and Cepko, C. L. (2004) 'Genetic analysis of the homeodomain transcription factor Chx10 in the retina using a novel multifunctional BAC transgenic mouse reporter', *Developmental Biology,* 271(2), pp. 388-402.

Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: a unified platform for automated protein structure and function prediction .', *Nature Protocols,* 5(725-738).

Rundshagen, U., Zuhlke, C., Schwinger, E. and Kasmann-Kellner, B. (2004) 'Mutations inf the MATP gene in five German patients affected by oculocutaneous albinism type 4.', *Human Mutation,* 23(2), pp. 106-110.

Rushton, W. A. (1965) 'The rod dark adaptation curve measured above cone threshold', *The Journal of Physiology,* 181(3), pp. 641-644.

Russell, S., Bennett, J., Wellman, J. A., Chung, D. C., Yu, Z.-F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., Cross, D., Marshall, K. A., Walshire, J., Kehoe, T. L., Reichert, H., Davis, M., Raffini, L., George, L. A., Hudson, F. P., Dingfield, L., Zhu, X., Haller, J. A., Sohn, E. H., Mahajan, V. B., Pfeifer, W., Weckmann, M., Johnson, C., Gewaily, D., Drack, A., Stone, E., Wachtel, K., Simonelli, F., Leroy, B. P., Wright, J. F., High, K. A. and Maguire, A. M. (2017) 'Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial', *Lancet,* 390(10097), pp. 849-860.

Saba, L., Meloni, A., Sardu, R., Travi, M., Primignani, P., Rosatelli, M. C. and Cao, A. (1992) 'A novel β-thalassemia mutation (G→A) at the initiation codon of the β-globin gene', *Human Mutation,* 1(5), pp. 420-422.

Sachidanandam, R., Weissman , D., Schmidt , S. C., Kakol, J. M., Stein, L. D., Marth , G., Sherry, S., Mullikin , J. C., Mortimore , B. J., Willey , D. L., Hunt , S. E., Cole , C. G., Coggill , P. C., Rice , C. M., Ning , Z., Rogers, J., Bentley, D. R., Kwok , P.-Y., Mardis , E. R., Yeh , R. T., Schultz , B., Cook , L., Davenport , R., Dante , M., Fulton , L., Hillier , L., Waterston, R. H., McPherson, J. D., Gilman , B., Schaffner , S., Van Etten , W. J., Reich , D., Higgins , J., Daly , M. J., Blumenstiel , B., Baldwin , J., Stange-Thomann , N., Zody , M. C., Linton , L., Lander, E. S. and Altshuler, D. (2001) 'A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms', *Nature,* 409(6822), pp. 928-933.

Saffra, N., Agarwal, S., Chiang, J., Masini, R. and Bertolucci, A. (2012) 'Spectral-domain optical coherence tomographic characteristics of autosomal recessive isolated foveal hypoplasia', *Archives of Ophthalmology,* 130(10), pp. 1324-1327.

Sandercoe, T. M., Geller, S. F., Hendrickson, A. E., Stone, J. and Provis, J. M. (2003) 'VEGF expression by ganglion cells in central retina before formation of the foveal depression in monkey retina: evidence of developmental hypoxia.', *The Journal of Comparative Neurology,* 462(1), pp. 42-54.

Sangermano, R., Khan, M., Cornelis, S. S., Richelle, V., Albert, S., Elmelik, D., Garanto, A., Qamar, R., Lugtenberg, D., van den Born, L. I., Collin, R. W. J. and Cremers, F. P. M. (2017) 'ABCA4 midigenes reveal the full splice spectrum of all reported non-canonical splice site variants in Stargardt disease', *Genome Research*.

Sarangarajan, R. and Boissy, R. E. (2001) 'Tyrp1 and Oculocutaneous Albinism Type 3', *Pigment Cell Research,* 14(6), pp. 437-444.

Schaefer, B. C. (1995) 'Revolutions in Rapid Amplification of cDNA Ends: New Strategies for Polymerase Chain Reaction Cloning of Full-Length cDNA Ends', *Analytical Biochemistry,* 227(2), pp. 255-273.

Scheper, G. C., van der Knaap, M. S. and Proud, C. G. (2007) 'Translation matters: protein synthesis defects in inherited disease', *Nature Reviews Genetics,* 8(9), pp. 711-723.

Schiöth, H. B., Roshanbin, S., Hägglund, M. G. A. and Fredriksson, R. (2013) 'Evolutionary origin of amino acid transporter families SLC32, SLC36 and SLC38 and physiological, pathological and therapeutic aspects', *Molecular Aspects of Medicine,* 34(2), pp. 571-585.

Schroeder, H. W., Orth, U., Eberhard, E.-K. and Gal, A. (2003) 'Familiäre Foveahypoplasie - Klinische Einordnung', *Klin Monatsbl Augenheilkd,* 220(8), pp. 559-562.

Schultz, D. W., Klein, M. L., Humpert, A. J., Luzier, C. W., Persun, V., Schain, M., Mahan, A., Runckel, C., Cassera, M., Vittal, V., Doyle, T. M., Martin, T. M., Weleber, R. G., Francis, P. J. and Acott, T. S. (2003) 'Analysis of the ARMD1 locus: evidence that a mutation in HEMICENTIN-1 is associated with age-related macular degeneration in a large family', *Human Molecular Genetics,* 12(24), pp. 3315-3323.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998) 'SMART, a simple modular architecture research tool: Identification of signaling domains', *Proceedings of the National Academy of Sciences,* 95(11), pp. 5857-5864.

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

Schwarz, J. M., Rodelsperger, C., Schuelke, M. and Seelow, D. (2010) 'MutationTaster evaluates disease-causing potential of sequence alterations', *Nature Methods,* 7(8), pp. 575-576.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) 'The FoldX web server: an online force field', *Nucleic Acids Research,* 33(Web Server issue), pp. W382-W388.

Sergouniotis, P. (2012) *Genetic and phenotypic heterogeneity in autosomal recessive retinal disease.* PhD, University College London, London.

Sergouniotis, Panagiotis I., Chakarova, C., Murphy, C., Becker, M., Lenassi, E., Arno, G., Lek, M., MacArthur, Daniel G., Bhattacharya, Shomi S., Moore, Anthony T., Holder, Graham E., Robson, Anthony G., Wolfrum, U., Webster, Andrew R. and Plagnol, V. (2014) 'Biallelic Variants in TTLL5, Encoding a Tubulin Glutamylase, Cause Retinal Dystrophy.', *The American Journal of Human Genetics,* 94(5), pp. 760-769.

Sharp, A. J., Mefford, H. C., Li, K., Baker, C., Skinner, C., Stevenson, R. E., Schroer, R. J., Novara, F., De Gregori, M., Ciccone, R., Broomer, A., Casuga, I., Wang, Y., Xiao, C., Barbacioru, C., Gimelli, G., Bernardina, B. D., Torniero, C., Giorda, R., Regan, R., Murday, V., Mansour, S., Fichera, M., Castiglia, L., Failla, P., Ventura, M., Jiang, Z., Cooper, G. M., Knight, S. J. and Romano, C. (2008) 'A

recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures', *Nature Genetics,* 40.

Shaw, C. J. and Lupski, J. R. (2004) 'Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease', *Human Molecular Genetics,* 13(suppl_1), pp. R57-R64.

Shaw, M. M., Falls, H. F. and Neel, J. V. (1960) 'Congenital aniridia.', *The American Journal of Human Genetics,* 12, pp. 389.

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M. and Sirotkin, K. (2001) 'dbSNP: the NCBI database of genetic variation', *Nucleic Acids Research,* 29(1), pp. 308-311.

Shi, Y. and Majewski, J. (2013) 'FishingCNV: a graphical software package for detecting rare copy number variations in exome-sequencing data', *Bioinformatics,* 29(11), pp. 1461-1462.

Shiels, A., Bennett, T. M., Knopf, H. L. S., Maraini, G., Li, A., Jiao, X. and Hejtmancik, J. F. (2008) 'The EPHA2 gene is associated with cataracts linked to chromosome 1p', *Molecular Vision,* 14, pp. 2042-2055.

Shigemizu, D., Momozawa, Y., Abe, T., Morizono, T., Boroevich, K. A., Takata, S., Ashikawa, K., Kubo, M. and Tsunoda, T. (2015) 'Performance comparison of four commercial human whole-exome capture platforms', *Scientific Reports,* 5, pp. 12742.

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., Day, I. N. M. and Gaunt, T. R. (2013) 'Predicting the Functional, Molecular, and Phenotypic Consequences of Amino Acid Substitutions using Hidden Markov Models', *Human Mutation,* 34(1), pp. 57-65.

Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., Teramura, H., Yamamoto, T., Komatsu, H., Miura, K., Ezura, H., Nishida, K., Ariizumi, T. and Kondo, A. (2017) 'Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion', *Nature Biotechnology,* 35, pp. 441.

Shin, J., Ming, G.-l. and Song, H. (2014) 'Decoding neural transcriptomes and epigenomes via high-throughput sequencing', *Nature Neuroscience,* 17, pp. 1463.

Shinawi, M., Schaaf, C. P., Bhatt, S. S., Xia, Z., Patel, A., Cheung, S. W., Lanpher, B., Nagl, S., Herding, H. S., Nevinny-Stickel, C., Immken, L. L., Patel, G. S., German, J. R., Beaudet, A. L. and Stankiewicz, P. (2009) 'A small recurrent deletion within 15q13.3 is associated with a range of neurodevelopmental phenotypes', *Nature Genetics,* 41(12), pp. 1269-1271.

Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., Giorda, R. and Pozzoli, U. (2004) 'Silencer elements as possible inhibitors of pseudoexon splicing', *Nucleic Acids Research,* 32(5), pp. 1783-1791.

Sistermans, E., de Wijs, I., de Coo, R., Smit, L., Menko, F. and van Oost, B. (1996) 'A (G-to-A) mutation in the initiation codon of the proteolipid protein gene causing a relatively mild form of Pelizaeus-Merzbacher disease in a Dutch family. ', *Human Mutation,* 97(3), pp. 337-339.

Small, K., Iber, J. and Warren, S. T. (1997) 'Emerin deletion reveals a common X chromosome inversion mediated by inverted repeats', *Nature Genetics,* 16.

Sohocki, M. M., Daiger, S. P., Bowne, S. J., Rodriquez, J. A., Northrup, H., Heckenlively, J. R., Birch, D. G., Mintz-Hittner, H., Ruiz, R. S., Lewis, R. A., Saperstein, D. A. and Sullivan, L. S. (2001) 'Prevalence of Mutations Causing Retinitis Pigmentosa and Other Inherited Retinopathies', *Human Mutation,* 17(1), pp. 42-51.

Spickett, C., Hysi, P., Hammond, C. J., Prescott, A., Fincham, G. S., Poulson, A. V., McNinch, A. M., Richards, A. J. and Snead, M. P. (2016) 'Deep intronic sequence variants in COL2A1 affect the alternative splicing efficiency of exon 2, and may confer a risk for Rhegmatogenous retinal detachment.', *Human Mutation,* 10, pp. 1085-1096.

Springer, A. D. (1999) 'New role for the primate fovea: a retinal excavation determines photoreceptor deployment and shape.', *Visual Neuroscience,* 16(4), pp. 629-636.

Springer, A. D. and Hendrickson, A. E. (2005) 'Development of the primate area of high acuity, 3: Temporal relationships between pit formation, retinal elongation and cone packing', *Visual Neuroscience,* 22(2), pp. 171-185.

Srivathsan, A., Baloğlu, B., Wang, W., Tan, W. X., Bertrand, D., Ng, A. H. Q., Boey, E. J. H., Koh, J. J. Y., Nagarajan, N. and Meier, R. (2018) 'A MinION-based pipeline for fast and cost-effective DNA barcoding', *bioRxiv*.

Stankiewicz, P. and Lupski, J. R. (2002) 'Genome architecture, rearrangements and genomic disorders', *Trends in Genetics,* 18(2), pp. 74-82.

Starcevic, M. and Dell'Angelica, E. C. (2004) 'Identification of Snapin and Three Novel Proteins (BLOS1, BLOS2, and BLOS3/Reduced Pigmentation) as Subunits of Biogenesis of Lysosome-related Organelles Complex-1 (BLOC-1)', *Journal of Biological Chemistry,* 279(27), pp. 28393-28401.

Stargardt, K. (1913) 'I. Über familiäre, progressive Degeneration in der Maculagegend des Auges', *Ophthalmologica,* 30(2-3), pp. 95-116.

Startek, M., Szafranski, P., Gambin, T., Campbell, I. M., Hixson, P., Shaw, C. A., Stankiewicz, P. and Gambin, A. (2015) 'Genome-wide analyses of LINE–LINE-mediated nonallelic homologous recombination', *Nucleic Acids Research,* 43(4), pp. 2188-2198.

Stavropoulos, D. J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R. K. C., Szego, M. J., Hayeems, R. Z., Shaul, R. Z., Brudno, M., Girdea, M., Frey, B., Alipanahi, B., Ahmed, S., Babul-Hirji, R., Porras, R. B., Carter, M. T., Chad, L., Chaudhry, A., Chitayat, D., Doust, S. J., Cytrynbaum, C., Dupuis, L., Ejaz, R., Fishman, L., Guerin, A., Hashemi, B., Helal, M., Hewson, S., Inbar-Feigenberg, M., Kannu, P., Karp, N., Kim, R. H., Kronick, J., Liston, E., MacDonald, H., Mercimek-Mahmutoglu, S., Mendoza-Londono, R., Nasr, E., Nimmo, G., Parkinson, N., Quercia, N., Raiman, J., Roifman, M., Schulze, A., Shugar, A., Shuman, C., Sinajon, P., Siriwardena, K., Weksberg, R., Yoon, G., Carew, C., Erickson, R., Leach, R. A., Klein, R., Ray, P. N., Meyn, M. S., Scherer, S. W., Cohn, R. D. and Marshall, C. R. (2016) 'Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine', *Npj Genomic Medicine,* 1, pp. 15012.

Steele-Stallard, H. B., Le Quesne Stabej, P., Lenassi, E., Luxon, L. M., Claustres, M., Roux, A.-F., Webster, A. R. and Bitner-Glindzicz, M. (2013) 'Screening for duplications, deletions and a common intronic mutation detects 35% of second mutations in patients with USH2A monoallelic mutations on Sanger sequencing', *Orphanet Journal of Rare Diseases,* 8, pp. 122-122.

Stenson, P. D., Mort, M., Ball, E. V., Shaw, K., Phillips, A. D. and Cooper, D. N. (2014) 'The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine', *Human Genetics,* 133(1), pp. 1-9.

Stinchcombe, J., Bossi, G. and Griffiths, G. M. (2004) 'Linking Albinism and Immunity: The Secrets of Secretory Lysosomes', *Science,* 305(5680), pp. 55.

Stone, J., Sandercoe, T. M. and Provis, J. M. (2005) 'Mechanisms of the formation and stability of retinal blood vessels.', in Tombran-Tink, J. and Barnstable, C. (eds.) *Ocular Angiogenesis: Diseases, Mechanisms and Therapeutics*. Totowa, NJ: Humana Press.

Strauss, O. (2005) 'The retinal pigment epithelium in visual function', *Physiological Reviews,* 85, pp. 845-881.

Summers, C. G., Connett, J. E., Holleschau, A. M., Anderson, J. L., De Becker, I., McKay, B. S. and Brilliant, M. H. (2014) 'Does levodopa improve vision in albinism? Results of a randomized, controlled clinical trial', *Clinical & Experimental Ophthalmology,* 42(8), pp. 713-721.

Sundberg, B. E., Wååg, E., Jacobsson, J. A., Stephansson, O., Rumaks, J., Svirskis, S., Alsiö, J., Roman, E., Ebendal, T., Klusa, V. and Fredriksson, R. (2008) 'The Evolutionary History and Tissue Mapping of Amino Acid Transporters Belonging to Solute Carrier Families SLC32, SLC36, and SLC38', *Journal of Molecular Neuroscience,* 35(2), pp. 179-193.

Sung, C.-H. and Chuang, J.-Z. (2010) 'The cell biology of vision', *The Journal of Cell Biology,* 190(6), pp. 953-963.

Suzuki, T., Li, W., Zhang, Q., Karim, A., Novak, E. K., Sviderskaya, E. V., Hill, S. P., Bennett, D. C., Levin, A. V., Nieuwenhuis, H. K., Fong, C.-T., Castellan, C., Miterski, B., Swank, R. T. and Spritz, R. A. (2002) 'Hermansky-Pudlak syndrome is caused by mutations in HPS4, the human homolog of the mouse light-ear gene', *Nature Genetics,* 30, pp. 321.

Takahashi, H., Sakuta, H., Shintani, T. and Noda, M. (2009) 'Functional mode of FoxD1/CBF2 for the establishment of temporal retinal specificity in the developing chick retina', *Developmental Biology,* 331(2), pp. 300-310.

Takahashi, H., Shintani, T., Sakuta, H. and Noda, M. (2003) 'CBF1 controls the retinotectal topographical map along the anteroposterior axis through multiple mechanisms', *Development,* 130(21), pp. 5203.

Talbot, S. A. and Marshall, W. H. (1941) 'Physiological Studies on Neural Mechanisms of Visual Localization and Discrimination', *American Journal of Ophthalmology,* 24(11), pp. 1255-1264.

Talkowski, M. E., Rosenfeld, J. A., Blumenthal, I., Pillalamarri, V., Chiang, C., Heilbut, A., Ernst, C., Hanscom, C., Rossin, E., Lindgren, A., Pereira, S., Ruderfer, D., Kirby, A., Ripke, S., Harris, D., Lee, J.-H., Ha, K., Kim, H.-G., Solomon, B. D., Gropman, A. L., Lucente, D., Sims, K., Ohsumi, T. K., Borowsky, M. L., Loranger, S., Quade, B., Lage, K., Miles, J., Wu, B.-L., Shen, Y., Neale, B., Shaffer, L. G., Daly, M. J., Morton, C. C. and Gusella, J. F. (2012) 'Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries', *Cell,* 149(3), pp. 525-537.

Tan, L. X., Toops, K. A. and Lakkaraju, A. (2016) 'Protective responses to sublytic complement in the retinal pigment epithelium', *Proceedings of the National Academy of Sciences,* 113(31), pp. 8789-8794.

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, H. K., Chao, L.-Y. and Saunders, G. F. (1997) 'Functional analysis of paired box missense mutations in the PAX6 gene.', *Human Molecular Genetics,* 6(3), pp. 381-386.

Tarpey, P., Thomas, S., Sarvananthan, N., Mallya, U., Lisgo, S., Talbot, C. J., Roberts, E. O., Awan, M., Surendran, M., McLean, R. J., Reinecke, R. D.,

Langmann, A., Lindner, S., Koch, M., Woodruff, G., Gale, R., Degg, C., Droutsas, K., Asproudis, I., Zubcov, A. A., Pieh, C., Veal, C. D., Machado, R. D., Backhouse, O. C., Baumber, L., Jain, S., Constantinescu, C. S., Brodsky, M. C., Hunter, D. G., Hertle, R. W., Read, R. J., Edkins, S., O'Meara, S., Parker, A., Stevens, C., Teague, J., Wooster, R., Futreal, P. A., Trembath, R. C., Stratton, M. R., Raymond, F. L. and Gottlob, I. (2006) 'Mutations in a novel member of the FERM family, FRMD7 cause X-linked idiopathic congenital nystagmus (NYS1)', *Nature Genetics,* 38(11), pp. 1242-1244.

Tassabehji, M., Newton, V. E. and Read, A. P. (1994) 'Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (*MITF*) gene.', *Nature Genetics,* (8), pp. 251-255.

Tattini, L., D'Aurizio, R. and Magi, A. (2015) 'Detection of Genomic Structural Variants from Next-Generation Sequencing Data', *Frontiers in Bioengineering and Biotechnology,* 3, pp. 92.

Thomas, M. G., Crosier, M., Lindsay, S., Kumar, A., Araki, M., Leroy, B. P., McLean, R. J., Sheth, V., Maconachie, G., Thomas, S., Moore, A. T. and Gottlob, I. (2014) 'Abnormal retinal development associated with FRMD7 mutations', *Human Molecular Genetics,* 23(15), pp. 4086-4093.

Thomas, M. G., Kumar, A., Mohammad, S., Proudlock, F. A., Engle, E. C., Andrews, C., Chan, W.-M., Thomas, S. and Gottlob, I. (2011) 'Structural Grading of Foveal Hypoplasia Using Spectral-Domain Optical Coherence Tomography: A Predictor of Visual Acuity?', *Ophthalmology,* 118(8), pp. 1653-1660.

Thomas, S., Proudlock, F. A., Sarvananthan, N., Roberts, E. O., Awan, M., McLean, R., Surendran, M., Anil Kumar, A. S., Farooq, S. J., Degg, C., Gale, R. P., Reinecke, R. D., Woodruff, G., Langmann, A., Lindner, S., Jain, S., Tarpey, P., Raymond, F. L. and Gottlob, I. (2008) 'Phenotypical characteristics of idiopathic infantile nystagmus with and without mutations in FRMD7', *Brain,* 131(5), pp. 1259-1267.

Thompson, D. A., Li, Y., McHenry, C. L., Carlson, T. J., Ding, X., Sieving, P. A., Apfelstedt-Sylla, E. and Gal, A. (2001) 'Mutations in the gene encoding lecithin retinol acyltransferase are associated with early-onset severe retinal dystrophy', *Nature Genetics,* 28, pp. 123.

Thompson, H., Andrews, W., Parnavelas, J. G. and Erskine, L. (2009) 'Robo2 is required for Slit-mediated intraretinal axon guidance', *Developmental Biology,* 335(2), pp. 418-426.

Thompson, H., Barker, D., Camand, O. and Erskine, L. (2006) 'Slits contribute to the guidance of retinal ganglion cell axons in the mammalian optic tract', *Developmental Biology,* 296(2), pp. 476-484.

Thorvaldsdóttir, H., Robinson, J. T. and Mesirov, J. P. (2013) 'Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration', *Briefings in Bioinformatics,* 14(2), pp. 178-192.

Ton, C. C. T., Hirvonen, H., Miwa, H., Weil, M. M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N. D., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L. C. and Saunders, G. F. (1991) 'Positional cloning and characterization of a paired box- and homeobox-containing gene from the aniridia region', *Cell,* 67(6), pp. 1059-1074.

Toral, M. A., Velez, G., Boudreault, K., Schaefer, K. A., Xu, Y., Saffra, N., Bassuk, A. G., Tsang, S. H. and Mahajan, V. B. (2017) 'Structural modeling of a novel SLC38A8 mutation that causes foveal hypoplasia.', *Molecular Genetics & Genomic Medicine,* 5(3), pp. 202-209.

Triplett, J. W. and Feldheim, D. A. (2012) 'Eph and ephrin signaling in the formation of topographic maps', *Seminars in cell & developmental biology,* 23(1), pp. 7-15.

Tsai, Y.-C., Greenberg, D., Powell, J., Hoijer, I., Ameur, A., Strahl, M., Ellis, E., Jonasson, I., Mouro Pinto, R., Wheeler, V., Smith, M. L., Gyllensten, U., Sebra, R., Korlach, J. and Clark, T. A. (2017) 'Amplification-free, CRISPR-Cas9 Targeted Enrichment and SMRT Sequencing of Repeat-Expansion Disease Causative Genomic Regions', *bioRxiv*.

Turner, D. L. and Cepko, C. L. (1987) 'A common progenitor for neurons and glia persists in rat retina late in development', *Nature,* 328, pp. 131.

Tzoulaki, I., White, I. M. S. and Hanson, I. M. (2005) 'PAX6 mutations: genotype-phenotype correlations', *BMC Genetics,* 6, pp. 27-27.

Usher, C. H. (1909) *On the inheritance of Retinitis Pigmentosa: With notes of Cases.* Sherratt & Hughes.

Vache, C., Besnard, T., le Berre, P., Garcia-Garcia, G., Baux, D., Larrieu, L., Abadie, C., Blanchet, C., Bolz, H. J., Millan, J., Hamel, C., Malcolm, S., Claustres, M. and Roux, A. F. (2012) 'Usher syndrome type 2 caused by activation of an USH2A pseudoexon: implications for diagnosis and therapy.', *Human Mutation,* 1, pp. 104-108.

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., Dekker, J. and Lander, E. S. (2010) 'Hi-C: a method to study the three-dimensional architecture of genomes', *Journal of Visualized Experiments,* (39), pp. 1869.

Van Cauwenbergh, C., Van Schil, K., Cannoodt, R., Bauwens, M., Van Laethem, T., De Jaegere, S., Steyaert, W., Sante, T., Menten, B., Leroy, B. P., Coppieters, F. and De Baere, E. (2016) 'arrEYE: a customized platform for high-resolution copy number analysis of coding and noncoding regions of known and candidate retinal dystrophy genes and retinal noncoding RNAs', *Genetics In Medicine,* 19, pp. 457.

Van de Sompele, S., Smith, C., Karali, M., Corton, M., Van Schil, K., Peelman, F., Cherry, T., Rosseel, T., Verdin, H., Derolez, J., Van Laethem, T., Khan, K. N., McKibbin, M., Toomes, C., Ali, M., Torella, A., Testa, F., Jimenez, B., Simonelli, F., De Zaeytijd, J., Van den Ende, J., Leroy, B. P., Coppieters, F., Ayuso, C., Inglehearn, C. F., Banfi, S. and De Baere, E. (2018) 'Biallelic sequence and structural variants in RAX2 are a novel cause for autosomal recessive inherited retinal disease', *Genetics in Medicine*.

van der Velden, A. W. and Thomas, A. A. M. (1999) 'The role of the 5′ untranslated region of an mRNA in translation regulation during development', *The International Journal of Biochemistry & Cell Biology,* 31(1), pp. 87-106.

van Genderen, M. M., Riemslag, F. C. C., Schuil, J., Hoeben, F. P., Stilma, J. S. and Meire, F. M. (2006) 'Chiasmal misrouting and foveal hypoplasia without albinism', *British Journal of Ophthalmology,* 90(9), pp. 1098.

van Heyningen, V. and Williamson, K. A. (2002) 'PAX6 in Sensory Development', *Human Molecular Genetics,* 11(10), pp. 1161-1167.

Van Schil, K., Klevering, B. J., Leroy, B. P., Pott, J. W. R., Bandah-Rozenfeld, D., Zonneveld-Vrieling, M. N., Sharon, D., den Hollander, A. I., Cremers, F. P. M., De Baere, E., Collin, R. W. J. and van den Born, L. I. (2015) 'A Nonsense Mutation in FAM161A Is a Recurrent Founder Allele in Dutch and Belgian Individuals With Autosomal Recessive Retinitis Pigmentosa', *Investigative Ophthalmology & Visual Science,* 56(12), pp. 7418-7426.

Van Schil, K., Naessens, S., Van de Sompele, S., Carron, M., Aslanidis, A., Van Cauwenbergh, C., Kathrin Mayer, A., Van Heetvelde, M., Bauwens, M., Verdin, H., Coppieters, F., Greenberg, M. E., Yang, M. G., Karlstetter, M., Langmann, T., De Preter, K., Kohl, S., Cherry, T. J., Leroy, B. P., Group, C. N. V. S. and De Baere, E. (2017) 'Mapping the genomic landscape of inherited retinal disease genes prioritizes genes prone to coding and noncoding copy-number variations', *Genetics In Medicine*.

Vasilyeva, T., Voskresenskaya, A., Kasmann-Kellner, B., Khlebnikova, O., Pozdeyeva, N., Bayazutdinova, G., Kutsev, S., EK, G., Semina, E., Marakhonov, A. and Zinchenko, R. (2017) 'Molecular analysis of patients with aneroid in Russian Federation broadens the spectrum of PAX6 mutations.', *Clinical Genetics*.

Vecino, E., Rodriguez, F. D., Ruzafa, N., Pereiro, X. and Sharma, S. C. (2016) 'Glia–neuron interactions in the mammalian retina', *Progress in Retinal and Eye Research,* 51, pp. 1-40.

Vetrini, F., Auricchio, A., Du, J., Angeletti, B., Fisher, D. E., Ballabio, A. and Marigo, V. (2004) 'The microphthalmia transcription factor (Mitf) controls expression of the ocular albinism type 1 gene: link between melanin synthesis and melanosome biogenesis', *Molecular and Cellular Biology,* 24(15), pp. 6550-6559.

Vincent, M. C., Gallai, R., Olivier, D., Speeg-Schatz, C., Flament, J., Calvas, P. and Dollfus, H. (2004) 'Variable phenotype related to a novel PAX6 mutation (IVS4+5G>C) in a family presenting congenital nystagmus and foveal hypoplasia.', *The American Journal of Ophthalmology,* 138(6), pp. 1016-1021.

Voaden, M. J., Lake, N., Marshall, J. and Morjaria, B. (1978) 'The utilization of glutamine by the retina: an autoradiographic and metabolic study', *Journal of Neurochemistry,* 31(4), pp. 1069-1076.

Von Dem Hagen, E. A. H., Houston, G. C., Hoffman, M. B. and Morland, A. B. (2007) 'Pigmentation predicts the shift in the line of decussation in humans with albinism.', *European Journal of Neuroscience,* (25), pp. 503-511.

Vulliemoz, S., Raineteau, O. and Jabaudon, D. (2005) 'Reaching beyond the midline: why are human brains cross wired?', *The Lancet Neurology,* 4(2), pp. 87-99.

Wallis, D. E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E. H., Rommens, J. and Muenke, M. (1999) 'Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly.', *Nature Genetics,* 22(2), pp. 196-198.

Walther, C. and Gruss, P. (1991) 'Pax-6, a murine paired box gene, is expressed in the developing CNS.', *Development,* 113(4), pp. 1435-1449.

Wang, A. L., Lukas, T. J., Yuan, M., Du, N., Tso, M. O. and Neufeld, A. H. (2009) 'Autophagy and exosomes in the aged retinal pigment epithelium: possible relevance to drusen formation and age-related macular degeneration', *PLoS One,* 4(1), pp. e4160-e4160.

Wang, G.-S. and Cooper, T. A. (2007) 'Splicing in disease: disruption of the splicing code and the decoding machinery.', *Nature Reviews Genetics,* 8, pp. 749-761.

Wang, Z., Rolish, M. E., Yeo, G., Tung, V., Mawson, M. and Burge, C. B. (2004) 'Systematic Identification and Analysis of Exonic Splicing Silencers', *Cell,* 119(6), pp. 831-845.

Wavre-Shapton, S. T., Tolmachova, T., Lopes da Silva, M., Futter, C. E. and Seabra, M. C. (2013) 'Correction: Conditional Ablation of the Choroideremia Gene Causes Age-Related Changes in Mouse Retinal Pigment Epithelium', *PLoS One,* 8(5).

Wei, A.-H., Zang, D.-J., Zhang, Z., Liu, X.-Z., He, X., Yang, L., Wang, Y., Zhou, Z.-Y., Zhang, M.-R., Dai, L.-L., Yang, X.-M. and Li, W. (2013) 'Exome Sequencing Identifies SLC24A5 as a Candidate Gene for Nonsyndromic Oculocutaneous Albinism.', *Journal of Investigative Dermatology,* 133(7), pp. 1834-1840.

Wei, A. H. and Li, W. (2013) 'Hermansky-Pudlak syndrome: pigmentary and non-pigmentary defects and their pathogenesis.', *Pigment Cell and Melanoma Research,* 26(2), pp. 176-192.

Weisschuh, N., Mayer, A. K., Strom, T. M., Kohl, S., Glöckle, N., Schubach, M., Andreasson, S., Bernd, A., Birch, D. G., Hamel, C. P., Heckenlively, J. R., Jacobson, S. G., Kamme, C., Kellner, U., Kunstmann, E., Maffei, P., Reiff, C. M., Rohrschneider, K., Rosenberg, T., Rudolph, G., Vámos, R., Varsányi, B., Weleber, R. G. and Wissinger, B. (2016) 'Mutation Detection in Patients with Retinal Dystrophies Using Targeted Next Generation Sequencing', *PLoS One,* 11(1), pp. e0145951.

Wen, Y., Liu, Y., Xu, Y., Zhao, Y., Hua, R., Wang, K., Sun, M., Li, Y., Yang, S., Zhang, X.-J., Kruse, R., Cichon, S., Betz, R. C., Nöthen, M. M., van Steensel, M. A. M., van Geel, M., Steijlen, P. M., Hohl, D., Huber, M., Dunnill, G. S., Kennedy, C., Messenger, A., Munro, C. S., Terrinoni, A., Hovnanian, A., Bodemer, C., de Prost, Y., Paller, A. S., Irvine, A. D., Sinclair, R., Green, J., Shang, D., Liu, Q., Luo, Y., Jiang, L., Chen, H.-D., Lo, W. H. Y., McLean, W. H. I., He, C.-D. and Zhang, X. (2009) 'Loss-of-function mutations of an inhibitory upstream ORF in the human hairless transcript cause Marie Unna hereditary hypotrichosis', *Nature Genetics,* 41, pp. 228.

Westbroek, W., Klar, A., Cullinane, A. R., Ziegler, S. G., Hurvitz, H., Ganem, A., Wilson, K., Dorward, H., Huizing, M., Tamimi, H., Vainshtein, I., Berkun, Y., Lavie, M., Gahl, W. A. and Anikster, Y. (2012) 'Cellular and clinical report of new Griscelli syndrome type III cases', *Pigment cell & melanoma research,* 25(1), pp. 47-56.

Wethmar, K., Smink, J. J. and Leutz, A. (2010) 'Upstream open reading frames: Molecular switches in (patho)physiology', *Bioessays,* 32(10), pp. 885-893.

Whiffin, N., Minikel, E., Walsh, R., O'Donnell-Luria, A. H., Karczewski, K., Ing, A. Y., Barton, P. J. R., Funke, B., Cook, S. A., MacArthur, D. and Ware, J. S. (2017) 'Using high-resolution variant frequencies to empower clinical genome interpretation', *Genetics in Medicine,* 19(10), pp. 1151-1158.

Widlund, H. R. and Fisher, D. E. (2003) 'Microphthalamia-associated transcription factor: a critical regulator of pigment cell development and survival', *Oncogene,* 22, pp. 3035.

Wiggs, J. L. and Pierce, E. A. (2013) 'Genetic Testing for Inherited Eye Disease. Who Benefits? ', *JAMA Ophthalmology,* 131(10).

Williams, G. P., Pathak-Ray, V. and Austin, M. W. (2007) 'The social impact of visual impairment', *The British Journal of Ophthalmology,* 91(7), pp. 986-986.

Williams, S. C., Altmann, C. R., Chow, R. L., Hemmati-Brivanlou, A. and Lang, R. A. (1998) 'A highly conserved lens transcriptional control element from the Pax-6 gene', *Mechanisms of Development,* 73(2), pp. 225-229.

Williams, S. E., Mann, F., Erskine, L., Sakurai, T., Wei, S., Rossi, D. J., Gale, N. W., Holt, C. E., Mason, C. A. and Henkemeyer, M. (2003) 'Ephrin-B2 and EphB1 Mediate Retinal Axon Divergence at the Optic Chiasm', *Neuron,* 39(6), pp. 919-935.

Wolf, A., Caliebe, A., Thomas, N. S. T., Ball, E. V., Mort, M., Stenson, P. D., Krawczak, M. and Cooper, D. N. (2011) 'Single base-pair substitutions at the

translation initiation sites of human genes as a cause of inherited disease.', *Human Mutation,* 32(10), pp. 1137-1143.

Wu, N., Doorenbos, M. and Dong Feng, C. (2016) 'Induced Pluripotent Stem Cells: Development in the Ophthalmologic Field.', *Stem Cells International,* 2016, pp. 7 pages.

Wutz, K., Sauer, C., Zrenner, E., Lorenz, B., Alitalo, T., Broghammer, M., Hergersberg, M., Chapelle, A. d. L., Weber, B. H. F., Wissinger, B., Meindl, A. and Pusch, C. M. (2002) 'Thirty distinct CACNA1F mutations in 33 families with incomplete type of XLCSNB and Cacna1f expression profiling in mouse retina', *The European Journal Of Human Genetics,* 10, pp. 449.

Xiao, M. and Hendrickson, A. (2000) 'Spatial and temporal expression of short, long/medium, or both opsin in human fetal cones.', *The Journal of Comparative Neurology,* 425(4), pp. 545-559.

Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L. and Pabo, C. O. (1999) 'Crystal structure of the human Pax6 paired domain–DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding', *Genes & Development,* 13(10), pp. 1263-1275.

Xu, W., Rould, M. A., Jun, S., Desplan, C. and Pabo, C. O. (1995) 'Crystal structure of a paired domain-DNA complex at 2.5 å resolution reveals structural basis for pax developmental mutations', *Cell,* 80(4), pp. 639-650.

Yang, J. and Zhang, Y. (2015) 'I-TASSER server: new development for protein structure and function predictions .', *Nucleic Acids Research,* 43, pp. W174–W181.

Yau, K.-W. and Hardie, R. C. (2009) 'Phototransduction motifs and variations', *Cell,* 139(2), pp. 246-264.

Yogalingam, G., Bonten, E. J., van de Vlekkert, D., Hu, H., Moshiach, S., Connell, S. A. and d'Azzo, A. (2008) 'Neuraminidase 1 Is a Negative Regulator of Lysosomal Exocytosis', *Developmental Cell,* 15(1), pp. 74-86.

Yokoi, T., Nishina, S., Fukami, M., Ogata, T., Hosono, K., Hotta, Y. and Azuma, N. (2016) 'Genotype–phenotype correlation of PAX6 gene mutations in aniridia', 3, pp. 15052.

Yuasa, J., Hirano, S., Yamagata, M. and Noda, M. (1996) 'Visual projection map specified by topographic expression of transcription factors in the retina', *Nature,* 382, pp. 632.

Zhang, C., Li, W.-H., Krainer, A. R. and Zhang, M. Q. (2008) 'RNA landscape of evolution for optimal exon and intron discrimination', *Proceedings of the National Academy of Sciences,* 105(15), pp. 5797.

Zhang, Q., Zhao, B., Li, W., Oiso, N., Novak, E. K., Rusiniak, M. E., Gautam, R., Chintala, S., O'Brien, E. P., Zhang, Y., Roe, B. A., Elliott, R. W., Eicher, E. M., Liang, P., Kratz, C., Legius, E., Spritz, R. A., O'Sullivan, T. N., Copeland, N. G., Jenkins, N. A. and Swank, R. T. (2003) 'Ru2 and Ru encode mouse orthologs of the genes mutated in human Hermansky-Pudlak syndrome types 5 and 6', *Nature Genetics,* 33, pp. 145.

Zhang, T., Hua, R., Xiao, W., Burdon, K. P., Bhattacharya, S. S., Craig, J. E., Shang, D., Zhao, X., Mackey, D. A., Moore, A. T., Luo, Y., Zhang, J. and Zhang, X. (2009) 'Mutations of the EPHA2 receptor tyrosine kinase gene cause autosomal dominant congenital cataract', *Human Mutation,* 30(5), pp. E603-E611.

Zhang, X. H. F. and Chasin, L. A. (2004) 'Computational definition of sequence motifs governing constitutive exon splicing', *Genes & Development,* 18(11), pp. 1241-1250.

Zhang, Y. (2008) 'I-TASSER server for protein 3D structure prediction.', *BMC Bioinformatics,* 9(40).

Zhao, Z., Tavoosidana, G., Sjölinder, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Singh Sandhu, K., Singh, U., Pant, V., Tiwari, V., Kurukuti, S. and Ohlsson, R. (2006) 'Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions', *Nature Genetics,* 38, pp. 1341.

Zuber, M. E., Gestri, G., Viczian, A. S., Barsacchi, G. and Harris, W. A. (2003) 'Specification of the vertebrate eye by a network of eye field transcription factors', *Development,* 130(21), pp. 5155.

Zuhlke, C., Stell, A. and Kasmann-Kellner, B. (2007) 'Genetics of oculocutaneous albinism.', *Ophthalmologe,* 104(8), pp. 674-680.

Appendix A Appendices

A.1 *SLC38A8* and *p53* reverse transcription PCR primers

5' RLM-RACE primer	Primer sequence (5'-3')	Product size (bp)
SLC38A8 inner primer	gagctctcactgagttgggg	-
SLC38A8 outer primer	ccttccagaaaagcctcaccc	-
5' RACE inner primer (supplied in kit)	cgcggatccgaacactgcgtttgctggctttgatg	-
5' RACE outer primer (supplied in kit)	gctgatggcgatgaatgaacactg	-
UTR-Primer_1-F	ggagggcaggagtgagaata	282 (when used with cDNA_ex1R primer)
UTR-Primer_2-F	gggagagaggattgatgggg	371 (when used with cDNA_ex1R primer)
UTR-Primer_3-F	agactgtccggcaccttaaa	477 (when used with cDNA_ex1R primer)
UTR-Primer_4-F	aatgactcactcacccctgg	590 (when used with cDNA_ex1R primer)
UTR-Primer_5-F	tgctcaggtcagaagtctca	714 (when used with cDNA_ex1R primer)
cDNA_ex1-R	cttccagaaaagcctcaccc	-

SLC38A8 RLM-RACE nested primers and *SLC38A8* 5' UTR primers for RT-PCR.

Primer	Primer sequence (5'-3')	Product (bp)
cDNA_*SLC38A8*_E x7-8	**F:** aatgatatggtcatcattgtgg **R:** aggatggtcagcggcatc	gDNA = 600 cDNA = 188
cDNA_*SLC38A8*_E x7-8_Nested	**F:** tgctgtctccatcgtaactgt **R:** cagctcctcctccagaagtc	gDNA = 493 cDNA = 81

Primer sequences spanning *SLC38A8* exons 7-8 for amplification of *SLC38A8* in RNA.

Primer	Primer sequence (5'-3')	Product (bp)
p53	**F:** agatattcccctgccctcaaca **R:** ctggagtcttccagtgtgat	gDNA = 1059 cDNA = 410

p53 primers

A.2 Code for WES/WGS bioinformatics analysis

WES QC commands:

bcftools stats

$ bcftools stats combined.variants.vcf > sample.stats.txt

$ plot-vcfstats sample.stats.txt -p sample.stats

Picard CollectMultipleMetrics

$ java -Xmx4g -jar /home/picard/picard-tools-1.141/picard.jar CollectMultipleMetrics I=finalbam.bam O=sample.bam_metrics.txt R=/home/ref/b37/human_g1k_v37.fasta

Samtools flagstat

$ samtools flagstat finalbam.bam > sample.flagstats.txt

WES Alignment

Trim fastq files:

$ trim_galore -q 20 --fastqc_args "--outdir /samplefolder" --gzip -o /samplefolder --paired sample_R1_001.fastq.gz sample_R2_001.fastq.gz

Align fastq files to reference genome GRCh37/hg19 and pipe SAM output directly into BAM format:

$ bwa mem -t 12 -M /home/ref/b37/human_g1k_v37.fasta sample_R1_001.fastq.gz sample_R2_001.fastq.gz -v 1 -R '@RG\tID:sample\tSM:sample_SM\tPL:ILLUMINA\tLB:sample' -M | samtools view -Sb - > sample.bwamem.bam

Sort aligned BAM file using Picard:

$ java -Xmx2g -jar /home/picard/picard-tools-2.5.0/picard.jar SortSam I=sample.bwamem.bam O=sample.bwamem.sort.bam SO=coordinate CREATE_INDEX=TRUE

Mark PCR duplicates using Picard:

$ java -Xmx4g -jar /home/picard/picard-tools-2.5.0/picard.jar MarkDuplicates I=sample.bwamem.sort.bam O=sample.bwamem.sort.dedup.bam M=sample.bwamem.sort.dedup.metrics CREATE_INDEX=TRUE

Create indel realigner targets using GATK:

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T RealignerTargetCreator -R /home/ref/b37/human_g1k_v37.fasta -known /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -known /home/ref/b37/1000G_phase1.indels.b37.vcf -I sample.bwamem.sort.dedup.bam -o sample.bwamem.sort.dedup.indelrealn.intervals

Perform indel realignment using GATK:

$ 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/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -known /home/ref/b37/1000G_phase1.indels.b37.vcf -I sample.bwamem.sort.dedup.bam -targetIntervals sample.bwamem.sort.dedup.indelrealn.intervals -o sample.bwamem.sort.dedup.indelrealn.bam

Recalibrate Base Quality Scores using GATK. Obtain the recalibration model:

$ java –Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T BaseRecalibrator -R /home/ref/b37/human_g1k_v37.fasta -knownSites /home/ref/b37/1000G_phase1.indels.b37.vcf -knownSites /home/ref/b37/dbSnp146.b37.vcf.gz -knownSites /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf -I sample.bwamem.sort.dedup.indelrealn.bam -o sample.bwamem.sort.dedup.indelrealn.recal.grp

Check the recalibration model:

$ java –Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar –T BaseRecalibrator –I sample.bwamem.sort.dedup.indelrealn.bam -R /home/ref/b37/human_g1k_v37.fasta –BQSR sample.bwamem.sort.dedup.indelrealn.recal.grp –o sample.bwamem.sort.dedup.indelrealn.postrecal.grp -knownSites /home/ref/b37/1000G_phase1.indels.b37.vcf -knownSites /home/ref/b37/dbSnp146.b37.vcf.gz -knownSites /home/ref/b37/Mills_and_1000G_gold_standard.indels.b37.sites.vcf

$ java –Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar –T AnalyzeCovariates –R /home/ref/b37/human_g1k_v37.fasta –before sample.bwamem.sort.dedup.indelrealn.recal.grp –after sample.bwamem.sort.dedup.indelrealn.postrecal.grp –plots sample.bwamem.sort.dedup.indelrealn.postrecal.plots.pdf

Apply the recalibration and print the final BAM file:

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T PrintReads -R /home/ref/b37/human_g1k_v37.fasta -I sample.bwamem.sort.dedup.indelrealn.bam -BQSR sample.bwamem.sort.dedup.indelrealn.recal.grp -o sample.bwamem.sort.dedup.indelrealn.recal.bam –disable_indel_quals

Call variants in g.vcf format using GATK HaplotypeCaller:

$ java -Xmx10g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T HaplotypeCaller -R /home/ref/b37/human_g1k_v37.fasta -D /home/ref/b37/dbSnp146.b37.vcf.gz --emitRefConfidence GVCF --variant_index_type LINEAR --variant_index_parameter 128000 -stand_call_conf 30 -stand_emit_conf 10 -I sample.bwamem.sort.dedup.indelrealn.recal.bam -o sample.raw.g.vcf

Perform genotyping using GATK GenotypeGVCFs:

$ java -Xmx8g -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 sample.raw.g.vcf --variant control.g.vcf -- variant control.g.vcf -o sample.raw.vcf --showFullBamList

Separately isolate SNPs and Indels from the genotyped vcf and perform hard filtering using GATK;

Isolate SNPs:

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T SelectVariants -R /home/ref/b37/human_g1k_v37.fasta -V sample.raw.vcf -selectType SNP -o sample_raw_snps.vcf

Isolate Indels:

java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T SelectVariants -R /home/ref/b37/human_g1k_v37.fasta -V sample.raw.vcf -selectType INDEL -o sample_raw_indels.vcf

Perform hard filtering on SNPs:

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T VariantFiltration -R /home/ref/b37/human_g1k_v37.fasta -V sample_raw_snps.vcf --filterExpression "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MappingQualityRankSum < -12.5 || ReadPosRankSum < -8.0" --filterName "snp_hard_filter" -o sample_filtered_snps.vcf

Perform hard filtering on Indels:

$ java -Xmx4g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T VariantFiltration -R /home/ref/b37/human_g1k_v37.fasta -V sample_raw_indels.vcf --filterExpression "QD < 2.0 || FS < 200.0 || ReadPosRankSum < -20.0" --filterName "indel_hard_filter" -o sample_filtered_indels.vcf

Combine hard filtered SNPs and Indels using GATK:

$ java -Xmx12g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T CombineVariants -R /home/ref/b37/human_g1k_v37.fasta --variant sample_filtered_snps.vcf --variant sample_filtered_indels.vcf --genotypemergeoption UNSORTED -o sample_combinedvariants.vcf

Filter on dbSNP146 at 0.01%:

$ perl /home/vcfhacks-v0.2.0/annotateSnps.pl -d /home/ref/b37/dbSnp146.b37.vcf.gz /home/ref/b37/clinvar_20160531.vcf.gz -f 0.01 -pathogenic -i sample_combinedvariants.vcf -o sample_dbSNP146.vcf -n

Filter on ExAC at 0.01%:

$ perl /home/vcfhacks-v0.2.0/filterVcfOnVcf.pl -i sample_dbSNP146.vcf -f /home/ref/ExAC/ExAC.r0.3.sites.vep.vcf.gz -o sample_dbSNP146_ExAC1.vcf -y 0.0001 –w

Annotate with VEP:

$ perl /home/variant_effect_predictor/variant_effect_predictor.pl --offline --vcf --dir_cache /home/variant_effect_predictor/vep_cache --dir_plugins /home/variant_effect_predictor/vep_cache/Plugins --everything --fasta /home/variant_effect_predictor/fasta/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz -i sample_dbSNP146_ExAC1.vcf -o sample_dbSNP146_ExAC1.vep.vcf

Filter using a BED file to retain variants in known ophthalmic disease genes:

$ perl /home/vcfhacks-v0.2.0/getVariantsByLocation.pl -i sample_dbSNP146_ExAC1.vep.vcf -b /data/umecl/BEDFILE.bed -o sample_dbSNP146_ExAC1.vep.BED.vcf

Filter to retain all functional variants:

$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl -i sample_dbSNP146_ExAC1.vep.BED.vcf -s SAMPLE -o sample_dbSNP146_ExAC1.vep.BED.functional.vcf -b

Filter & Rank on CADD Score:

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl -c /data/shared/cadd/v1.3/*.gz -i sample_dbSNP146_ExAC1.vep.BED.functional.vcf -o sample_dbSNP146_ExAC1.vep.BED.functional.cadd.vcf

Ensembl Gene Annotator:

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb --i sample_dbSNP146_ExAC1.vep.BED.functional.cadd.vcf -o sample_dbSNP146_ExAC1.vep.BED.functional.cadd.vcf.geneanno

Convert the annotated file into an Excel spreadsheet:

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i sample_dbSNP146_ExAC1.vep.BED.functional.cadd.vcf.geneanno --vep --gene_anno -o sample_dbSNP146_ExAC1.vep.BED.functional.cadd.vcf.geneanno.simple.xlsx

<u>Filter to retain all rare, functional biallelic variants (compound heterozygous and homozygous)</u>

Filter to retain all biallelic variants:

$ perl /home/vcfhacks-v0.2.0/findBiallelic.pl -i sample_dbSNP146_ExAC1.vep.vcf -s SAMPLE -o sample_ dbSNP146_ExAC1.vep.biallelic.vcf

Filter to retain all functional variants:

$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl -i genotyped_sample_ dbSNP146_ExAC1.vep.biallelic.vcf -s SAMPLE -o genotyped_sample_ dbSNP146_ExAC1.vep.biallelic.functional.vcf -b

Filter & Rank on CADD score:

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl -c /data/shared/cadd/v1.3/*.gz -i sample_ dbSNP146_ExAC1.vep.biallelic.functional.vcf -o sample_ dbSNP146_ExAC1.vep.biallelic.functional.cadd.vcf

Annotate genes:

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb --i sample_ dbSNP146_ExAC1.vep.biallelic.functional.cadd.vcf -o sample_ dbSNP146_ExAC1.vep.biallelic.functional.cadd.vcf.geneanno

Convert the annotated file into an Excel spreadsheet:

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i sample_ dbSNP146_ExAC1.vep.biallelic.functional.cadd.vcf.geneanno --vep --gene_anno -o sample_ dbSNP146_ExAC1.vep.biallelic.functional.cadd.vcf.geneanno.simple.xlsx

<u>Filter to retain all rare, functional heterozygous variants</u>

Filter to retain all heterozygous variants:

$ perl /home/vcfhacks-v0.2.0/getHetVariants.pl –i sample_dbSNP146_ExAC1.vep.vcf –s SAMPLE –o sample_dbSNP146_ExAC1.vep.het.vcf

Filter to retain all functional variants:

$ perl /home/vcfhacks-v0.2.0/getFunctionalVariants.pl -i genotyped_sample_ dbSNP146_ExAC1.vep.het.vcf -s SAMPLE -o genotyped_sample_ dbSNP146_ExAC1.vep.het.functional.vcf -b

Filter & Rank on CADD score:

$ perl /home/vcfhacks-v0.2.0/rankOnCaddScore.pl -c /data/shared/cadd/v1.3/*.gz -i sample_ dbSNP146_ExAC1.vep.het.functional.vcf -o sample_ dbSNP146_ExAC1.vep.het.functional.cadd.vcf

Annotate genes:

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb --i sample_ dbSNP146_ExAC1.vep.het.functional.cadd.vcf -o sample_ dbSNP146_ExAC1.vep.het.functional.cadd.vcf.geneanno

Convert the annotated file into an Excel spreadsheet:

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i sample_ dbSNP146_ExAC1.vep.het.functional.cadd.vcf.geneanno --vep --gene_anno -o sample_ dbSNP146_ExAC1.vep.het.functional.cadd.vcf.geneanno.simple.xlsx

WES GATK Depth of Coverage

Calculate the depth of coverage over the entire genome, or a specified region of the genome using a bed file:

$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T DepthOfCoverage -R /home/ref/b37/human_g1k_v37.fasta -I sample.bwamem.sort.dedup.indelrealn.recal.bam -o sample_depthofcoverage.txt -L /home/ref/b37/AllRefSeqGenes.b37.bed -ct 5 -ct 10 -ct 15 -ct 20 -ct 25 -ct 30

Compute the depth of coverage over all *SLC38A8* exons in Case F1335 (alongside 5 control exomes):

$ java -Xmx8g -jar /home/GATK/GenomeAnalysisTK-3.5-0/GenomeAnalysisTK.jar -T DepthOfCoverage -R /home/ref/b37/human_g1k_v37.fasta -o F1335_duplication_depthofcoverage.txt -I F1335.bwamem.sort.dedup.indelrealn.recal.bam -I CONTROL1.bwamem.sort.dedup.indelrealn.recal.cln.bam -I CONTROL2.bwamem.sort.dedup.indelrealn.recal.cln.bam -I CONTROL3.bwamem.sort.dedup.indelrealn.recal.cln.bam -I CONTROL4.bwamem.sort.dedup.indelrealn.recal.cln.bam -I CONTROL5.bwamem.sort.dedup.indelrealn.recal.cln.bam -L SLC38A8_exons.BED -ct 5 -ct 10 -ct 15 -ct 20 -ct 25 -ct 30

SAMBLASTER commands (Case F1335):

$ bwa mem -t 12 -M /data/umecl/UKIRDC/hg19_index_base F1335_0_R1.fastq.gz F1335_0_R2.fastq.gz | /data/umecl/UKIRDC/samblaster/samblaster -M | samtools view -Sb - > /data/umecl/UKIRDC/WGS/ F1335.out.bam

$ bwa mem -t 12 -M /data/umecl/UKIRDC/hg19_index_base F1335_0_R1.fastq.gz F1335_0_R2.fastq.gz | /data/umecl/UKIRDC/samblaster/samblaster -M -d F1335.disc.sam -s F1335.split.sam | samtools view -Sb - > /data/umecl/UKIRDC/WGS/F1335.out.bam

$ samtools sort F1335.disc.sam > F1335.disc.sorted.bam

$ samtools index F1335.disc.sorted.bam

A.3 Code for CNV analysis on WES data

Define the group of BAM files:

```
> bam_files <- c("sample1.bam", "sample2.bam", …)
```

Create count data from all BAM files:

```
> 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:

```
> counts.dafr <- as(counts[, colnames(counts)], 'data.frame')
```

Define the test sample:

```
> test_sample <- counts$test_sample.bam
```

Define the reference samples:

```
> ref_samples < -c('sample1.bam', 'sample2.bam'…)
> ref_set < -as.matrix(counts.dafr[,ref_samples])
```

Optimize the choice of aggregate reference set:

```
> choice <- select.reference.set(test.counts=test_sample, reference.counts=ref_set,
bin.length=(counts.dafr$end-counts.dafr$start)/1000, n.bins.reduced=10000)
```

Construct the reference set:

```
> matrix <- (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=counts.dafr$space, start=counts.dafr$start, end=counts.dafr$end,
name=counts.dafr$names) 244
```

Load the set of common CNVs identified by (Conrad et al., 2010):

```
> data(Conrad.hg19)
```

Annotate CNV calls:

```
> all.exons <- AnnotateExtra(x=all.exons,
reference.annotation=Conrad.hg19.common.CNVs, min.overlap=0.5,
column.name='Conrad.hg19')
```

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=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)
```

A.4 BreakDancerMax code

Replace read groups in your bam file:

```
$ java -jar /home/picard/picard-tools-2.5.0/picard.jar AddOrReplaceReadGroups
I=/data/umecl/UKIRDC/sample.bwamem.sort.dedup.indelrealn.recal.cln.bam
O=4070.breakdancer.bam RGID=SAMPLE_WES RGLB=SAMPLE_WES RGPL=illumina
RGPU=SAMPLE_WES RGSM=SAMPLE_WES
```

Convert the bam file into a configuration file ready to run the BreakDancer program:

```
$ perl /home/breakdancer-1.1_2011_02_21/perl/bam2cfg.pl sample.breakdancer.bam >
sample.breakdancer.config
```

Run BreakDancerMax:

```
$ /home/breakdancer-1.1_2011_02_21/cpp/breakdancer_max sample.breakdancer.config >
sample.breakdancer.txt
```

A.5 DELLY2 code

```
$ /data/umecl/delly/src/delly call -t INV -g /data/umecl/UKIRDC/human_g1k_v37.fasta -o
F1310.b37.delly.inv.bcf
/data/umecl/UKIRDC/genome_F1310_delly/F1310.bwamemb37.sort.dedup.indelrealn.recal.
bam
```

$ /data/umecl/delly/src/bcftools/bcftools view F1310.b37.delly.inv.bcf > F1310.b37.delly.inv.vcf

$ /data/umecl/svprops/src/svprops F1310.b37.delly.inv.bcf

$ /data/umecl/svprops/src/sampleprops F1310.b37.delly.inv.bcf

$ perl /home/variant_effect_predictor/variant_effect_predictor.pl --offline --vcf --dir_cache /home/variant_effect_predictor/vep_cache --dir_plugins /home/variant_effect_predictor/vep_cache/Plugins --everything --fasta /home/variant_effect_predictor/fasta/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz -i F1310.b37.delly.inv.vcf -o F1310.b37.delly.inv.vep.vcf

$ perl /home/vcfhacks-v0.2.0/geneAnnotator.pl -d /home/vcfhacks-v0.2.0/data/geneAnnotatorDb --i F1310.b37.delly.inv.vep.vcf -o F1310.b37.delly.inv.vep.vcf.geneanno

$ perl /home/vcfhacks-v0.2.0/annovcfToSimple.pl -i F1310.b37.delly.inv.vep.vcf.geneanno --vep --gene_anno -o F1310.b37.delly.inv.vep.vcf.geneanno.simple.xlsx --functional

A.6 BED file list of 270 genes compiled from RetNet, albinism genes and candidate gene lists with average % capture on Agilent SureSelect V5/V6

Gene	V5 (%)	V6 (%)		Gene	V5 (%)	V6 (%)
ABCA4	100	100		MAPK3	100	100
ABCC6	96.8	100		MC1R	100	100
ABHD12	100	100		MERTK	100	100
ACBD5	100	100		MFN2	100	100
ADAMTS18	100	100		MFSD8	100	100
ADGRA3	99.9	100		MIR204	100	100
ADGRV1	100	100		MKKS	100	100
ADIPOR1	100	100		MKS1	100	100
AGBL5	100	100		MTTP	100	100
AHI1	100	100		MVK	100	100

AHR	100	100		MYO7A	100	100
AIPL1	100	100		NBAS	100	100
ALMS1	100	100		NDP	100	100
ARL2BP	99.9	100		NEK2	100	100
ARL3	100	100		NEUROD1	100	100
ARL6	99.2	99.2		NMNAT1	100	100
ASRGL1	100	100		NPHP1	100	100
ATF6	100	100		NPHP3	100	100
ATOH7	100	100		NPHP4	100	100
ATXN7	100	100		NR2E3	100	100
BBIP1	100	100		NR2F1	100	100
BBS1	100	100		NRL	100	100
BBS10	100	100		NYX	100	100
BBS12	100	100		OAT	99.3	100
BBS2	100	100		OFD1	99.6	100
BBS4	100	100		OPA1	100	100
BBS5	100	100		OPA3	100	100
BBS7	100	100		OPN1LW	75.4	89.8
BBS9	100	100		OPN1MW	75.4	89.8
BEST1	100	100		OPN1SW	100	100
C12orf65	100	100		OR2W3	100	100
C1QTNF5	100	100		OTX2	100	100
C21orf2	100	100		PANK2	100	100

C2orf71	100	100		PAX2	100	100
C8orf37	100	100		PAX6	100	100
CA4	100	100		PCDH15	100	100
CABP4	100	100		PCYT1A	100	100
CACNA1F	100	100		PDE6A	100	100
CACNA2D4	100	100		PDE6B	100	100
CAMK2A	100	100		PDE6C	100	100
CAPN5	100	100		PDE6G	100	100
CC2D2A	100	100		PDZD7	100	100
CDH23	100	100		PEX1	100	100
CDH3	100	100		PEX2	100	100
CDH5	100	100		PEX6	100	100
CDHR1	100	100		PEX7	100	100
CEP164	100	100		PGK1	99.1	100
CEP250	100	100		PHYH	100	100
CEP290	100	100		PITPNM3	100	100
CERKL	100	100		PITX2	100	100
CFH	100	100		PLA2G5	100	100
CHM	100	100		PLK4	100	100
CIB2	100	100		PNPLA6	100	100
CLN3	100	100		POC1B	100	100
CLRN1	100	100		POMGNT1	100	100
CLUAP1	100	100		PRCD	100	100

CNGA1	98.2	98.9		PRDM13	100	100
CNGA3	100	100		PROM1	100	100
CNGB1	100	100		PRPF3	100	100
CNGB3	100	100		PRPF4	100	100
CNNM4	100	100		PRPF6	100	100
COL11A1	100	100		PRPF8	100	100
COL2A1	100	100		PRPF31	100	100
COL9A1	100	100		PRPH2	100	100
CRB1	100	100		RAB28	100	100
CRYAA	100	100		RAX2	100	100
CRYBB2	100	100		RB1	100	100
CRYBB3	100	100		RBP3	100	100
CRYGC	100	100		RBP4	100	100
CRYGD	100	100		RCBTB1	100	100
CRX	100	100		RD3	100	100
CSPP1	100	100		RDH11	100	100
CTNNA1	100	100		RDH12	100	100
CYP4V2	100	100		RDH5	100	100
DHDDS	100	100		RGR	100	100
DHX38	100	100		RGS9	100	100
DMD	100	100		RGS9BP	100	100
DRAM2	100	100		RHO	100	100
DTHD1	100	100		RIMS1	100	100

EFEMP1	100	100		RLBP1	100	100
ELOVL4	100	100		ROM1	100	100
EMC1	100	100		RP1	100	100
EXOSC2	100	100		RP1L1	99.9	100
EYS	100	100		RP2	100	100
FAM161A	100	100		RP9	100	100
FLVCR1	100	100		RPE65	100	100
FOXD1	0	0		RPGR	75.1	75.1
FOXC1	100	100		RPGRIP1	100	100
FRMD7	100	100		RPGRIP1L	100	100
FSCN2	100	100		RS1	100	100
FZD4	100	100		RTN4IP1	100	100
GDF6	100	100		SAG	100	100
GNAT1	100	100		SDCCAG8	100	100
GNAT2	100	100		SEMA4A	100	100
GNB3	100	100		SLC24A5	100	100
GNPTG	100	100		SLC24A1	100	100
GPR143	100	100		SLC25A46	100	100
GPR179	100	100		SLC38A8	100	100
GRM6	100	100		SLC7A14	100	100
GUCA1A	100	100		SLC45A2	100	100
GUCA1B	100	100		SNRNP200	100	100
GUCY2D	100	100		SPATA7	100	100

HARS	100	100		SPP2	100	100
HERC2	79.3	92.8		TEAD1	100	100
HGSNAT	100	100		TIMM8A	100	100
HK1	100	100		TIMP3	100	100
HMCN1	100	100		TMEM126A	100	100
HMX1	100	100		TMEM237	100	100
IDH3A	100	100		TOPORS	100	100
IDH3B	100	100		TREX1	100	100
IFT140	100	100		TRIM32	100	100
IFT172	100	100		TRIM44	100	100
IFT27	100	100		TRNT1	100	100
IMPDH1	99	100		TRPM1	100	100
IMPG1	100	100		TSPAN12	100	100
IMPG2	100	100		TTC8	100	100
INPP5E	100	100		TTLL5	100	100
INVS	100	100		TTPA	100	100
IQCB1	100	100		TUB	100	100
ITM2B	100	100		TULP1	100	100
JAG1	100	100		TYR	100	100
KCNJ13	100	100		TYRP1	100	100
KCNV2	100	100		UNC119	100	100
KIAA1549	99.7	99.7		USH1C	100	100
KIF11	100	100		USH1G	100	100

KIZ	100	100		USH2A	100	100
KLHL7	100	100		VCAN	100	100
LAMA1	100	100		VPS13B	100	100
LCA5	100	100		WDPCP	100	100
LGR4	100	100		WDR19	100	100
LRAT	100	100		WFS1	100	100
LRMDA	100	100		WHRN	100	100
LRP5	100	100		ZNF408	100	100
LZTFL1	100	100		ZNF423	100	100
MAK	100	100		ZNF513	100	100

A.7 Primer sequences used for PCR and Sanger sequencing all coding regions of *PAX6* and *PAX6* *cis*-regulatory elements

Exon	Forward sequence (5'-3')	Reverse sequence (5'-3')	Product (bp)	PCR conditions
4	ACCTCGGTTGGGAGTTCAG	GTCGCGAGTCCCTGTGTC	217	St/62A_T
5	ATCTTCCTCTTCCTTCTTCTCC	GAAATGAAGAGAGGGCGTTG	290	St/60A_T
5a	GGGCTACAAATGTAATTTTAAGAAA	CTCACACATCCGTTGGACAC	236	St/60A_T
6	AGTGCTGGACAATCAAAACG	GGTGGGAGGAGGTAAAGAGG	400	St/60A_T
7	GGTTGTGGGTGAGCTGAGAT	AAGCCCTGAGAGGAAATGGT	333	St/63A_T
8	AGACTACACCAGGCCCCTTT	CACTGAAAAGATGCCCAGAGA	387	St/60A_T
9	AGGTGGGAACCAGTTTGATG	CAAGCACCTCTGTCTCTAGGAA	240	St/60A_T
10	CAGTGGAGGTGCCAAGGT	CAGTAGTCAGAGTGAGAGTCAGAGC	300	St/60A_T
11	TCAATGTCTTTAAACCTGTTTGC	CTCTCAAGGGTGCAGACACA	289	St/60A_T
12	CAGACTTGTTGGCAGAGTTCC	TAAACACGCCCTCCCATAAG	345	St/60A_T
13	TTTCTGAAGGTGCTACTTTTATTTG	CGGCTCTAACAGCCATTTTT	373	St/58A_T*

Abbreviation are as follows: St = standard PCR with Taq polymerase and 1.5 mM $MgCl_2$, A_T = annealing temperature (in °C). * Note that the PCR program of *PAX6* exon 13 differed slightly from the standard PCR reaction: 95°C for 5 minutes, 95°C for 1 minute, 55°C for 2 minutes, 72°C for 2 minutes (x40 cycles), final extension step of 72°C for 7 minutes.

Element	Human coordinates (GRCh37)	Primer sequence (5'-3')	PCR conditions	Reference
E-200 distal	chr11:32053010-32053783	tcagggtgttctgtgacct tggcccacttcactggtca	St/60A$_T$	(Ravi *et al.*, 2013)
E-200 proximal	chr11:32052714-32053362	ggcacaagatagaaagccctg gaaagtgagagccatccca	TD/58A$_T$	(Ravi *et al.*, 2013)
CTCF6 element	chr11:31892452-31893062	cactcagaggcagagcaga tggatgaagaagccagaa	St/55A$_T$	Unpublished
EE ectodermal enhancer	chr11:31843104-31843692	aagttttcccggacctctgt aactcgatcacatggacct	TD/56A$_T$	(Williams *et al.*, 1998)
OCE1	chr11:31837809-31838293	gccctggggaaatctgaata gtgacaggtgttgggacct	St/58A$_T$	Unpublished
NRE	chr11:31825361-31826132	atctcacaacctcatccct aggccttgctctgttgga	TD/58A$_T$	(Kammandel *et al.*, 1999)
7CE1	chr11:31820755-31821453	ggtcaatcgctagtctcacca cgaggtagcttggtcaagaa	HSMM/TD/55A$_T$	(Kleinjan *et al.*, 2004)
E60A	chr11:31784779-31785426	aagccctttcctgtcgcttg atgggctacctgtgcaca	St/60A$_T$	(McBride *et al.*, 2011)
E100	chr11:31734245-31734929	aaatagcactggatgctgtct gaaatgattgaggcagaaggt	St/60A$_T$	(Griffin *et al.*, 2002)
SIMO	chr11:31685581-31686174	gaactcaatgtagtgttttcagcc gactttctcaggaagaaatcagg	St/55A$_T$	(Kleinjan *et al.*, 2001)
HS2	chr11:31677439-31677991	ggctacatgactagacacatgc tacgtgttgatgaccacaca	St/60A$_T$	(Kleinjan *et al.*, 2001)
HS3	chr11:31676764-31677347	caaagaatcatcagaagttgg tcaaatattttgggaacttacct	St/55A$_T$	(Kleinjan *et al.*, 2001)
HS5	chr11:31671019-31671690	ctgacaagaggccaaagctc cgaggaatctgccgtatatga	St/60A$_T$	(McBride *et al.*, 2011)
HS8A	chr11:31662868-31663475	gtgaagtctgtagcccaaga agctcacgccctgccttga	St/60A$_T$	(Kleinjan *et al.*, 2001)
HS8B	chr11:31662110-31663778	aatgttttcgcaccctatg gcatgtcacagcaaaacagc	St/58A$_T$	(Kleinjan *et al.*, 2001)

Abbreviations are as follows: St = standard PCR with Taq polymerase and 1.5 mM $MgCl_2$. TD = touchdown PCR, HSMM = HotShot master mix, A_T = annealing temperature (in °C).

A.8 Homozygous and heterozygous variants identified in *PAX6 cis*-regulatory elements

cis element	Patient	Variant (HGVS)	Position on chr11	dbSNP rs ID	Allele count & frequency (gnomAD)	No. Hom (gnomAD)	FATHMM non-coding score
OCE1	F1371 (Het)	c.-317+1196G>C (NM_001127612.1)	31,838,161	rs1806184	NA	NA	0.14332
	F1369, F1402, F1405 (Het) F1373, F1071 (Hom)	c.-317+1229C>T (NM_001127612.1)	31,838,128	rs4440995	31743/154382 0.2056	3589	0.47668
	F1369, F1402, F1405 (Het) F1373, F1071 (Hom)	c.-317+1408G>A (NM_001127612.1)	31,837,949	rs1540318	6279/30878 0.2033	645	0.15129
HS8B	All patients (Hom)	c.931-6633C>T (NM_001288725.1)	31,662,656	rs7480296	28809/30886 0.9328	13533	0.59371
HS5	F1373 (Het)	c.1040-120A>G (NM_001288725.1)	31,671,543	rs75155674	250/30846 0.008105	3	0.98938
E200-distal	F1371, F1402, F1298 (Het)	g.32053419A>G (NC_000011.9)	32,053,419	rs11031557	3422/30934 0.1106	224	0.27340
	F1372 (Het)	g.32053271G>A (NC_000011.9)	32,053,271	rs11031556	333/30978 0.01075	1	0.97956
E200-proximal	F1371, F1071, F1298, F1403 (Het)	g.32053007A>T (NC_000011.9)	32,053,007	rs1002229	5036/30926 0.1628	555	0.96450
	F1071 (Het)	g.32053046T>A (NC_000011.9)	32,053,046	rs148075639	114/30974 0.003681	0	0.96728
	F1372 (Het)	g.32053271G>A (NC_000011.9)	32,053,271	rs11031556	333/30978 0.01075	1	0.97956
E100	F1371, F1373, F1352, F1369, F1298, F1402, F1404, F1405, F1071, F1419 (Het) F1371 (Hom)	c.1147-50218T>C (NM_001288725.1)	31,734,741	rs7926476	8951/30914 0.2895	1410	0.99489

CTCF6	F1373, F1402, F1371, F1298, F1288, F1071, F1419, F1421 (Het) F1374, F1372, F1403, F1404, F1287 (Hom)	g.31892512C>T (NC_000011.9)	31,892,512	rs662155	25375/30926 0.8205	10538	0.16798
NRE	F1369, F1373, F1071, F1402 (Het) F1372, F1374, F1352, F1371, F1298, F1287, F1288, F1405, F1404, F1403, F1421 (Hom)	c.11-1136A>C (NM_000280)	31,825,518	rs694617	26920/30906 0.8710	11765	0.55452

Their location on chromosome 11 is shown, using human reference genome build GRCh37, Ensembl transcript ID ENST00000419022 (RefSeq NM_000280).

A.9 Rare heterozygous candidate variants identified in patients F1405 and F1352

Gene	RefSeq	Chr position	cDNA change	AA change	Variant Consequence	No. reads WT/Mut	dbSNP rs ID	ExAC	gnomAD	No. Hom	PolyPhen-2 score	SIFT score	CADD score
VSX2	NM_182894.2	14:74,727,428	c.892C>T	p.(Gln298Ter)	Nonsense	37/27	.	.	0/206716	0	.	.	39
EPHA2	NM_004431.4	1:16,464,512	c.1148T>G	p.(Val383Gly)	Missense	57/62	rs7510403034	.	1/245434 4.07e-6	0	benign	damaging	17.9

A.10 *FOXD1* PCR primers and conditions, Sanger sequencing primers and conditions and SDM primers

Amplicon 1

PCR primers (5'-3'):

HuFOXD1-1Fb: CGGCGCAGGAGTTATAAAGTCGGCG

HuFOXD1-1Rb: CGCCGCGCAGCAGCAGAGACTCG

Nested PCR primers:

HuFOXD1-1nest1R: CGTTGGGTGGGAGCAG

HuFOXD1-1nest2R: AAGCTGCCGTTGTCGAAC

Both used with HuFOXD1-1Fb: CGGCGCAGGAGTTATAAAGTCGGCG

PCR conditions:

NEB Q5 DNA Polymerase with Q5 reaction Buffer and 5X enhancer.

PCR program:

98°C for 30 seconds, 98°C for 7 seconds, 68.6°C for 15 seconds, 72°C for 18 seconds (x35 cycles), final extension step of 72°C for 1 minute 30 seconds.

Sanger sequencing enhancers:

NEB Q5 5X enhancer added into Sanger sequencing reaction.

Amplicon 1 (2)

PCR primers (5'-3'):

TF2-Forward: AGCAGTCGTTGAGCGAGAGGTT

TF2-Reverse: CCTGAGCACTGAGATGTCCGAT

TF4-Forward: AGTCCGCCGACATGTTCG

TF4-Reverse: TGAAGCCGCCCTACTGCTATA

PCR conditions:

KAPA DNA Polymerase with KAPA GC-Buffer, 1.3M Betaine and 5% DMSO.

PCR program:

TD-PCR program as described in Section X, $\mathbf{A_T}$ of 62°C.

Sanger sequencing primers (5'-3'):

174-Forward: GTACGCCGGGGAGGACGAGCTGGA

TF-3-Reverse: CGAGCTGGAGGATCTGGAG

TF-5-Forward: CGCTTGAAGCGCTTCCTC

Amplicon 2

PCR primers (5'-3'):

HuFOXD1-2Forward: GCAACTACTGGACGCTGGACCCGGAG

HuFOXD1-2Reverse: CGGGCTGTTGACATTTTGTCCGAG

Nested PCR primers:

HuFOXD1-2nest1Forward: CCGACATGTTCGACAACG

HuFOXD1-2nest1Reverse: CGCAGCGGCGAAAAT

PCR conditions:

KAPA Polymerase with GC-Buffer, 1.3M Betaine, 5% DMSO, 50pmol/μl 7-deaza-dGTP.

PCR program:

98°C for 30 seconds, 98°C for 10 seconds, 56. 9°C for 30 seconds, 72°C for 10 seconds (x35 cycles), final extension step of 72°C for 1 minute.

Sanger sequencing primers (5'-3'):

2R2b: CGCGCCTGGAGGAGCGAACAAAACA

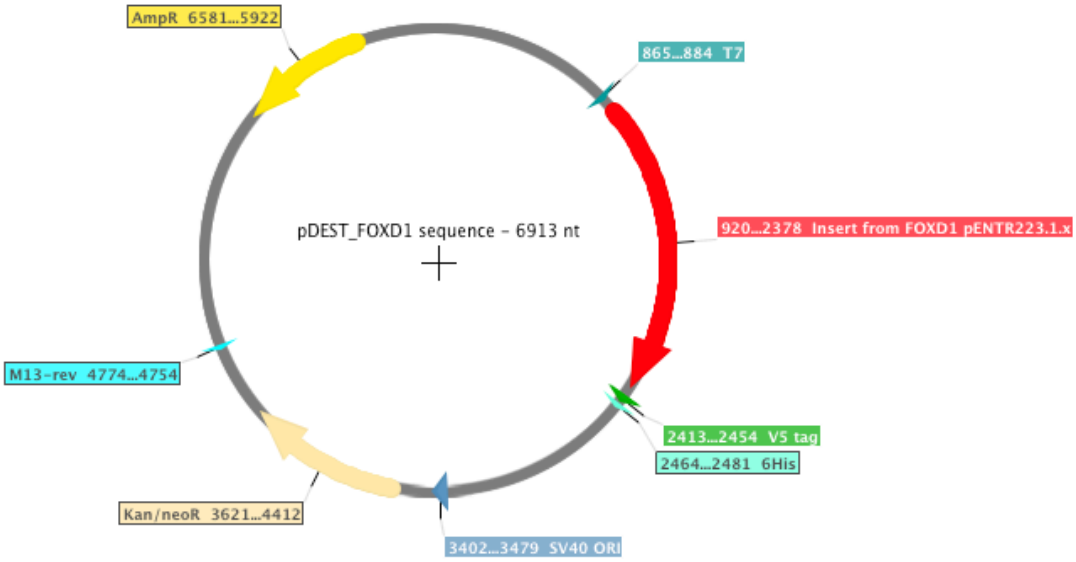
OM: TTAACAATTGGAAATCCTAGCAG

980-R: GAGCGGGTGCGGCCGGTAGC

SDM Primers

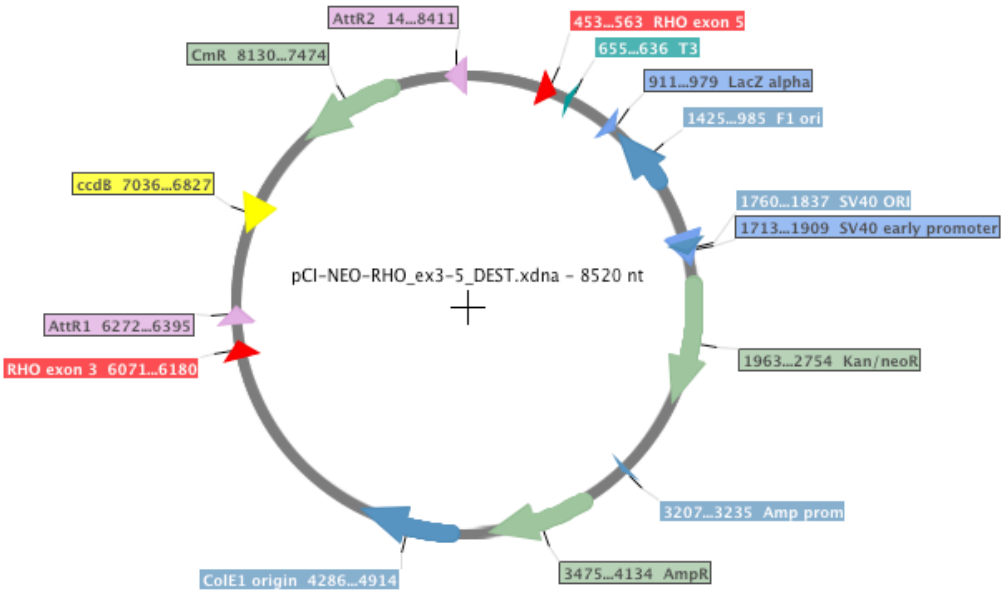
Primer	Sequence (5'-3')
FOXD1-SDM-F	gtagttgcccttgtccgggttgccggg
FOXD1-SDM-R	cccggcaacccggacaagggcaactac

A.11 Expression clones



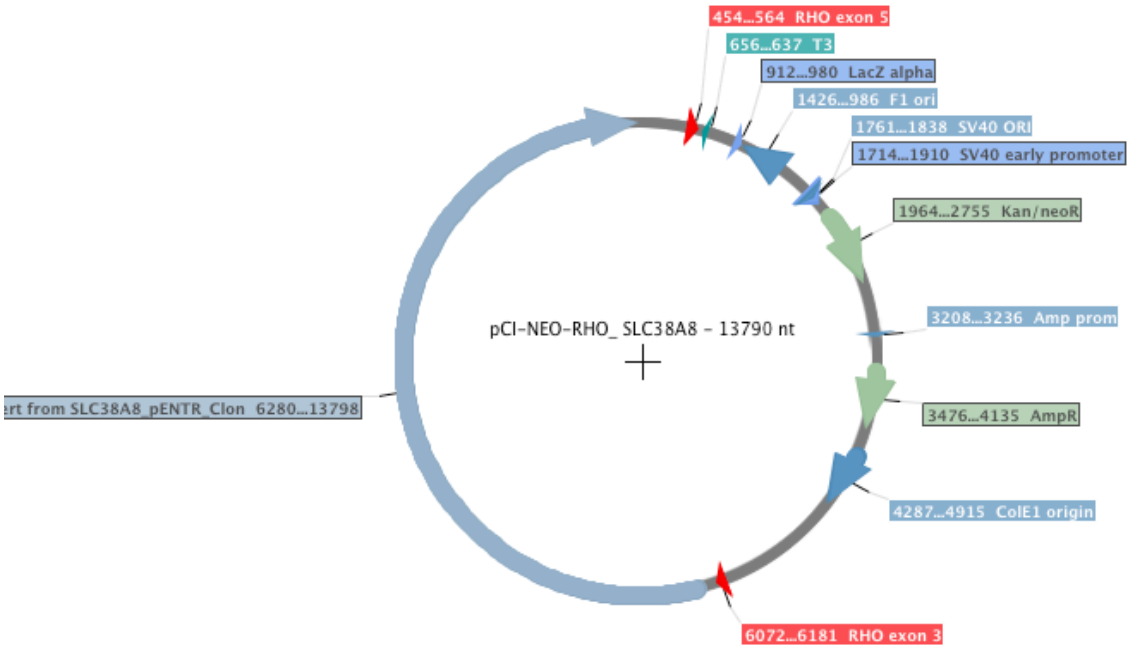
Schematic representation of the pDEST_*FOXD1* expression clone.

Highlighted in red is the *FOXD1* open reading frame. Primers for Sanger sequencing are as follows: Blue; M13R, Turquoise; T7. The His-tag is highlighted in pale green and the V5 tag is in green. The Ampicillin resistance is highlighted in bright yellow and the Kanamycin/Neomycin resistance in pale yellow.



Schematic representation of the pCI-Neo-*RHO*-ex3_5/DEST.

The splice vector was a kind gift of Erwin van Wijk. Highlighted in red are the *RHO* exons 3 and 5 (exon 4 and flanking intronic sequences removed by digestion with *Eco*NI and *Pfl*MI and blunted using large fragment DNA polymerase I).



Schematic representation of the pCI-NEO-*RHO_SLC38A8*.

The *SLC38A8* open reading frame is highlighted in blue. The *RHO* exons 3 and 5 are coloured red. Ampicillin and kanamycin are highlighted in green.

A.12 *SLC38A8* primer sequences

Exon	Forward sequence (5'-3')	Reverse sequence (5'-3')	Product (bp)	PCR enhancers	PCR conditions
1	gagctctcactgagttgggg	cctcattagctgagcggg	437	10% D	HS/65A_T
2	tgtggagagtgtgcccg	ggagcctcccttccctac	449	-	HS/60A_T
3	gtggatgtccaaccccag	caaatgtgaaggccaattcc	350	-	HS/60A_T
4	acctcattgtcaagggcg	ctctgcagtgagccacgag	312	-	HS/60A_T
5	ttgcagccatgctctgttac	ccctgacagagaaaccaagg	266	-	HS/60A_T
6	tcaggctatttcaagagaaacctc	tgaggccttttcctatgcac	398	-	HS/60A_T
6 (SNP deletion)		ctactgtcactccccaccctc	287	-	HS/60A_T
7	aaggaaaatcttaggtggtaaagc	agtcccacagccgcgtag	352	-	HS/60A_T
8	tgtggaacaggagcactgac	tcctacccatatgtggctcc	422	-	HS/60A_T
9	ccttccagctccactgtgtt	cagatgccctcatactgggt	380	-	HS/60A_T
10	ttgatgactccatactgggc	gaggaaaagaaatggcatcg	294	-	HS/60A_T

Abbreviations are as follows: HS = HotShot mastermix, A_T = annealing temperature (in °C), D = DMSO.

A.13 Non-pathogenic intronic and coding region variants identified in *SLC38A8*

Patient	Location on Chr16	cDNA & aa change	dbSNP ID	Allele count & frequency (gnomAD)	No. reported homozygotes
F1377, F1369, F1298, F1071, F1404, F1405, F1371 (Het) F1288, F1335, F1372, F1351, F1352, F1368, F1402 (Hom)	84,075,563	c.189+11G>T	rs1105355	135618/248466 0.5458	37248
F1371 (Het)	84,075,482	c.189+92G>A	rs57933451	2072/30900 0.06706	126
F1371 (Het)	84,075,440	c.189+134G>C	rs76310525	1471/30906 0.04760	45
F1352, F1403, F1298, F1071, F1421 (Het) F1369, F1310 (Hom)	84,070,514	c.190-9C>T	rs1876960	95698/260968 0.3667	18203
F1352, F1403, F1298, F1071, F1421 (Het) F1369, F1310 (Hom)	84,070,510	c.190-5C>T	rs1876962	98069/264594 0.3706	18400
F1352, F1403, F1298, F1071, F1419, F1421 (Het) F1369, F1310 (Hom)	84,070,500	c.195G>C p.(Ser65=)	rs1317524	98915/269200 0.3674	18527
F1352, F1403, F1298, F1071, F1421 (Het) F1369, F1310 (Hom)	84,070,238	c.388+69G>T	rs8057543	11930/30190 0.3952	2451
F1378, F1372, F1351, F1071, F1419 (Het) F1371, F1335, F1402, F1405 (Hom)	84,070,179	c.388+128A>T	rs34702590	7143/30908 0.2311	794
F1371 (Het)	84,067,023	c.440C>G p.(Ala147Gly)	rs56802364	2822/275206 0.01025	144
F1071, F1421 (Het) F1310 (Hom)	84,067,004	c.459G>A p.(Leu153=)	rs57576928	23930/276494 0.08655	1320
F1371 (Het)	84,066,858	c.530+75G>A	rs75604026	929/30840 0.03012	39
F1405 (Hom)	84,065,790	c.531-220_531-218delCAA	rs140016057	21367/30569 0.6990	7526

F1371 (Het)	84,065,586	c.531-13G>A	rs11865628	2839/276270 0.01028	142
F1071 (Het), F1310 (Hom)	84,063,219	c.633-63C>G	rs61432810	2269/30966 0.07327	114
F1371, F1071 (Het) F1310 (Hom)	84,063,215	c.633-59C>T	rs11863267	3183/30940 0.1029	170
F1371 (Het)	84,063,130	c.659G>C p.(Ser220Thr)	rs11862366	7820/277006 0.02823	291
F1310 (Hom)	84,063,036	c.690+65C>T	rs62045928	4095/30942 0.1323	291
F1371 (Het)	84,056,442	c.744C>G p.(Ser248Cys)	rs11861325	2951/274852 0.01074	158
F1351, F1369, F1374, F1377, F1402, F1403, F1298, F1071, F1421, F1405 (Het)	84,056,326-84,056,327	c.805+53_805+54delAA	rs5818479	0.23063 (1000 Genomes)	-
F1368 (Hom)	84,050,970	c.806-78C>A	rs28669064	2064/30964 0.06672	60
F1372 (Het)	84,050,209	c.1077C>T p.(Leu359=)	rs77876966	12388/276570 0.04479	377
F1374, F1351, F1405 (Het) F1402, F1071 (Hom)	84,050,037	c.1162+87C>G	rs12929392	7466/30904 0.2416	1011
F1377, F1369, F1372 (Het)	84,046,769	c.1163-112C>T	rs13336030	4991/30896 0.1615	989
F1377, F1369, F1372 (Het)	84,046,762	c.1163-105G>A	rs11864211	4993/30894 0.1616	991
F1377, F1421 (Het)	84,046,715	c.1163-58T>C	rs11864146	5229/30939 0.1690	567
F1352, F1369, F1371, F1377 (Het) F1372, F1421 (Hom)	84,046,499	c.1214+107G>A	rs11864095	9521/30852 0.3086	1829
F1371, F1372 (Het)	84,046,490	c.1214+116C>T	rs11864162	3098/30916 0.1002	206

A.14 Primers used to sequence the inversion in *SLC38A8* identified in case F1310

Primer	Sequence (5'-3')	PCR conditions
***SLC38A8*_BP1**	cttttagcaggaggcaccag	St/65A_T
***SLC38A8*_BP2**	ggtgatggaaggagacagga	St/65A_T
***SLC38A8*_BP3**	ccagtgttggaggtcaaggt	St/65A_T
***SLC38A8*_BP4**	aggcctgtccagtcagtgag	St/65A_T

Abbreviations are as follows: St = standard PCR, A_T = annealing temperature (in °C).

***SLC38A8*_BP1 with *SLC38A8*_BP3:** 961 bp in WT (first breakpoint)

***SLC38A8*_BP2 with *SLC38A8*_BP4**: 869 bp in WT (second breakpoint)

***SLC38A8*_BP1 with *SLC38A8*_BP2:** 946 bp in patient F1310 (first breakpoint)

***SLC38A8*_BP3 with *SLC38A8*_BP4:** 884 bp in patient F1310 (second breakpoint)

A.15 GATEway vectors for *SLC38A8* Midigene assay

attB sites are capitalized. Abbreviations are as follows: A_T = annealing temperature (in °C).

Primer	Sequence (5'-3')	Product (bp)	PCR conditions
*SLC38A8*_BAC-1F	GGGGACAAGTTTGTACAAAAAAGCAGGCT atccttgtgctatttgggatc	7554	68A_T
*SLC38A8*_BAC-1R	GGGGACCACTTTGTACAAGAAAGCTGGGT cagcagtcaatgtcctttgaa		

A.16 *SLC38A8* Midigene Primer Sequences

SDM_691-301_F: tacaaacacccatgccccttcccctctctc

SDM_691-301_R: gagagaggggaaaggggcatgggtgtttgta

*SLC38A8*_5F: ttgcagccatgctctgttac

*SLC38A8*_5R: ccctgacagagaaaccaagg

*SLC38A8*_6F: tcaggctatttcaagagaaacctc

*SLC38A8*_6R: tgaggccttttcctatgcac

*RHO*_3F: cggaggtcaacaacgagtct

*RHO*_5R: gtggtgagcatgcagttcc

T7 (forward): taatacgactcactataggg

CMV: cgcaaatgggcggtaggcgtg

SV40R: gaaatttgtgatgctattgc

A.17 ExomeDepth reference choices per WES sample and the number of CNVs detected by the software

Patient	No. ExomeDepth references chosen	No. CNVs detected
F1071	6/11	6
F1288	7/8	5
F1298	6/11	16
F1335	7/8	58
F1352	6/11	2
F1369	6/11	7
F1371	11/11	100
F1372	7/14	4
F1373	7/14	12
F1374	8/15	17
F1405	11/11	48

A.18 Heterozygous and homozygous polymorphisms identified in the coding region of *FOXD1*

Patient	Location on Chr5	cDNA change	aa change	dbSNP rs ID	Allele count & frequency (gnomAD)	No. Hom	CADD score
F1369 (Het)	72,743,951	c.237G>A	p.(Leu79=)	rs552853109	288/76964 0.003742	5	12.96
F1335 (Het)	72,743,389	c.798G>C	p.(Pro266=)	.	.	.	10.58
F1335, F1298 (Het)	72,743,030-72,743,044	c.1146_1160delGCAGGCCGCCGCC	p.(Gln383_Ala387del)	rs771204220	762/25206 0.03023	12	NA
F1071, F1405, F1403, F1419 (Hom) F1335, F1298, F1377, F1404, F1421 (Het)	72,742,882	c.1308A>G	p.(Ser436=)	rs2972191	59341/144878 0.4096	13426	9.692

A.19 Primer sequences used for PCR and Sanger sequencing all coding regions of *VSX2*

Exon	Forward sequence (5'-3')	Reverse sequence (5'-3')	Product (bp)
1	CCAGAGCATTAGACACCGGA	AGGAACTTTTCCGCCTGG	594
2	AGGCTTCCTGGGGAGACAG	GTTGTCGGCGAAAATAGGGT	312
3	AGAGGAAGGCGGTTCTGTCT	ATGGGCATCTGGAACCCT	283
4	ACTCCAAGCCTACAAGGGGT	TGCCTCAAGGCTCTGTCC	344
5	TGCTCGTCCTTAATTCTGGC	CCTTGCCTCAGAGAGCATC	512

A.20 Primer sequences used for PCR and Sanger sequencing all coding regions of *LAMP1*

Exon	Forward sequence (5'-3')	Reverse sequence (5'-3')	Product (bp)
7-8	CCAATGACCATTCACGTTTG	CCACACGGAAGCGACTAAAC	513